The Annals of Applied Probability
2019, Vol. 29, No. 6, 3489–3540
https://doi.org/10.1214/19-AAP1486

A MARTINGALE APPROACH FOR FRACTIONAL BROWNIAN MOTIONS AND RELATED PATH DEPENDENT PDES

BY FREDERI VIENS[1] AND JIANFENG ZHANG[2]

Michigan State University and University of Southern California

In this paper, we study dynamic backward problems, with the computation of conditional expectations as a special objective, in a framework where the (forward) state process satisfies a Volterra type SDE, with fractional Brownian motion as a typical example. Such processes are neither Markov processes nor semimartingales, and most notably, they feature a certain time inconsistency which makes any direct application of Markovian ideas, such as flow properties, impossible without passing to a path-dependent framework. Our main result is a functional Itô formula, extending the seminal work of Dupire (*Quant. Finance* **19** (2019) 721–729) to our more general framework. In particular, unlike in (*Quant. Finance* **19** (2019) 721–729) where one needs only to consider the stopped paths, here we need to concatenate the observed path up to the current time with a certain smooth observable curve derived from the distribution of the future paths. This new feature is due to the time inconsistency involved in this paper. We then derive the path dependent PDEs for the backward problems. Finally, an application to option pricing and hedging in a financial market with rough volatility is presented.

CONTENTS

Received December 2017; revised October 2018.
[1]Supported in part by NSF Grants DMS-1407762, 1734183 and 1811779.
[2]Supported in part by NSF Grant DMS-1413717.
MSC2010 subject classifications. Primary 60G22; secondary 60H20, 60H30, 35K10, 91G20.
Key words and phrases. Fractional Brownian motion, Volterra SDE, Monte Carlo methods, path dependent PDEs, functional Itô formula, rough volatility, time inconsistency.

1. Introduction.

1.1. *Background and heuristic description of the main ideas.* This paper introduces a new technique for analyzing functionals of non-Markov processes using ideas which are borrowed from the Markovian case, but necessarily require taking into account path dependence. In the Markovian case, consider the conditional expectation

$$Y_t = \mathbb{E}\big[g(X_T)|\mathcal{F}_t^X\big], \quad 0 \le t \le T,$$

where g is a continuous function, X is a Markov diffusion process and $\{\mathcal{F}_t^X\}_{0\le t\le T}$ is the filtration generated by X, it is well known that Y_t is a deterministic function of X_t only

$$Y_t = u(t, X_t),$$

and that function u solves a parabolic PDE, at least in weak form.

When $g(X_T)$ is replaced with a more general $\mathcal{F}_T^X$-measurable random variable ξ and/or X is a non-Markov diffusion process, Y_t will depend on the entire path of X up to time t:

$$Y_t = u(t, X_{[0,t]}) \quad \text{where } X_{[0,t]} := \{X_r\}_{0\le r\le t}, \tag{1.1}$$

and u solves a so-called Path Dependent PDE (PPDE, for short), which was first proposed by Peng [33]. A powerful tool to study PPDEs is Dupire's [16] functional Itô calculus; see also Cont and Fournie [10–12]. A successful viscosity theory for (fully) nonlinear PPDEs was established by Ekren et al. [17–19, 37]. We also refer to Lukoyanov [29], Peng and Wang [35], Peng and Song [34] and Cosso and Russo [13] for some different approaches and the book Zhang [40] for more references in this direction. We shall emphasize that, in all above works, the state process X is a diffusion process, in particular, it is a semimartingale under all involved probabilities.

In this paper, we are interested in extending the above path-dependent analysis to more "heavily" non-Markov processes X, beyond the semimartingale framework. Typical examples of such non-Markov X are Gaussian processes with memory properties, such as the fractional Brownian motion (fBm). When X is fBm or similar processes, if there is a hope to replicate PDE-type ideas for representing Y_t even in the state dependent case $\xi = g(X_T)$, then any representation using a deterministic function u will necessarily depend on the entire path of X up to t, namely in the form of (1.1).

How to figure out this dependence, and how to find the deterministic function u in as explicit a way as possible (theoretically or numerically), is what this paper is about. To illustrate our main idea, in Section 2 we consider a special case that

$$X_t = \int_0^t K(t, r)\, dW_r, \tag{1.2}$$

where W is a Brownian motion and K is a deterministic kernel, and thus X is a Volterra-type Gaussian process. This is in particular the case for fBM. Our main idea is to introduce the following simple but crucial auxiliary process Θ with two time variables:

$$\Theta_s^t := \int_0^t K(s, r)\, dW_r. \tag{1.3}$$

This process Θ enjoys many nice properties, as we explain in details below.

We first note that, for any fixed s, the process $t \in [0, s] \mapsto \Theta_s^t$ is a martingale. The existing theory of PPDEs relies heavily on the semimartingale theory, while X is not a semimartingale. So by adding Θ as our "state process," we will be be able to exploit its (semi)martingale property, and thus recover the PPDE language. We remark that, for $s > t$, we have the orthogonal decomposition: $X_s = \Theta_s^t + [X_s - \Theta_s^t]$. This elementary property is a common computational tool in stochastic analysis, used in many studies regarding fBm and related processes (see the textbook by Nualart, [31], Chapter 5). However, we believe our paper is the first instance where this property is applied in the context of PPDEs; the reason for this may be that the property is usually invoked to exploit the independent part $X_s - \Theta_s^t$ of the decomposition, and the martingale property of Θ is rarely exploited.

Next, the process X typically violates the standard flow property, which is another major obstacle for using PDEs and PPDEs. The introduction of the martingale component Θ is the key for recovering the flow property, or say (X, Θ) together will enjoy certain "Markov" property. More precisely, we shall rewrite (1.1) as

$$Y_t = u\big(t, X_{[0,t)} \otimes \Theta_{[t,T]}^t\big), \tag{1.4}$$

and show that this function u satisfies a PPDE. While in the standard literature on PPDEs as mentioned earlier, u depends only on the stopped path $X_{[0,t]}$, in our situation u will depend on the path $\Theta_{[t,T]}^t$ as well. This is the major difference between Dupire's functional Itô calculus and our extension. We also note that, when $K(t, r) = K(r)$, then $\Theta_s^t = X_t$ for $t \le s \le T$, and thus (1.4) reduces back to (1.1).

Moreover, the introduction of Θ is also crucial for numerical computation of the u in (1.4). On one hand, writing u as the solution to a PPDE enables us to extend the existing numerical methods for standard PDEs to PPDEs naturally, which will be carried out in a separate project. On the other hand, we note that the function

u in (1.4) is continuous under mild conditions, which is important for numerical purpose. However, Θ is typically discontinuous in X, so if we write Y as a function of X only, the function u in (1.1) could be discontinuous, and thus its numerical methods would be less efficient.

Finally, we discuss the tractability of the process Θ, which is important both for numerical purpose and for applications, and we shall discuss it more when we consider a financial application in the next subsection and Section 5. First, we note that Θ^t_s is $\mathcal{F}^W_t$-measurable, so mathematically all our analysis will have no measurability issues. However, in many applications people may not observe W. Fortunately, for the models we will consider, Θ^t_s will be measurable to the observed information. In particular, when X is a fBM, actually we have $\mathbb{F}^X = \mathbb{F}^W$ since one can represent W as a function of X through certain transform operator; see the textbook by Nualart, [31], Chapter 5. Also, see Mocioalca and Viens [30] for the case that K is of convolution form, which covers the so-called Riemann–Liouville fBm. Then it is legitimate to use Θ^t_s at time t, provided we observe $X_{[0,t]}$.

Having explained the ideas in details, in Section 3 we turn to the general framework where X solves a Volterra SDE; see (3.1) below. In this case, the corresponding Θ will be a semimartingale, and still shares all the nice properties discussed above. Our main technical result is a functional Itô formula for functions u of the form (1.4), extending Dupire's [16] functional Itô calculus which involves only the stopped process $X_{[0,t]}$. We remark that in Dupire's calculus the spatial derivatives involve only perturbations of X_t, but not of the path before t: $X_{[0,t)}$. The main feature of our extension is again that the state variable contains the auxiliary process Θ, and our path derivatives will involve only the perturbation of $\Theta^t_{[t,T]}$, which is exactly in the spirit of Dupire. This is important because on one hand $\Theta^{\cdot}_s$ is a semimartingale so an Itô formula involving its derivatives is possible, and on the other hand $X_\cdot$ is not a semimartingale so its derivatives are not helpful for Itô calculus and should be avoided. We note that Dupire's calculus serves as an alternative to the Malliavin calculus, and appears as a simpler calculus of variations, when questions of measurability with respect to current information are crucial. Said differently, the Malliavin calculus can be viewed as an overkill from the standpoint of keeping track of this adaptability, since it applies equally well to anticipating processes.

Section 4 applies our functional Itô formula to solve the backward problems in such framework and obtain naturally the PPDEs. We shall formulate it as a nonlinear Backward SDE (BSDE), whose linear version is essentially the conditional expectation $\mathbb{E}[g(X_\cdot)|\mathcal{F}_t]$ as discussed in Section 2. Such nonlinear problems have many applications, especially in finance, stochastic control and probabilistic numerical methods. We identify the corresponding semilinear PPDE, and assuming a classical solution exists, it yields immediately a solution to the BSDE. This strategy, known as a nonlinear Feynman–Kac formula, goes back to the original work of Peng [36]. Section 4 also provides a brief discussion of fully nonlinear PPDEs, corresponding to stochastic optimization problems with diffusion controls in our framework.

In Section 5, we apply our methodology to the option pricing and hedging problem in a rough volatility model, motivated by the recent work El Euch and Rosenbaum [20]. We discuss this and more general financial applications in the next subsection.

1.2. *Application in rough volatility models*. Consider a standard stock price model under risk neutral probability:

$$dS_t = \sigma_t S_t \, dB_t, \tag{1.5}$$

where B is a Brownian motion, σ is the volatility process and we are assuming zero interest rate for simplicity. A number of recent studies have questioned the possibility of assuming that σ is a Markov process. In continuous time, the first paper to work with this assumption is Comte and Renault [9], in which σ is assumed to be driven by an fBm. So-called continuous-time long-memory models of that sort have been the main source of highly non-Markov volatility model. The paper, Chronopoulou and Viens [8], can be consulted for references to such works, and pertains to validating and calibrating this type of model from option data. Fractional stochastic volatility models continue to draw lots of interest. A notable work is Gatheral, Jaisson and Rosenbaum [24], which finds market evidence that volatility's high-frequency behavior could be modeled as a rough path, for example, based on fBm with $H \in (0, 1/2)$, and thus introduced the rough volatility models; see also Bennedsen, Lunde and Pakkanen [4]. Among many others, Abi Jaber, Larsson and Pulido [1] and Gatheral and Keller-Ressel [25] studied affine variance models, where σ^2 is modeled as a convolution type linear Volterra SDE, which in particular includes the rough Heston model studied in El Euch and Rosenbaum [20, 21]; Bayer, Friz and Gatheral [3] studied a rough Bergomi model; Cuchiero and Teichmann [14] studied affine Volterra processes with jumps; Gulisashvili, Viens and Zhang [26] provided an asymptotic analysis applying to short-time fBm-modeling of volatility for fixed-income securities near maturity; and Fouque and Hu [22] studied a portfolio optimization problem in a model with fractional Ornstein–Uhlenbeck process.

We remark that our model of general Volterra SDEs covers all the models mentioned above, except the jump model in [14] which we believe can be dealt with by extending our work to PPDEs on càdlàg paths; see Keller [27] where the state process is a standard jump diffusion. Several works in the literature, for example, [3, 20, 25] have already used the forward variance processes $\hat{\Theta}^t_s := \mathbb{E}[\sigma_s^2|\mathcal{F}_t]$, which is closely related to our process Θ^t_s. Indeed, in the affine variance models, they can be transformed from one to the other, as we will see in Section 5. However, for general models, especially when the drift of the Volterra SDE is nonlinear, we believe our process Θ is intrinsic and is more convenient. Moreover, we note that most works in the literature either focus on modeling the financial market, or on pricing contingent claims in rough volatility models. We shall provide a systematic study on the hedging of contingent claims in general rough volatility models, motivated by the work [20]. Finally, we allow the backward process to be nonlinear, which appears naturally in applications, for example, when the borrowing and lending interest rates are different or when a control is involved; and we allow the payoff to be path dependent, thus including Asian options and lookback options.

We now focus on the hedging issue. Consider the model (1.5) and assume σ^2 is modeled through certain Volterra SDE which induce the crucial auxiliary process Θ. Then the price of the contingent claim will take the form $Y_t = u(t, S_t, \sigma^2_{[0,t)} \otimes_t \Theta^t_{[t,T]})$ and u solves a PPDE. We note that the market is typically incomplete when S is the only tradable instruments. Applying our functional Itô formula, we will see in Section 5 that the contingent claims can be hedged by using S and Θ (in appropriate sense), and the hedging portfolios are exactly in the spirit of the Δ-hedging, as derivatives of the function u with respect to S and Θ, respectively. Then the issue boils down to whether or not we can hedge Θ by using tradable assets in the market. Note that Θ is defined through σ^2, so the key is to understand the variance σ^2 or the volatility σ.

First, note that by observing S continuously in time, mathematically we may compute σ from it, or say σ is $\mathbb{F}^S$-measurable. In fact, based on this fact, in the financial market there are proxies for σ, such as the VIX index from the Chicago Board of Options Exchange, which is a proxy for the volatility on the S&P500 index. This VIX index has become so mature that even skeptics when it comes to volatility quotes will argue that if S in (1.5) is a model for the S&P500, then both S and σ are observable stochastic processes. Consequently, depending on the market and on the assumptions one is willing to make regarding volatility quotes, we may assume that (S, σ) is observed. This clearly has advantage compared to computing σ from S when numerical methods are concerned.

Next, note that in the simple setting (1.2)–(1.3) and assume for simplicity that $\sigma = X$, we have $\Theta^t_s = \mathbb{E}[\sigma_s|\mathcal{F}_t]$. Mathematically, to compute Θ from the observed information σ, one needs to first compute W from σ by using certain transfer operator as mentioned before, and then do the pathwise stochastic integration in (1.3). Numerically this will be very expensive. However, in the financial context, the expression $\mathbb{E}[\sigma_s|\mathcal{F}_t]$ above is also known as the forward volatility at time t with horizon $s > t$, and that is a market observable. For the S&P500 and many other equities, it is directly quoted by use of implied volatility.

Finally, for the rough volatility models we will consider in Section 5, we will be able to replicate Θ^t_s by using the forward variance $\mathbb{E}[\sigma^2_s|\mathcal{F}_t]$ (similar to the forward volatility as we just discussed), which can be further replicated (approximately) by using the variance swaps. See more details in Section 5. We can therefore conclude that we are able to hedge the contingent claims by using the tradable assets S and the forward variance $\mathbb{E}[\sigma^2_s|\mathcal{F}_t]$.

2. The flow property of fBm.

This section provides simple heuristics in the case of $X =$ fBm for easily tractable examples.

2.1. *Martingale decomposition of fBm.* For simplicity in this section, we restrict to one-dimensional processes only. Let B^H be a fBm with Hurst parameter $H \in (0,1)$. As we mentioned in the Introduction, Chapter 5 in the textbook by Nualart [31] explains that there exists a Brownian motion (standard Wiener process) W and an explicitly known deterministic kernel $K(t,s) > 0$ such that

$$B_t^H = \int_0^t K(t,r)\,dW_r \quad \text{and} \quad \mathbb{F}^{B^H} = \mathbb{F}^W =: \mathbb{F}, \tag{2.1}$$

where the notation $\mathbb{F}^X$ is the filtration of $X : \mathbb{F}^X = \{\mathcal{F}_t^X : t \geq 0\}$. The inclusion $\mathbb{F}^{B^H} \subset \mathbb{F}^W$ is immediate. The reverse inclusion comes from the existence of a bijective transfer operator to express W as a Wiener integral with respect to B^H. We remark that, among others, one main feature of fBM is the violation of the standard flow property which can be viewed as certain time inconsistency: for $0 \leq t < s \leq T$,

$$B_s^H \neq \tilde{B}_s^{t,H} \quad \text{where } \tilde{B}_s^{t,H} := B_t^H + \int_t^s K(s,r)\,dW_r. \tag{2.2}$$

We are interested in backward problems. For $\xi = g(B_\cdot^H) \in \mathbb{L}^2(\mathcal{F}_T)$, denote

$$Y_t := \mathbb{E}[\xi|\mathcal{F}_t]. \tag{2.3}$$

Clearly, Y is a martingale. Our goal is to characterize Y from the PPDE point of view. Due to (2.1), it is clear that

$$Y_t = u_1\big(t, B_{[0,t]}^H\big) = u_2(t, W_{[0,t]}), \tag{2.4}$$

for some measurable functions u_1, u_2. Since B^H is not a semimartingale (when $H \neq \frac{1}{2}$), it is difficult to derive a PPDE for u_1. On the other hand, since W is a standard Brownian motion, formally u_2 should satisfy a path dependent heat equation. Indeed, this is true if u_2 is continuous in W in the topology of uniform convergence. However, provided g is continuous in B^H, since B^H is discontinuous in W in pathwise sense, it is unlikely that u_2 will have desired pathwise regularity.

To get around of this, we will utilize the following simple but crucial decomposition:

$$B_s^H = \Theta_s^t + I_s^t := \int_0^t K(s,r)\,dW_r + \int_t^s K(s,r)\,dW_r, \quad 0 \leq t \leq s \leq T, \tag{2.5}$$

where

$$\Theta_s^t := \int_0^t K(s,r)\,dW_r = \mathbb{E}_t\big[B_s^H\big] \tag{2.6}$$

is $\mathcal{F}_t$ measurable, and I_s^t is independent of $\mathcal{F}_t$. We note that when s is fixed, the process $t \in [0,s] \mapsto \Theta_s^t$ is a $\mathbb{F}$-martingale.

REMARK 2.1. (i) If we use the rough path norms instead of the uniform convergence topology, then B^H can be continuous in W. However, this requires a weaker regularity on the function u_2 in (2.4), which would induce serious difficulties for the functional Itô formula (3.16) below, and we do not see how to use the existing PPDE theory to exploit the rough-path dependence of B^H on W. Further exploration in this direction may be worthwhile, but is beyond the scope of this paper.

(ii) Alternatively one may view the u_2 in (2.4) as a weak solution to the path dependent heat equation, in Sobolev sense without requiring pathwise regularity as in Cont [12]. However, the regularity itself is interesting and important, for example, when one considers numerical methods. Moreover, in applications typically one observes B^H. Although in theory one may obtain W from B^H through the bijective transfer operator, such operator is not explicit, and thus in practice it is not convenient to use the information W. As we will see soon, we shall express Y through Θ which is more trackable in many applications.

2.2. *A state dependent case*. To begin by illustrating our idea of using the process $t \mapsto \Theta_s^t$ in the simplest possible context, we consider a very special case:

$$\xi = g(B_T^H).$$

By the martingale/orthogonal decomposition (2.5), we have

$$Y_t = u(t, \Theta_T^t) \quad \text{where } u(t,x) := \mathbb{E}[g(x + I_T^t)].$$

Assuming g is smooth, then the regularity of u is clear. Moreover, since $t \mapsto \Theta_T^t$ is a martingale, applying the standard Itô formula we obtain

$$dY_t = \left[\partial_t u(t, \Theta_T^t) + \frac{1}{2}\partial_{xx}^2 u(t, \Theta_T^t) K^2(T,t)\right] dt + \partial_x u(t, \Theta_T^t) K(T,t)\, dW_t.$$

Noticing that Y is a martingale by definition (2.3), this implies that the drift term above must vanish, that is,

$$\partial_t u(t,x) + \frac{1}{2} K^2(T,t) \partial_{xx}^2 u(t,x) = 0, \qquad u(T,x) = g(x). \tag{2.7}$$

This very simple (backward) heat equation with a time-dependent diffusion coefficient shows how to compute the martingale Y by tracking the observed martingale process $t \mapsto \Theta_T^t$ and solving a deterministic problem (2.7) for u.

REMARK 2.2. If one were merely interested in finding a PDE representation of Y at a given time, say Y_0, a natural solution, in analogy to the Brownian case, would emerge. Define

$$\tilde{u}(t,x) := \mathbb{E}[g(x + B_T^H - B_t^H)] = \mathbb{E}[g(x + B_{T-t}^H)],$$

where the second equality holds because of the stationarity of increments of fBm. Note that $x + B^H_{T-t} \sim Normal(x, (T-t)^{2H})$, then

$$\tilde{u}(t,x) = \int_{\mathbb{R}} g(y) p^H(T-t, y-x)\,dt \quad \text{where } p^H(t,x) := \frac{1}{\sqrt{2\pi} t^H} e^{-\frac{x^2}{2t^{2H}}}.$$

One may check straightforwardly that $\partial_t p^H(t,x) = Ht^{2H-1}\partial_{xx} p^H(t,x)$. Then $\tilde{u}$ solves the following PDE:

$$\partial_t \tilde{u} + H(T-t)^{2H-1}\partial_{xx}\tilde{u}(t,x) = 0, \qquad \tilde{u}(T,x) = g(x).$$

This PDE was already obtained by Decreusefond and Ustunel [15] (see also Baudoin and Coutin [2] for a more general result in this direction) and it does not look very different from the previous one we identified. Note that we still have $\tilde{u}(T, B^H_T) = \xi$ and $\tilde{u}(0, B^H_0) = Y_0$, however, $\tilde{u}(t, B^H_t)$ is not a martingale and in particular $\tilde{u}(t, B^H_t) \neq Y_t$ for $0 < t < T$. This is due to the fact that the natural decomposition $B^H_T = B^H_t + (B^H_T - B^H_t)$ is not an orthogonal decomposition, namely B^H_t and $B^H_T - B^H_t$ are not independent.

The standard technique in the Brownian case W, for which this decomposition over increments works so well, happens to coincide with the use of Θ since $\Theta^t_T = W_t$ and $I^t_T = W_T - W_t$. But when trying the same increments trick with B^H_t instead of W, since $\tilde{u}(t, B^H_t) \neq Y_t$, the PDE above does not help track the value of conditional expectations dynamically. Thus the orthogonal decomposition (2.5) is preferable for purposes, such as in stochastic finance, where $t \mapsto Y_t$ needs to be evaluated dynamically: we want to use a single PDE (or PPDE) to represent all values of Y and for this, we need to track Θ.

2.3. *A simple path dependent case.* We now consider the case

$$\xi = g(B^H_T) + \int_0^T f(t, B^H_t)\,dt. \tag{2.8}$$

As we alluded to in the Introduction, this is a typical example useful in finance, for instance as a model of a portfolio utility with legacy (g) and consumption (f) terms, or as a contingent claim with straightforward path dependence such as in Asian options. Because of the explicit path dependence, we contend that our framework based on tracking Θ can handle this dependence without any additional effort beyond what needs to be deployed to handle the stochastic path dependence in B^H.

REMARK 2.3. In the special case $H = 1/2$, that is, B^H is a standard Brownian motion, which we denote by W, we have

$$\tilde{Y}_t := Y_t - \int_0^t f(s, W_s)\,ds = \mathbb{E}\Big[g(W_T) + \int_t^T f(s, W_s)\,ds \Big| \mathcal{F}_t\Big] = \tilde{u}(t, W_t),$$

where $\tilde{u}$ satisfies a standard backward heat equation with additive forcing:

$$\partial_t \tilde{u} + \frac{1}{2}\partial^2_{xx}\tilde{u} + f(t,x) = 0, \qquad \tilde{u}(T,x) = g(x).$$

In the general fBm case, however, the above $\tilde{Y}_t$ is not Markovian anymore. Instead, we use the same idea that leads to Θ to the expression for ξ: we decompose the conditional expectation of the integral part in Y_t from (2.3) into a term which is observable at time t, and other terms. Those other terms are not independent of the past, but we can apply the results of the previous section directly to them, whether at the terminal time, or for each s in the integral from t to T. Thanks to the explicit PDE (2.7) from that section, we obtain

$$\begin{aligned} Y_t &= \int_0^t f(s, B^H_s)\,ds + \mathbb{E}[g(B^H_T)|\mathcal{F}_t] + \int_t^T \mathbb{E}[f(s, B^H_s)|\mathcal{F}_t]\,ds \\ &= \int_0^t f(s, B^H_s)\,ds + u_g(T; t, \Theta^t_T) + \int_t^T u_f(s; t, \Theta^t_s)\,ds, \end{aligned}$$

where

$$\begin{aligned} &\partial_t u_g(T;t,x) + \frac{1}{2}K^2(T,t)\partial^2_{xx}u_g(T;t,x) = 0, \quad 0 \le t \le T, \\ &\partial_t u_f(s;t,x) + \frac{1}{2}K^2(s,t)\partial^2_{xx}u_f(s;t,x) = 0, \quad 0 \le t \le s, \qquad (2.9) \\ &u_g(T;T,x) = g(x), \qquad u_f(s;s,x) = f(s,x). \end{aligned}$$

Therefore, we discover that Y_t can be expressed as a single deterministic function $u(t,\cdot)$ of the concatenated path which equals B^H up to time t and equals Θ^t afterwards:

$$\begin{aligned} &Y_t = u\left(t, \left\{\int_0^{t\wedge s} K(s,r)\,dW_r\right\}_{0\le s\le T}\right) = u(t, B^H \otimes_t \Theta^t) \\ &\text{where } (\omega \otimes_t \theta)_s := \omega_s \mathbf{1}_{[0,t)}(s) + \theta_s \mathbf{1}_{[t,T]}(s), \qquad (2.10) \\ &u(t, \omega \otimes_t \theta) := \int_0^t f(s,\omega_s)\,ds + u_g(T;t,\theta_T) + \int_t^T u_f(s;t,\theta_s)\,ds. \end{aligned}$$

The last line above clearly shows how $u(t,\cdot)$ is an explicit function of the entire concatenated path $\omega \otimes_t \theta$. From that standpoint, at least in this example, we have succeeded in representing the conditional expectation Y dynamically thanks to a single deterministic path-dependent functional, by tracking Θ. We remark that the above function u is continuous in $(t, \omega \otimes_t \theta)$ under mild and natural conditions. This is reassuring for our goal, which is to introduce appropriate time and path derivatives and then derive a path dependent PDE for the above u and in more general contexts. Moreover, such regularity is important when one considers numerical methods, even though it is not the focus of this paper.

Let us first take a look heuristically at the above example. Differentiating u formally,

$$\partial_t u(t, \omega \otimes_t \theta) = f(t, \omega_t) + \partial_t u_g(T; t, \theta_T) - u_f(t; t, \theta_t) + \int_t^T \partial_t u_f(s; t, \theta_s)\, ds.$$

Note that

$$u_f(t; t, \theta_t) = f(t, \theta_t) = f(t, \omega_t) \quad \text{provided } \theta_t = \omega_t,$$

resulting a corresponding cancellation in $\partial_t u$. This condition $\theta_t = \omega_t$ simply requires continuity of the path $\omega \otimes_t \theta$ at the point $s = t$. This is certainly the case for us in this section since $\Theta_t^t = B_t^H$, and will remain true for any continuous Gaussian process and for non-Gaussian processes of interest, as can be seen in (3.3) below. Then, by (2.9) we have

$$\begin{aligned}
\partial_t u(t, \omega \otimes_t \theta) &= \partial_t u_g(T; t, \theta_T) + \int_t^T \partial_t u_f(s; t, \theta_s)\, ds \\
&= -\frac{1}{2} K^2(T, t) \partial_{xx}^2 u_g(T; t, \theta_T) \\
&\quad - \frac{1}{2} \int_t^T K^2(s, t) \partial_{xx}^2 u_f(s; t, \theta_s)\, ds \\
&= -\frac{1}{2} K^2(T, t) \partial_{\theta_T \theta_T}^2 u(t, \omega \otimes_t \theta) \\
&\quad - \frac{1}{2} \int_t^T K^2(s, t) \partial_{\theta_s \theta_s}^2 u(t, \omega \otimes_t \theta)\, ds,
\end{aligned} \tag{2.11}$$

with the terminal condition $u(T, \omega \otimes_T \theta) = g(\omega_T) + \int_0^T f(t, \omega_t)\, dt$. The last equality in (2.11) comes from the expression for u in (2.10). In that line, the notation $\partial_{\theta_s \theta_s}^2 u$ means that this is a derivative with respect to the value of the path $\omega \otimes_t \theta$ at time $s \in [t, T]$.

In any case, at least heuristically, one sees that u itself appears to solve a PPDE in some sense. We will make this sense precise in the next section, in Theorem 4.1. One observation is that the spatial derivatives of u above are only with respect to θ, not to ω, which is crucial in our context because ω corresponds to B^H which is not a semimartingale, and we wish to use the semimartingale property to transport stochastic objects (such as conditional expectations) to their deterministic representations.

REMARK 2.4. (i) The kernel K involves two time variables, and thus B^H is not a semimartingale (when $H \neq \frac{1}{2}$). Moreover, B^H violates the standard flow property, see (2.2), and is by nature time inconsistent. The introduction of the term Θ is the key for recovering the flow property for B^H, which is crucial for deriving the corresponding PPDE.

(ii) In the standard functional Itô calculus of Dupire [16], the function u depends only on the stopped path $X_{t\wedge\cdot}$, or equivalently, in their situation Θ^t is flat. So our main result in the next section is indeed a nontrivial extension of Dupire's result.

(iii) We also note that the occurrence of derivatives of u with respect only to the "θ" portion of the path does not contradict Dupire's functional Itô calculus, because as just mentioned, in his setting $\theta_s = \omega_t$ for all $s \in [t, T]$, and thus the derivatives with respect to θ reduce to the derivative with respect to ω_t alone.

3. Functional Itô formula. In this section, we expand our framework's reach by considering more general processes X, beyond the Gaussian class. Assume X is a solution to the d-dimensional Volterra SDE:

$$X_t = x + \int_0^t b(t; r, X_\cdot)\, dr + \int_0^t \sigma(t; r, X_\cdot)\, dW_r, \quad 0 \le t \le T, \tag{3.1}$$

where W is a standard (possibly multidimensional) Wiener process, and b and σ have appropriate dimensions and are adapted in the sense that $\varphi(t; r, X_\cdot) = \varphi(t; r, X_{r\wedge\cdot})$ for $\varphi = b, \sigma$. As in (2.2), one main feature of such SDE is that it violates the flow property. That is, $X_s \neq \tilde{X}^t_s$ for $0 \le t < s \le T$, where $\tilde{X}^t$ is the solution to the following SDE:

$$\tilde{X}^t_s = X_t + \int_t^s b\big(s; r, X \otimes_t \tilde{X}^t_\cdot\big)\, dr + \int_t^s \sigma\big(s; r, X \otimes_t \tilde{X}^t\big)\, dW_r, \quad t \le s \le T.$$

Throughout the paper, the following assumption will always be in force.

ASSUMPTION 3.1. (i) The SDE (3.1) admits a weak solution (X, W).
(ii) $\mathbb{E}[\sup_{0\le t\le T} |X_t|^p] < \infty$ for all $p \ge 1$.

The condition (ii) is technical, and in order not to distract our main focus, we postpone its discussion to the Appendix. For condition (i), there have been many works on well-posedness of Volterra SDEs; see, for example, Berger and Mizel [5, 6]. In this paper, we prefer not to restrict to specific conditions so as to allow for the most generality, and in applications any reasonable model should admit at least one solution. However, we would like to mention that, since in most applications X is the observable state process and W is just used to model the distribution of X, it suffices to consider a weak solution. Moreover, no uniqueness of weak solution is needed. So from now on, we will always fix a weak solution (X, W), and slightly unlike in Section 2, we shall always use the full filtration:

$$\mathbb{F} = \mathbb{F}^{X,W}. \tag{3.2}$$

In this framework, the analogue of the martingale term in the decomposition of X can be defined using exactly the same idea as in the Gaussian case, by basing it on (3.1) rather than (2.1). Thus we denote

$$\Theta^t_s := x + \int_0^t b(s; r, X_\cdot)\, ds + \int_0^t \sigma(s; r, X_\cdot)\, dW_r, \quad t \le s \le T, \tag{3.3}$$

where $t \mapsto \Theta^t_s$ is a semimartinagle. Moreover, we denote

$$\check{X}^t_s := X \otimes_t \Theta^t \quad \text{namely } \check{X}^t_s := X_s \mathbf{1}_{[0,t)}(s) + \Theta^t_s \mathbf{1}_{[t,T]}(s). \tag{3.4}$$

To simplify the notation, quite often we omit the X in b and σ, and simply write $\varphi(t,s) = \varphi(t; s, X.)$ for $\varphi = b, \sigma$.

REMARK 3.2. As we will see in the paper, we shall write the interested value process Y_t as a function of the paths $X \otimes_t \Theta^t$, which we observe and the function u is typically continuous under mild conditions. We note that Θ^t_s is also a function of $X_{[0,t]}$. However, this dependence is typically discontinuous under uniform convergence. For example, set

$$b = 0; \qquad \sigma(t; s, \omega) = 1, \quad t \in \left[0, \frac{T}{2}\right];$$
$$\sigma(t; s, \omega) = 1 + \left[t - \frac{T}{2}\right]\sigma_0(s, \omega), \quad t \in \left(\frac{T}{2}, T\right],$$

for some appropriate function σ_0. Then $X_{[0,\frac{T}{2}]} = W_{[0,\frac{T}{2}]}$, and, for $0 < t \le \frac{T}{2}$, $\Theta^t_T = W_t + \frac{T}{2}\int_0^t \sigma_0(s, W.)\,dW_s$. This involves a stochastic integral and is typically discontinuous in pathwise sense. Consequently, if we rewrite Y_t as a function of $X_{[0,t]}$ only, the function could be discontinuous. Besides theoretical interest, such regularity is crucial when one studies numerical methods for the related problems.

3.1. *The path derivatives.* As in Dupire [16], though in the end all paths are continuous, since we employ some piecewise-continuous approximations, we must extend the sample space to the càdlàg space D^0. Denote

$$\Omega := C^0([0,T], \mathbb{R}^d), \qquad \overline{\Omega} := D^0([0,T], \mathbb{R}^d), \qquad \Omega_t := C^0([t,T], \mathbb{R}^d);$$
$$\Lambda := [0,T] \times \Omega, \qquad \overline{\Lambda} := \{(t,\omega) \in [0,T] \times \overline{\Omega} : \omega|_{[t,T]} \in \Omega_t\};$$
$$\|\omega\|_T := \sup_{0 \le t \le T} |\omega_t|, \qquad \mathbf{d}((t,\omega),(t',\omega')) := |t - t'| + \|\omega - \omega'\|_T.$$

Here, we change our notation slightly compared to what we had used in the illustrative examples of the previous section. Indeed, we see here that ω is defined on $[0,T]$, whereas previously we used the letter ω for paths on $[0,t)$. The correspondence between these two conventions is that what we now call $\omega \mathbf{1}_{[0,t)}$ and $\omega \mathbf{1}_{[t,T]}$ correspond to the ω and θ in (2.10), respectively. Though the old convention was natural because it highlighted the concatenation of the path of X up to t with the path of its observable martingale component Θ^t after t, the new convention does not presume that this is the structure of the full path ω, and allows more compact notation. Moreover, we emphasize that in this subsection all the terms are deterministic and there is no probability involved.

The space we really care about is Λ, however, for technical reasons we need to allow ω to be discontinuous on $[0, t]$. We note that Dupire's framework is covered by our setup since Dupire's space is the subset of those ω which are constant on $[t, T]$. We also note that, for any $(t, \omega) \in \overline{\Lambda}$ and $t' > t$, we have $(t', \omega) \in \overline{\Lambda}$. Let $C^0(\overline{\Lambda})$ denote the set of functions $u : \overline{\Lambda} \to \mathbb{R}$ continuous under $\mathbf{d}$. For $u \in C^0(\overline{\Lambda})$, define

$$\partial_t u(t, \omega) := \lim_{\delta \downarrow 0} \frac{u(t + \delta, \omega) - u(t, \omega)}{\delta} \quad \text{for all } (t, \omega) \in \overline{\Lambda}, \tag{3.5}$$

provided the limit exists. We note that here $\partial_t u$ is actually the right time derivative.

We next define the spatial derivative with respect to ω. Given $(t, \omega) \in \overline{\Lambda}$, we define $\partial_\omega u(t, \omega)$ as the Fréchet derivative with respect to $\omega \mathbf{1}_{[t,T]}$ which is a linear operator on Ω_t:

$$\begin{aligned} &u(t, \omega + \eta \mathbf{1}_{[t,T]}) - u(t, \omega) \\ &\quad = \langle \partial_\omega u(t, \omega), \eta \rangle + o(\|\eta \mathbf{1}_{[t,T]}\|_T) \quad \text{for any } \eta \in \Omega_t. \end{aligned} \tag{3.6}$$

It is clear that this is equal to the Gateux derivative:

$$\langle \partial_\omega u(t, \omega), \eta \rangle = \lim_{\varepsilon \to 0} \frac{u(t, \omega + \varepsilon \eta \mathbf{1}_{[t,T]}) - u(t, \omega)}{\varepsilon} \quad \text{for any } \eta \in \Omega_t. \tag{3.7}$$

We emphasize that the above perturbation is only on $[t, T]$, not on $[0, t)$. This is consistent with Dupire's derivative. For any $s < t$ and $\eta \in \Omega_s$, we will take the convention that

$$\langle \partial_\omega u(t, \omega), \eta \rangle := \langle \partial_\omega u(t, \omega), \eta \mathbf{1}_{[t,T]} \rangle. \tag{3.8}$$

DEFINITION 3.3. Let $u \in C^0(\overline{\Lambda})$ such that $\partial_\omega u$ exists for all $(t, \omega) \in \overline{\Lambda}$.

(i) We say $\partial_\omega u$ has polynomial growth if there exist constants $C > 0$, $\kappa > 0$ such that

$$|\langle \partial_\omega u(t, \omega), \eta \rangle| \le C[1 + \|\omega\|_T^\kappa] \|\eta \mathbf{1}_{[t,T]}\|_T \quad \text{for all } (t, \omega) \in \overline{\Lambda}, \eta \in \Omega. \tag{3.9}$$

(ii) We say $\partial_\omega u$ is continuous if, for any $\eta \in \Omega$, the mapping $(t, \omega) \in \overline{\Lambda} \mapsto \langle \partial_\omega u(t, \omega), \eta \rangle$ is continuous under $\mathbf{d}$.

Throughout the paper, we use κ to denote a generic order of polynomial growth, which may vary from line to line. We note that, when $\partial_\omega u$ is continuous, it is clear that the mapping $\lambda \in [0, 1] \mapsto u(t, \omega + \lambda \eta \mathbf{1}_{[t,T]})$ is continuously differentiable, and thus

$$u(t, \omega + \eta \mathbf{1}_{[t,T]}) - u(t, \omega) = \int_0^1 \langle \partial_\omega u(t, \omega + \lambda \eta \mathbf{1}_{[t,T]}), \eta \rangle \, d\lambda, \tag{3.10}$$

for any $\eta \in \Omega_t$. Define further the second derivative $\partial^2_{\omega\omega} u(t,\omega)$ as a bilinear operator on $\Omega_t \times \Omega_t$:

$$\begin{aligned}(3.11)\qquad &\langle \partial_\omega u(t, \omega + \eta_1 \mathbf{1}_{[t,T]}), \eta_2\rangle - \langle \partial_\omega u(t,\omega), \eta_2\rangle \\ &\quad = \langle \partial^2_{\omega\omega} u(t,\omega), (\eta_1, \eta_2)\rangle + o(\|\eta_1 \mathbf{1}_{[t,T]}\|_T),\end{aligned}$$

for any $\eta_1, \eta_2 \in \Omega_t$. Similarly, define $\langle \partial^2_{\omega\omega} u(t,\omega), (\eta_1,\eta_2)\rangle$ for $\eta_1, \eta_2 \in \Omega_s$ as in (3.8), and define the polynomial growth and continuity in the spirit of Definition 3.3, with $\|\eta \mathbf{1}_{[t,T]}\|_T$ in (3.9) replaced with $\|\eta_1 \mathbf{1}_{[t,T]}\|_T \|\eta_2 \mathbf{1}_{[t,T]}\|_T$.

DEFINITION 3.4. We say $u \in C^{1,2}(\overline{\Lambda}) \subset C^0(\overline{\Lambda})$ if $\partial_t u$, $\partial_\omega u$, $\partial^2_{\omega\omega} u$ exist and are continuous on $\overline{\Lambda}$. Let $C^{1,2}_+(\overline{\Lambda})$ be the subset of $C^{1,2}(\overline{\Lambda})$ such that all the derivatives have polynomial growth, and $\langle \partial^2_{\omega\omega} u, (\eta,\eta)\rangle$ is locally uniformly continuous in ω with polynomial growth, that is, there exist $\kappa > 0$ and a bounded modulus of continuity function ρ such that, for any $(t,\omega), (t,\omega') \in \overline{\Lambda}$ and $\eta \in \Omega_t$,

$$\begin{aligned}(3.12)\qquad &\big|\langle \partial^2_{\omega\omega} u(t,\omega) - \partial^2_{\omega\omega} u(t,\omega'), (\eta,\eta)\rangle\big| \\ &\quad \le \big[1 + \|\omega\|_T^\kappa + \|\omega'\|_T^\kappa\big] \|\eta \mathbf{1}_{[t,T]}\|_T^2 \rho(\|\omega - \omega'\|_T).\end{aligned}$$

REMARK 3.5. Cont and Fournier [11] established the functional Itô formula in their framework for all $u \in C^{1,2}(\overline{\Lambda})$, by using the standard localization techniques with stopping times. In their framework, only $(t, \omega \mathbf{1}_{[0,t]})$ is involved, and thus it is sufficient to consider the stopped paths $\omega_{t\wedge\cdot}$. However, in our framework, the whole path of ω on $[0,T]$ is involved, we have difficulty to apply the localization techniques directly. Thus in this paper we require slightly stronger conditions by restricting u to $C^{1,2}_+(\overline{\Lambda})$. We shall leave the possible relaxation of these conditions in future research.

EXAMPLE 3.6. The first example below is in the framework of the path-dependent case of Section 2.3. The second covers Dupire's case.

(i) If $u(t,\omega) = g(\omega_T) + \int_t^T f(s,\omega_s)\,ds$ and f, g are smooth, then

$$\partial_t u(t,\omega) = -f(t,\omega_t),$$

$$\langle \partial_\omega u(t,\omega), \eta\rangle = \partial_x g(\omega_T)\cdot \eta_T + \int_t^T \partial_x f(s,\omega_s)\cdot \eta_s\,ds,$$

$$\langle \partial^2_{\omega\omega} u(t,\omega), (\eta^1,\eta^2)\rangle = \partial^2_{xx} g(\omega_T) : [\eta^1_T(\eta^2_T)^\top] + \int_t^T \partial^2_{xx} f(s,\omega_s) : [\eta^1_s(\eta^2_s)^\top]\,ds.$$

Here, $A_1 : A_2 := \mathrm{tr}(A_1 A_2^\top)$ for two matrices A_1, A_2.

(ii) If u is adapted in the sense that, after time t, the path of ω is frozen, that is, $u(t,\omega) = v(t, \omega \mathbf{1}_{[0,t)} + \omega_t \mathbf{1}_{[t,T]})$ for some function v, then

$$\partial_t u(t,\omega) = \partial_t v(t,\omega),$$
$$\langle \partial_\omega u(t,\omega), \eta \rangle = \partial_\omega v(t,\omega) \cdot \eta_t,$$
$$\langle \partial^2_{\omega\omega} u(t,\omega), (\eta^1, \eta^2) \rangle = \partial^2_{\omega\omega} v(t,\omega) : [\eta^1_t (\eta^2_t)^\top],$$

where $\partial_t v$, $\partial_\omega v$, $\partial^2_{\omega\omega} v$ are Dupire's path derivatives.

The following result is similar to Cont and Fournié [10].

PROPOSITION 3.7. *Let $u_1, u_2 \in C^{1,2}_+(\overline{\Lambda})$. Assume $u_1 = u_2$ on Λ, then $\partial_t u_1 = \partial_t u_2$, $\langle \partial_\omega u_1, \eta \rangle = \langle \partial_\omega u_2, \eta \rangle$, $\langle \partial^2_{\omega\omega} u_1, (\eta, \eta) \rangle = \langle \partial^2_{\omega\omega} u_2, (\eta, \eta) \rangle$ on Λ, for all $\eta \in \Omega$.*

The proof is closely related to the functional Itô formula below, so we postpone it and combine with the proof of Theorem 3.10. We believe it is possible to show that $\langle \partial^2_{\omega\omega} u_1, (\eta_1, \eta_2) \rangle = \langle \partial^2_{\omega\omega} u_2, (\eta_1, \eta_2) \rangle$ for all $\eta_1, \eta_2 \in \Omega$ under possibly weaker regularity conditions on u_1, u_2. We do not pursue such generality in this paper. We now define the following.

DEFINITION 3.8. Let $C^{1,2}_+(\Lambda)$ denote the collection of functions $u : \Lambda \to \mathbb{R}$ such that there exists $\tilde{u} \in C^{1,2}_+(\overline{\Lambda})$ satisfying $\tilde{u} = u$ on Λ. In this case, we define the path derivatives: $\partial_t u := \partial_t \tilde{u}$, $\partial_\omega u := \partial_\omega \tilde{u}$, $\partial^2_{\omega\omega} u := \partial^2_{\omega\omega} \tilde{u}$ on Λ.

By Proposition 3.7, for any $\eta \in \Omega$, clearly $\partial_t u$, $\partial_\omega u$, and $\langle \partial^2_{\omega\omega} u, (\eta, \eta) \rangle$ are uniquely determined on Λ, regardless of the choice of $\tilde{u}$ in Definition 3.8.

3.2. *Functional Itô formula in the regular case.* As noted in the Introduction, the use of fBm causes two difficulties: (i) it is non-Markovian, because of the two-variable kernel $K(t,r)$; (ii) for $H < 1/2$, the kernel is singular, that is, $\lim_{r\to t} K(t,r) = \infty$. The SDE (3.1) has the same issues, stemming from the same properties of b and σ as functions of (t,r). The possible dependence of b and σ on the path X (which we mostly omit in the notation below) add to the path dependence. To understand the problem better, in this subsection we focus on the lack of a Markov property and additional path dependence, and postpone the singularity issue to the next subsection; thus we assume b and σ have no singularity as r tends to t. We remark that in this "regular" case, X is typically a semimartingale:

$$dX_t = b(t;t)\,dt + \sigma(t;t)\,dW_t + \left[\int_0^t \partial_t b(t;r)\,dr + \int_0^t \partial_t \sigma(t;r)\,dW_r\right] dt, \tag{3.13}$$

provided that $\partial_t b(t;r)$, $\partial_t \sigma(t;r)$ exist and have good integrability properties near the time diagonal. We shall remark that, even for $H > \frac{1}{2}$, the fBM B^H does not satisfy the above expression because the corresponding $\partial_t \sigma(t;r)$ is not square integrable.

In this subsection, we assume the following.

ASSUMPTION 3.9. $\partial_t b(t;r,\cdot)$, $\partial_t \sigma(t;r,\cdot)$ exist for $t \in [r,T]$, and for $\varphi = b, \sigma, \partial_t b, \partial_t \sigma$,

$$|\varphi(t;r,\omega)| \le C_0\big[1 + \|\omega\|_T^{\kappa_0}\big] \quad \text{for some constants } C_0, \kappa_0 > 0. \tag{3.14}$$

Recall (3.3), (3.4) and (3.13). Under Assumptions 3.1 and 3.9, it is obvious that

$$\mathbb{E}\big[\|\check{X}^t\|_T^p\big] \le C_p \quad \text{for all } p \ge 1, 0 \le t \le T. \tag{3.15}$$

Our main result is the following functional Itô formula.

THEOREM 3.10. *Let Assumptions 3.1 and 3.9 hold and $u \in C_+^{1,2}(\Lambda)$. Then $\mathbb{P}$-a.s.,*

$$\begin{aligned} du(t,\check{X}^t) &= \partial_t u(t,\check{X}^t)\,dt + \frac{1}{2}\langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X},\sigma^{t,X})\rangle\,dt \\ &\quad + \langle \partial_\omega u(t,\check{X}^t), b^{t,X}\rangle\,dt + \langle \partial_\omega u(t,\check{X}^t), \sigma^{t,X}\rangle\,dW_t, \end{aligned} \tag{3.16}$$

where, for $\varphi = b, \sigma$, $\varphi_s^{t,\omega} := \varphi(s;t,\omega)$ emphasizes the dependence on $s \in [t,T]$.

The main idea of the proof follows that of Dupire [16] and Cont and Fournie [11]. However, here we have to deal with the two time variables, and for that purpose we need a few technical lemmas. The first one is a direct consequence of the proof of Kolmogorov's continuity criterion; see, for example, Revuz and Yor [38], Chapter I, Theorem 2.1.

LEMMA 3.11. *Let $\tilde{X}$ be a process on $[0,T]$ with $\tilde{X}_0 = 0$, and $\alpha, \beta > 0$ be constants. Assume*

$$\mathbb{E}\big[|\tilde{X}_t - \tilde{X}_{t'}|^{2p}\big] \le C_p \beta^p |t - t'|^{\alpha p} \quad \textit{for all } 0 \le t < t' \le T,\ p \ge 1.$$

Then, for each $p \ge 1$, there exists another constant $\tilde{C}_p > 0$, which may depend on T, α, p, the dimension d and the above C_p, but does not depend on β, such that

$$\mathbb{E}\big[\|\tilde{X}\|_T^{2p}\big] \le \tilde{C}_p \beta^p.$$

The following result will be crucial for the functional Itô formula.

LEMMA 3.12. *Let Assumptions 3.1 and 3.9 hold. Fix n and set $h := 2^{-n}T$, $t_i := ih$, $i = 0, \dots, 2^n$,*

$$X_t^n := \sum_{i=0}^{2^n-1} \Theta_t^{t_i} \mathbf{1}_{[t_i,t_{i+1})}(t) + X_T \mathbf{1}_{\{T\}}(t). \tag{3.17}$$

Then, for any $p \geq 1$,

$$\mathbb{E}[\|\check{X}^t - \check{X}^{t'}\|_T^{4p}] \leq C_p|t' - t|^p, \qquad \mathbb{E}[\|X - X^n\|_T^8] \leq C2^{-n}. \tag{3.18}$$

PROOF. We start with the first inequality of (3.18). Let $t < t'$ and denote

$$\tilde{X}_s := \check{X}_s^{t'} - \check{X}_s^t = \int_{s\wedge t}^{s\wedge t'} b(s;r)\,dr + \int_{s\wedge t}^{s\wedge t'} \sigma(s;r)\,dW_r.$$

We claim that, for any $s < s'$ and any $p \geq 1$,

$$I_p := \mathbb{E}[|\tilde{X}_s - \tilde{X}_{s'}|^{2p}] \leq C_p(t' - t)^{\frac{p}{2}}(s' - s)^{\frac{p}{2}}. \tag{3.19}$$

Then the first inequality of (3.18) follows from Lemma 3.11. We prove (3.19) in three cases.

Case 1. $s \leq t$. Then $\tilde{X}_s = 0$ and thus, by Assumptions 3.1 and 3.9,

$$\begin{aligned}
I_p = \mathbb{E}[|\tilde{X}_{s'}|^{2p}] &\leq C_p\mathbb{E}\left[\left|\int_{s'\wedge t}^{s'\wedge t'} b(s';r)\,dr\right|^{2p} + \left|\int_{s'\wedge t}^{s'\wedge t'} |\sigma(s';r)|^2\,dr\right|^p\right] \\
&\leq C_p\mathbb{E}\left[\left|\int_{s'\wedge t}^{s'\wedge t'} [1 + \|X\|_T^{\kappa_0}]\,dr\right|^{2p} + \left|\int_{s'\wedge t}^{s'\wedge t'} [1 + \|X\|_T^{2\kappa_0}]\,dr\right|^p\right] \\
&\leq C_p[s' \wedge t' - s' \wedge t]^p \leq C_p(t' - t)^{\frac{p}{2}}(s' - s)^{\frac{p}{2}}.
\end{aligned}$$

Case 2. $t < s \leq t'$. Then

$$\begin{aligned}
I_p &= \mathbb{E}\left[\left|\int_t^s [b(s;r)\,dr + \sigma(s;r)\,dW_r] - \int_t^{s'\wedge t'} [b(s';r)\,dr + \sigma(s';r)\,dW_r]\right|^{2p}\right] \\
&\leq C_p\mathbb{E}\left[\left|\int_t^s [b(s;r) - b(s';r)]\,dr + \int_t^s [\sigma(s;r) - \sigma(s';r)]\,dW_r\right|^{2p}\right. \\
&\qquad \left. + \left|\int_s^{s'\wedge t'} b(s';r)\,dr + \int_s^{s'\wedge t'} \sigma(s';r)\,dW_r\right|^{2p}\right] \\
&\leq C_p\mathbb{E}\left[\left|\int_t^s [s' - s][1 + \|X\|_T^{\kappa_0}]\,dr\right|^{2p} + \left|\int_t^s [s' - s]^2[1 + \|X\|_T^{2\kappa_0}]\,dr\right|^p\right. \\
&\qquad \left. + \left|\int_s^{s'\wedge t'} [1 + \|X\|_T^{\kappa_0}]\,dr\right|^{2p} + \left|\int_s^{s'\wedge t'} [1 + \|X\|_T^{2\kappa_0}]\,dr\right|^p\right] \\
&\leq C_p[(s - t)^p(s' - s)^{2p} + (s' \wedge t' - s)^p] \leq C_p(t' - t)^{\frac{p}{2}}(s' - s)^{\frac{p}{2}}.
\end{aligned}$$

Case 3. $s > t'$. Then

$$\begin{aligned} I_p &= \mathbb{E}\Big[\Big|\int_t^{t'} [b(s;r)\,dr + \sigma(s;r)\,dW_r] - \int_t^{t'} [b(s';r)\,dr + \sigma(s';r)\,dW_r]\Big|^{2p}\Big] \\ &\le C_p \mathbb{E}\Big[\Big|\int_t^{t'} [s'-s][1+\|X\|_T^{\kappa_0}]\,dr\Big|^{2p} + \Big|\int_t^{t'} [s'-s]^2[1+\|X\|_T^{2\kappa_0}]\,dr\Big|^{p}\Big] \\ &\le C_p(t'-t)^p(s'-s)^{2p} \le C_p(t'-t)^{\frac{p}{2}}(s'-s)^{\frac{p}{2}}. \end{aligned}$$

So in all the cases, we have proved (3.19).

To see the second inequality in (3.18), for each i, note that

$$X_t - \Theta_t^{t_i} = \int_{t_i}^t b(t;r)\,dr + \int_{t_i}^t \sigma(t;r)\,dW_r, \quad t \ge t_i.$$

By Case 2 above, we see that

$$\mathbb{E}\big[\big|[X_t - \Theta_t^{t_i}] - [X_{t'} - \Theta_{t'}^{t_i}]\big|^{2p}\big] \le C_p h^{\frac{p}{2}} |t-t'|^{\frac{p}{2}}, \quad t_i \le t < t' \le t_{i+1}.$$

Then by Lemma 3.11, we have $\mathbb{E}[\sup_{t_i \le t \le t_{i+1}} |X_t - \Theta_t^{t_i}|^{4p}] \le C_p h^p$. Thus

$$\mathbb{E}[\|X - X^n\|_T^8] \le \sum_{i=0}^{2^n-1} \mathbb{E}\Big[\sup_{t_i \le t \le t_{i+1}} |X_t - X_t^n|^8\Big] \le Ch^2 2^n = C2^{-n},$$

completing the proof. □

We need another lemma dealing with two variable functions/processes.

LEMMA 3.13. *Let Assumptions 3.1 and 3.9 hold, $u \in C^0(\overline{\Lambda})$, and $0 \le t_1 < t_2 \le T$.*

(i) *If $\partial_\omega u$ is continuous and has polynomial growth, then*

$$(3.20)\quad \begin{aligned} \Big\langle \partial_\omega u(t_2, \check{X}^{t_1}), \int_{t_1}^{t_2} b(\cdot;r)\,dr \Big\rangle &= \int_{t_1}^{t_2} \langle \partial_\omega u(t_2, \check{X}^{t_1}), b^{r,X}\rangle\,dr; \\ \Big\langle \partial_\omega u(t_2, \check{X}^{t_1}), \int_{t_1}^{t_2} \sigma(\cdot;r)\,dW_r \Big\rangle &= \int_{t_1}^{t_2} \langle \partial_\omega u(t_2, \check{X}^{t_1}), \sigma^{r,X}\rangle\,dW_r. \end{aligned}$$

Here, assuming W is k-dimensional, $\langle \partial_\omega u, \sigma\rangle\,dW_r := \sum_{i=1}^k \langle \partial_\omega u, \sigma^i\rangle\,dW_r^i$, where σ^i is the ith column of σ.

(ii) *If* $\partial^2_{\omega\omega} u$ *is continuous and has polynomial growth, then*

$$\Big\langle \partial^2_{\omega\omega} u(t_2, \check{X}^{t_1}), \Big(\int_{t_1}^{t_2} \sigma(\cdot; r)\, dW_r, \int_{t_1}^{t_2} \sigma(\cdot; r)\, dW_r\Big)\Big\rangle$$

$$\begin{aligned} &= \sum_{i=1}^{k} \int_{t_1}^{t_2} \langle \partial^2_{\omega\omega} u(t_2, \check{X}^{t_1}), (\sigma^{i,t,X}, \sigma^{i,t,X})\rangle\, dt \\ &\quad + \int_{t_1}^{t_2} \Big\langle \partial^2_{\omega\omega} u(t_2, \check{X}^{t_1}), \Big(\int_{t_1}^{t} \sigma(\cdot; r)\, dW_r, \sigma^{t,X}\Big)\Big\rangle dW_t \\ &\quad + \int_{t_1}^{t_2} \Big\langle \partial^2_{\omega\omega} u(t_2, \check{X}^{t_1}), \Big(\sigma^{t,X}, \int_{t_1}^{t} \sigma(\cdot; r)\, dW_r\Big)\Big\rangle dW_t. \end{aligned} \tag{3.21}$$

PROOF. For notational simplicity, we assume $d = k = 1$, and omit the variable $(t_2, \check{X}^{t_1})$ inside $\partial_\omega u$ and $\partial^2_{\omega\omega} u$.

(i) We prove the second equality in three steps. The first one follows similar arguments.

Step 1. Assume $\sigma(s; r) = \sum_{i=0}^{n-1} \sigma(s; r_i)\mathbf{1}_{[r_i, r_{i+1})}$ for some $t_1 = r_0 < \cdots < r_n = t_2$. Since $\partial_\omega u$ is linear, the second equality of (3.20) is obvious.

Step 2. Assume, for some constants $C, \kappa > 0$ and for all $t_1 \le r < r' \le t_2$,

$$\big|\sigma(s; r) - \sigma(s; r')\big| + \big|\partial_s \sigma(s; r) - \partial_s \sigma(s; r')\big| \le C\big[1 + \|\omega\|_T^\kappa\big]\big|r - r'\big|. \tag{3.22}$$

Denote $\sigma_n(s; r) := \sum_{i=0}^{2^n-1} \sigma(s; r_i)\mathbf{1}_{[r_i, r_{i+1})}$, where $r_i := t_1 + i(t_2 - t_1)2^{-n}$, $i = 0, \dots, 2^{-n}$. Then $\sup_{t_2 \le s \le T}[|\sigma_n(s; r) - \sigma(s; r)| + |\partial_s \sigma_n(s; r) - \partial_s \sigma(s; r)| \le C[1 + \|\omega\|_T^\kappa]2^{-n}$ for all $r \in [t_1, t_2]$. By (3.9), this implies $\lim_{n\to\infty} \langle \partial_\omega u, \sigma_n^{r,X}\rangle = \langle \partial_\omega u, \sigma^{r,X}\rangle$, which together with the dominated convergence theorem, implies further that

$$\lim_{n\to\infty} \mathbb{E}\Big[\Big|\int_{t_1}^{t_2} \langle \partial_\omega u, \sigma_n^{r,X}\rangle\, dW_r - \int_{t_1}^{t_2} \langle \partial_\omega u, \sigma^{r,X}\rangle\, dW_r\Big|^2\Big] = 0. \tag{3.23}$$

Moreover, denote $\tilde{X}_s := \int_{t_1}^{t_2} [\sigma_n(s; r) - \sigma(s; r)]\, dW_r$. Then, for $t_2 \le s < s' \le T$ and $p \ge 1$,

$$\begin{aligned} \mathbb{E}[|\tilde{X}_s - \tilde{X}_{s'}|^{2p}] &= \mathbb{E}\Big[\Big|\int_{t_1}^{t_2} \int_s^{s'} \partial_s \sigma_n(l; r) - \partial_s \sigma(l; r)\, dl\, dW_r\Big|^{2p}\Big] \\ &\le C_p \mathbb{E}\Big[\Big|\int_{t_1}^{t_2} \Big|\int_s^{s'} \partial_s \sigma_n(l; r) - \partial_s \sigma(l; r)\, dl\Big|^2 dr\Big|^p\Big] \\ &\le C_p 2^{-2pn}(s' - s)^{2p}. \end{aligned}$$

Applying Lemma 3.11, we get $\mathbb{E}[\sup_{t_2 \le s \le T} |\tilde{X}_s|^2] \le C2^{-2n}$. Then

$$\lim_{n\to\infty} \sup_{t_2 \le s \le T} |\tilde{X}_s| = 0, \quad \mathbb{P}\text{-a.s.},$$

and thus

$$\lim_{n\to\infty}\Big\langle \partial_\omega u, \int_{t_1}^{t_2} \sigma_n(\cdot;r)\,dW_r\Big\rangle = \Big\langle \partial_\omega u, \int_{t_1}^{t_2} \sigma(\cdot;r)\,dW_r\Big\rangle, \quad \mathbb{P}\text{-a.s.} \tag{3.24}$$

By Step 1, the second equality of (3.20) holds for σ_n. Then by (3.23) and (3.24) we obtain the desired equality for σ.

Step 3. Denote

$$\sigma_\varepsilon(s;r) := \frac{1}{\varepsilon}\int_{(r-\varepsilon)^+}^{r} \sigma(s;l)\,dl \quad \text{and thus} \quad \partial_s\sigma_\varepsilon(s;r) := \frac{1}{\varepsilon}\int_{(r-\varepsilon)^+}^{r} \partial_s\sigma(s;l)\,dl.$$

It is clear that $\lim_{\varepsilon\to 0}\mathbb{E}[\int_{t_1}^{t_2}[|\sigma_\varepsilon(s;r)-\sigma(s;r)|^p + |\partial_s\sigma_\varepsilon(s;r)-\partial_s\sigma(s;r)|^p]\,dr] = 0$ for all $p \geq 1$. Fix some p large enough, then there exists $\varepsilon_n \downarrow 0$ such that

$$\mathbb{E}\bigg[\int_{t_1}^{t_2}\big[\big|\sigma_{\varepsilon_n}(s;r)-\sigma(s;r)\big|^p + \big|\partial_s\sigma_{\varepsilon_n}(s;r)-\partial_s\sigma(s;r)\big|^p\big]\,dr\bigg] \leq 2^{-n}.$$

Now following the arguments of Step 2 as well as that of Lemma 3.11, one can show that

$$\lim_{n\to\infty}\mathbb{E}\bigg[\bigg|\int_{t_1}^{t_2}\langle\partial_\omega u, \sigma_{\varepsilon_n}^{r,X}\rangle\,dW_r - \int_{t_1}^{t_2}\langle\partial_\omega u, \sigma^{r,X}\rangle\,dW_r\bigg|^2\bigg] = 0;$$

$$\lim_{n\to\infty}\Big\langle \partial_\omega u, \int_{t_1}^{t_2} \sigma_{\varepsilon_n}(\cdot;r)\,dW_r\Big\rangle = \Big\langle \partial_\omega u, \int_{t_1}^{t_2} \sigma(\cdot;r)\,dW_r\Big\rangle \quad \text{a.s.}$$

Clearly, σ_ε satisfies the conditions in Step 2, and thus the second equality of (3.20) holds for each σ_ε. Then the above limits imply the desired equality for σ.

(ii) Combining the arguments in Steps 2 and 3 in (i) above, it suffices to prove (3.21) in the case σ is piecewise constant in r: $\sigma(s;r) = \sum_{i=0}^{n-1}\sigma(s;r_i)\mathbf{1}_{[r_i,r_{i+1})}$ for some $t_1 = r_0 < \cdots < r_n = t_2$. In this case, $\int_{t_1}^{t_2}\sigma(s;r)\,dW_r = \sum_{i=0}^{n-1}[W_{r_{i+1}} - W_{r_i}]\sigma(s;r_i)$. Denote $W_{s,t} := W_t - W_s$ and $I_{ij} := \langle\partial^2_{\omega\omega}u, (\sigma^{r_i,X}, \sigma^{r_j,X})\rangle$. Since $\partial^2_{\omega\omega}u$ is bilinear, we see that

$$\begin{aligned}
&\int_{t_1}^{t_2}\Big\langle\partial^2_{\omega\omega}u, \Big(\int_{t_1}^{t}\sigma(\cdot;r)\,dW_r, \sigma^{t,X}\Big)\Big\rangle\,dW_t \\
&\quad = \sum_{i=0}^{n-1}\int_{r_i}^{r_{i+1}}\Big\langle\partial^2_{\omega\omega}u, \Big(\sum_{j=0}^{i-1}\sigma^{r_j,X}W_{r_j,r_{j+1}} + \sigma^{r_i,X}W_{r_i,t}, \sigma^{r_i,X}\Big)\Big\rangle\,dW_t \\
&\quad = \sum_{i=0}^{n-1}\bigg[\sum_{j=0}^{i-1} I_{ji}W_{r_i,r_{i+1}}W_{r_j,r_{j+1}} + I_{ii}\int_{r_i}^{r_{i+1}} W_{r_i,t}\,dW_t\bigg].
\end{aligned}$$

Then, by similar arguments for the last term of (3.21), the right-hand side of (3.21) becomes

$$\begin{aligned}
&\sum_{i=0}^{n-1} I_{ii}[r_{i+1}-r_i] + \sum_{i=0}^{n-1}\Bigg[\sum_{j=0}^{i-1}[I_{ji}+I_{ij}]W_{r_i,r_{i+1}}W_{r_j,r_{j+1}} \\
&\qquad + 2I_{ii}\int_{r_i}^{r_{i+1}} W_{r_i,t}\,dW_t\Bigg] \\
&\quad = \sum_{i=0}^{n-1} I_{ii}W^2_{r_i,r_{i+1}} + \sum_{0\le j<i\le n-1} W_{r_i,r_{i+1}}W_{r_j,r_{j+1}}[I_{ij}+I_{ji}] \\
&\quad = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} W_{r_i,r_{i+1}}W_{r_j,r_{j+1}}I_{ij},
\end{aligned}$$

which is equal to the left-hand side of (3.21). □

We are now ready to prove our main result.

PROOF OF THEOREM 3.10 AND PROPOSITION 3.7. As announced, we shall prove these two results together. This does not create a circular argument. The first step is to prove Theorem 3.10 on $\overline{\Lambda}$, which is not related to Proposition 3.7, then we prove Proposition 3.7, and finally we invoke Proposition 3.7 to draw the conclusion of Theorem 3.10 as it applies to Λ instead of $\overline{\Lambda}$, where the derivatives in (1.2) are uniquely determined because of Proposition 3.7. For notational simplicity in this proof, we assume all processes are scalar. The multidimensional case can be proved without any significant difficulty. Moreover, we emphasize again that we denote by κ the generic polynomial growth order which may vary from line to line.

Step 1. By abusing the notation slightly, in this step we assume $u \in C^{1,2}_+(\overline{\Lambda})$ and prove (3.16) for such a function. This step does not refer to Proposition 3.7 since it works in $C^{1,2}_+(\overline{\Lambda})$. Without loss of generality, we shall only prove the result for $u(T,X) - u(0,0)$. Fix n and consider the setting in (3.17). Then

$$u(T,X) - u(0,0) = u(T,X) - u\big(T,X^n\big) + \sum_{i=0}^{2^n-1}\big[I_i^1 + I_i^2\big], \tag{3.25}$$

where

$$\begin{aligned}
I_i^1 &:= u\big(t_{i+1}, X^n \otimes_{t_i} \Theta^{t_i}\big) - u\big(t_i, X^n \otimes_{t_i} \Theta^{t_i}\big); \\
I_i^2 &:= u\big(t_{i+1}, X^n \otimes_{t_{i+1}} \Theta^{t_{i+1}}\big) - u\big(t_{i+1}, X^n \otimes_{t_i} \Theta^{t_i}\big).
\end{aligned}$$

First, by the second inequality in (3.18) we have

$$\mathbb{E}\Big[\sum_{n=1}^{\infty}\|X-X^n\|_T^8\Big] \le C\sum_{n=1}^{\infty} 2^{-n} < \infty \quad \text{and thus} \quad \lim_{n\to\infty}\|X-X^n\|_T = 0, \quad \mathbb{P}\text{-a.s.}$$

Since u is continuous, we have

$$\lim_{n\to\infty}\big[u(T,X)-u\big(T,X^n\big)\big]=0, \quad \mathbb{P}\text{-a.s.} \tag{3.26}$$

Next, by definition (3.5) and the continuity of $\partial_t u$ we have

$$I_i^1 = \int_{t_i}^{t_{i+1}} \partial_t u\big(t, X^n \otimes_{t_i} \Theta^{t_i}\big)\,dt.$$

By (3.18), one can easily show that

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\int_{t_i}^{t_{i+1}} \big\|X^n \otimes_{t_i} \Theta^{t_i} - X\otimes_t \Theta^t\big\|_T\,dt = 0, \quad \mathbb{P}\text{-a.s.} \tag{3.27}$$

and thus, again by the continuity of $\partial_t u$ together with polynomial growth of $\partial_t u$ and (3.15),

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1} I_i^1 = \int_0^T \partial_t u\big(t, X\otimes_t \Theta^t\big)\,dt, \quad \mathbb{P}\text{-a.s.} \tag{3.28}$$

Moreover, note that

$$X^n \otimes_{t_i} \Theta^{t_i} = X^n \otimes_{t_{i+1}} \Theta^{t_i} =: X^{n,i}. \tag{3.29}$$

Denote $\Delta\Theta^{t_i} := \Theta^{t_{i+1}} - \Theta^{t_i}$, then

$$\begin{aligned} I_i^2 &= u\big(t_{i+1}, X^n \otimes_{t_{i+1}} \Theta^{t_{i+1}}\big) - u\big(t_{i+1}, X^{n,i}\big) \\ &= \int_0^1 \big\langle \partial_\omega u\big(t_{i+1}, X^n \otimes_{t_{i+1}} \big[\Theta^{t_i} + \lambda\Delta\Theta^{t_i}\big]\big), \Delta\Theta^{t_i}\big\rangle\,d\lambda \\ &= I_i^{2,1} + I_i^{2,2} + I_i^{2,3}, \end{aligned} \tag{3.30}$$

where

$$\begin{aligned} I_i^{2,1} &:= \big\langle \partial_\omega u\big(t_{i+1}, X^{n,i}\big), \Delta\Theta^{t_i}\big\rangle \\ I_i^{2,2} &:= \frac{1}{2}\big\langle \partial^2_{\omega\omega} u\big(t_{i+1}, X^{n,i}\big), \big(\Delta\Theta^{t_i}, \Delta\Theta^{t_i}\big)\big\rangle \\ I_i^{2,3} &:= \big\langle \partial^2_{\omega\omega} u\big(t_{i+1}, X^n \otimes_{t_{i+1}} \big[\Theta^{t_i} + \lambda^*\Delta\Theta^{t_i}\big]\big) - \partial^2_{\omega\omega} u\big(t_{i+1}, X^{n,i}\big), \\ &\qquad \big(\Delta\Theta^{t_i}, \Delta\Theta^{t_i}\big)\big\rangle, \end{aligned}$$

for some appropriate (random) λ^* taking values on [0, 1]. Note that

$$\Delta\Theta_s^{t_i} = \int_{t_i}^{t_{i+1}} b(s;r)\,dr + \int_{t_i}^{t_{i+1}} \sigma(s;r)\,dW_r, \quad s \ge t_{i+1}. \tag{3.31}$$

Since $X^n \otimes_{t_{i+1}} \Theta^{t_i} = X^n \otimes_{t_i} \Theta^{t_i}$ is $\mathcal{F}_{t_i}$-measurable, similar to Lemma 3.13(i), we have

$$I_i^{2,1} = \int_{t_i}^{t_{i+1}} \langle \partial_\omega u(t_{i+1}, X^{n,i}), b^{r,X} \rangle \, dr + \int_{t_i}^{t_{i+1}} \langle \partial_\omega u(t_{i+1}, X^{n,i}), \sigma^{r,X} \rangle \, dW_r.$$

Recall (3.29) and (3.27). By Assumptions 3.1 and 3.9, since $\partial_\omega u$ is continuous, we have

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} \int_{t_i}^{t_{i+1}} \big| \langle \partial_\omega u(t_{i+1}, X^{n,i}), b^{r,X} \rangle - \langle \partial_\omega u(r, \check{X}^r), b^{r,X} \rangle \big| \, dr = 0;$$

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} \int_{t_i}^{t_{i+1}} \big| \langle \partial_\omega u(t_{i+1}, X^{n,i}), \sigma^{r,X} \rangle - \langle \partial_\omega u(r, \check{X}^r), \sigma^{r,X} \rangle \big|^2 \, dr = 0.$$

Therefore, with convergence in $\mathbb{L}^2$,

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} I_i^{2,1} = \int_0^T \langle \partial_\omega u(t, \check{X}^t), b^{t,X} \rangle \, dt + \int_0^T \langle \partial_\omega u(t, \check{X}^t), \sigma^{t,X} \rangle \, dW_t. \tag{3.32}$$

We now consider $I_i^{2,2}$. In the spirit of Lemma 3.13(ii) we can prove

$$2I_i^{2,2} = I_i^{2,2,1} + I_i^{2,2,2} + I_i^{2,2,3}, \tag{3.33}$$

where

$$\begin{aligned}
I_i^{2,2,1} :=& \Big\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \Big(\int_{t_i}^{t_{i+1}} b(\cdot; r)\, dr, \int_{t_i}^{t_{i+1}} b(\cdot; r)\, dr \Big) \Big\rangle \\
&+ \Big\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \Big(\int_{t_i}^{t_{i+1}} b(\cdot; r)\, dr, \int_{t_i}^{t_{i+1}} \sigma(\cdot; r)\, dW_r \Big) \Big\rangle \\
&+ \Big\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \Big(\int_{t_i}^{t_{i+1}} \sigma(\cdot; r)\, dW_r, \int_{t_i}^{t_{i+1}} b(\cdot; r)\, dr \Big) \Big\rangle \\
I_i^{2,2,2} :=& \int_{t_i}^{t_{i+1}} \Big\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \Big(\int_{t_i}^{t} \sigma(\cdot; r)\, dW_r, \sigma^{t,X} \Big) \Big\rangle dW_t \\
&+ \int_{t_i}^{t_{i+1}} \Big\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \Big(\sigma^{t,X}, \int_{t_i}^{t} \sigma(\cdot; r)\, dW_r \Big) \Big\rangle dW_t \\
I_i^{2,2,3} :=& \int_{t_i}^{t_{i+1}} \langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), (\sigma^{t,X}, \sigma^{t,X}) \rangle \, dt.
\end{aligned}$$

One can similarly show that, with convergence in $\mathbb{L}^2$,

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} I_i^{2,2,3} = \int_0^T \langle \partial^2_{\omega\omega} u(t, \check{X}^t), (\sigma^{t,X}, \sigma^{t,X}) \rangle \, dt. \tag{3.34}$$

By the martingale property,

$$
\begin{aligned}
\mathbb{E}\Bigg[\Bigg|\sum_{i=0}^{2^n-1} I_i^{2,2,2}\Bigg|^2\Bigg]
&= \sum_{i=0}^{2^n-1} \mathbb{E}\Bigg[\int_{t_i}^{t_{i+1}} \bigg|\bigg\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \bigg(\int_{t_i}^{t} \sigma(\cdot; r)\, dW_r, \sigma^{t,X}\bigg)\bigg\rangle\bigg|^2 dt \\
&\quad + \int_{t_i}^{t_{i+1}} \bigg|\bigg\langle \partial^2_{\omega\omega} u(t_{i+1}, X^{n,i}), \bigg(\sigma^{t,X}, \int_{t_i}^{t} \sigma(\cdot; r)\, dW_r\bigg)\bigg\rangle\bigg|^2 dt\Bigg] \\
&\le C \sum_{i=0}^{2^n-1} \mathbb{E}\Bigg[\big[1 + \|X^{n,i}\|_T^{\kappa} + \|X\|_T^{\kappa}\big] \int_{t_i}^{t_{i+1}} \sup_{t_{i+1} \le s \le T} \bigg|\int_{t_i}^{t} \sigma(s; r)\, dW_r\bigg|^2 dt\Bigg] \\
&\le C \sum_{i=0}^{2^n-1} \Bigg(\mathbb{E}\Bigg[2^{-n} \int_{t_i}^{t_{i+1}} \sup_{t_{i+1} \le s \le T} \bigg|\int_{t_i}^{t} \sigma(s; r)\, dW_r\bigg|^4 dt\Bigg]\Bigg)^{\frac{1}{2}}.
\end{aligned}
$$

Note that, for $t_i \le t \le t_{i+1} \le s < s' \le T$ and $p \ge 1$,

$$
\begin{aligned}
&\mathbb{E}\Bigg[\bigg|\int_{t_i}^{t} \sigma(s; r)\, dW_r - \int_{t_i}^{t} \sigma(s'; r)\, dW_r\bigg|^{2p}\Bigg] \\
&\quad \le C_p \mathbb{E}\Bigg[\bigg[\int_{t_i}^{t} \big|\sigma(s; r) - \sigma(s'; r)\big|^2 dr\bigg]^p\Bigg] \\
&\quad \le C_p \mathbb{E}\Bigg[\bigg[\int_{t_i}^{t} \bigg[\int_{s}^{s'} \big|\partial_l \sigma(l, r)\big|\, dl\bigg]^2 dr\bigg]^p\Bigg] \le C_p (t - t_i)^p (s' - s)^{2p}.
\end{aligned}
$$

Applying Lemma 3.11, we have

$$
\mathbb{E}\Bigg[\sup_{t_{i+1} \le s \le T} \bigg|\int_{t_i}^{t} \sigma(s; r)\, dW_r\bigg|^4 dt\Bigg] \le C(t - t_i)^2 \le C 2^{-2n}. \tag{3.35}
$$

Then

$$
\begin{aligned}
\mathbb{E}\Bigg[\Bigg|\sum_{i=0}^{2^n-1} I_i^{2,2,2}\Bigg|^2\Bigg] &\le C \sum_{i=0}^{2^n-1} \bigg(2^{-n} \int_{t_i}^{t_{i+1}} 2^{-2n}\, dt\bigg)^{\frac{1}{2}} \\
&= C \sum_{i=0}^{2^n-1} 2^{-2n} = C 2^{-n} \to 0.
\end{aligned} \tag{3.36}
$$

Moreover, by (3.35) again,

$$\begin{aligned}
\mathbb{E}\Bigg[\bigg|\sum_{i=0}^{2^n-1} I_i^{2,2,1}\bigg|^2\Bigg] &\le 2^n \sum_{i=0}^{2^n-1} \mathbb{E}\big[I_i^{2,2,1}|^2\big] \\
&\le C2^n \sum_{i=0}^{2^n-1} \mathbb{E}\Big[\big[1 + \|X^{n,i}\|_T^\kappa + \|X\|_T^\kappa\big] \\
&\quad \times \Big[2^{-4n} + 2^{-2n} \sup_{t_{i+1}\le s\le T} \Big|\int_{t_i}^{t_{i+1}} \sigma(s;r)\,dW_r\Big|^2\Big]\Big] \\
&\le C2^{-2n} + C2^{-n} \sum_{i=0}^{2^n-1} \Big(\mathbb{E}\Big[\sup_{t_{i+1}\le s\le T} \Big|\int_{t_i}^{t_{i+1}} \sigma(s;r)\,dW_r\Big|^4\Big]\Big)^{\frac{1}{2}} \\
&\le C2^{-2n} + C2^{-n} \sum_{i=0}^{2^n-1} 2^{-n} \le C2^{-n}.
\end{aligned}$$

Plug this and (3.34), (3.36) into (3.33), we obtain

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} I_i^{2,2} = \frac{1}{2}\int_0^T \langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X}, \sigma^{t,X})\rangle\,dt. \tag{3.37}$$

Finally, denote $\|\Delta\Theta_i\| := \sup_{t_{i+1}\le s\le T} |\Delta\Theta_s^{t_i}|$. For any $\varepsilon > 0$, by (3.12) we have

$$\big|I_i^{2,3}\big| \le \rho\big(\|\Delta\Theta_i\|\big)\|\Delta\Theta_i\|^2 \le \rho(\varepsilon)\|\Delta\Theta_i\|^2 + C\varepsilon^{-1}\|\Delta\Theta_i\|^3.$$

By (3.31), similar to (3.35) we have

$$\mathbb{E}\big[\big|I_i^{2,3}\big|\big] \le \rho(\varepsilon)\mathbb{E}\big[\|\Delta\Theta_i\|^2\big] + C\varepsilon^{-1}\mathbb{E}\big[\|\Delta\Theta_i\|^3\big] \le C\rho(\varepsilon)2^{-n} + C\varepsilon^{-1}2^{-\frac{3}{2}n}.$$

Then

$$\mathbb{E}\Bigg[\sum_{i=0}^{2^n-1}\big|I_i^{2,3}\big|\Bigg] \le C\rho(\varepsilon) + C\varepsilon^{-1}2^{-\frac{n}{2}}.$$

By first sending $n\to\infty$ and then $\varepsilon\to 0$, we obtain $\lim_{n\to\infty}\mathbb{E}[\sum_{i=0}^{2^n-1}|I_i^{2,3}|] = 0$. Plug this and (3.32), (3.37) into (3.30), we get

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=0}^{2^n-1} I_i^2 &= \int_0^T \langle \partial_\omega u(t,\check{X}^t), b^{t,X}\rangle\,dt + \int_0^T \langle \partial_\omega u(t,\check{X}^t), \sigma^{t,X}\rangle\,dW_t \\
&\quad + \frac{1}{2}\int_0^T \langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X},\sigma^{t,X})\rangle\,dt, \quad \mathbb{P}\text{-a.s.}
\end{aligned}$$

This, together with (3.26) and (3.28), proves (3.16) for $u \in C_+^{1,2}(\overline{\Lambda})$.

Step 2. We now prove Proposition 3.7. Set $u := u^1 - u^2 \in C^{1,2}_+(\overline{\Lambda})$ and thus $u = 0$ on Λ. By definition in (3.5), it is clear that $\partial_t u = 0$ on Λ. Fix $(t_0, \omega^0) \in \Lambda$ and $\eta \in \Omega_{t_0}$. Define

$$\begin{aligned} \tilde{X}_t &:= \omega^0_t \mathbf{1}_{[0,t_0)}(t) + \left[\omega^0_{t_0} + \int_{t_0}^t \eta_r \, dW_r\right]\mathbf{1}_{[t_0,T]}(t), \\ \tilde{\Theta}^t_s &:= \omega^0_s + \int_{t_0}^t \eta_r \, dW_r, \end{aligned} \qquad t_0 \le t \le s \le T.$$

Note that $\tilde{X} \otimes_t \tilde{\Theta}^t$ is continuous, thus $u(t, \tilde{X} \otimes_t \tilde{\Theta}^t) = 0$, $t_0 \le t \le T$. By Step 1, which does not require Proposition 3.7, the process $u(t, \tilde{X} \otimes_t \tilde{\Theta}^t)$ satisfies (3.16) on $[t_0, T]$, and thus

$$\langle \partial_\omega u(t, \tilde{X} \otimes_t \tilde{\Theta}^t), \eta \rangle = 0, \qquad \langle \partial^2_{\omega\omega} u(t, \tilde{X} \otimes_t \tilde{\Theta}^t), (\eta, \eta) \rangle = 0, \quad t_0 \le t \le T, \mathbb{P}\text{-a.s.}$$

In particular, noting that $\tilde{X} \otimes_{t_0} \tilde{\Theta}^{t_0} = \omega^0$, then for $t = t_0$ we have

$$\langle \partial_\omega u(t_0, \omega^0), \eta \rangle = 0, \qquad \langle \partial^2_{\omega\omega} u(t_0, \omega^0), (\eta, \eta) \rangle = 0, \quad \mathbb{P}\text{-a.s.}$$

Since η is arbitrary, we prove Proposition 3.7.

Step 3. Finally, it is clear that (3.16) holds for $u \in C^{1,2}_+(\Lambda)$. In particular, by Proposition 3.7 (or say Step 2 above), the path derivatives in the right-hand side of (3.16) do not depend on the choice of $\tilde{u} \in C^{1,2}_+(\overline{\Lambda})$. □

REMARK 3.14. An alternative way to define path derivatives is directly through (3.16) by positing that this functional Itô formula holds, and the derivatives will be uniquely defined in appropriate sense. This is the approach in [18, 19] for PPDEs in a semmartingale framework. In this way, we may avoid involving the càdlàg space $\overline{\Lambda}$.

3.3. *The singular case.* We now consider the case where $b(t;t)$ and $\sigma(t;t)$ may blow up. We shall assume the following growth condition which is modeled after the behavior of the kernel for Gaussian processes with self-similarity parameter $H \in (0, 1/2)$, and is therefore satisfied by fBm with H in that range (and is more true for fBM with larger hurst parameter by setting $H = \frac{1}{2}$ at below).

ASSUMPTION 3.15. For $\varphi = b, \sigma$, $\partial_t \varphi(t; s, \cdot)$ exists for $t \in (s, T]$, and there exists $0 < H < \frac{1}{2}$ such that, for any $0 \le s < t \le T$,

$$\begin{aligned} |\varphi(t; s, \omega)| &\le C_0[1 + \|\omega\|_T^{\kappa_0}](t-s)^{H-\frac{1}{2}}, \\ |\partial_t \varphi(t; s, \omega)| &\le C_0[1 + \|\omega\|_T^{\kappa_0}](t-s)^{H-\frac{3}{2}}. \end{aligned} \tag{3.38}$$

We remark that, in this case $b^{t,X}$, $\sigma^{t,X}$ are not in Ω_t and thus cannot serve as the test function in the right-hand side of (3.16). To overcome this difficulty, we assume the following conditions on u which roughly mean that $u(t,\omega)$ does not depend on $\{\omega_s\}_{t\le s\le t+\delta}$ for some small $\delta>0$, or depends very weakly on the paths in the sense that u's derivatives become increasingly smaller as one approaches the diagonal.

DEFINITION 3.16. We say $u\in C^{1,2}_+(\Lambda)$ vanishes diagonally with rate $\alpha\in(0,1)$, denoted as $u\in C^{1,2}_{+,\alpha}(\Lambda)$ if there exist an extension of u in $C^{1,2}_+(\overline{\Lambda})$, still denoted as u, a polynomial growth order κ, and a bounded modulus of continuity function ρ satisfying: for any $0\le t<T$, $0<\delta\le T-t$ and $\eta,\eta_1,\eta_2\in\Omega_t$ with the supports of η, η_1 and η_2 contained in $[t,t+\delta]$:

(i) for any $\omega\in\overline{\Omega}$ such that $\omega\mathbf{1}_{[t,T]}\in\Omega_t$,

$$\begin{aligned}&\big|\langle\partial_\omega u(t,\omega),\eta\rangle\big|\le C\big[1+\|\omega\|_T^\kappa\big]\|\eta\|_T\delta^\alpha,\\&\big|\langle\partial^2_{\omega\omega}u(t,\omega),(\eta_1,\eta_2)\rangle\big|\le C\big[1+\|\omega\|_T^\kappa\big]\big\||\eta_1||\eta_2|\big\|_T\delta^{2\alpha};\end{aligned}\tag{3.39}$$

(ii) for any other $\omega'\in\overline{\Omega}$ such that $\omega'\mathbf{1}_{[t,T]}\in\Omega_t$,

$$\begin{aligned}&\big|\langle\partial_\omega u(t,\omega)-\partial_\omega u(t,\omega'),\eta\rangle\big|\\&\quad\le\big[1+\|\omega\|_T^\kappa+\|\omega'\|_T^\kappa\big]\|\eta\|_T\rho\big(\|\omega-\omega'\|_T\big)\delta^\alpha,\\&\big|\langle\partial^2_{\omega\omega}u(t,\omega)-\partial^2_{\omega\omega}u(t,\omega'),(\eta_1,\eta_2)\rangle\big|\\&\quad\le\big[1+\|\omega\|_T^\kappa+\|\omega'\|_T^\kappa\big]\big\||\eta_1||\eta_2|\big\|_T\rho\big(\|\omega-\omega'\|_T\big)\delta^{2\alpha}.\end{aligned}\tag{3.40}$$

These conditions will allow us to truncate the coefficients b,σ near the diagonal, and control the error made by this truncation. For $\varphi=b,\sigma$ and $\delta>0$, we introduce the truncated functions:

$$\varphi^\delta(t;s,\omega):=\varphi\big(t\vee(s+\delta);s,\omega\big).$$

We also again use the notation $\varphi^{\delta,s,\omega}$ for the path $t\in[s,T]\mapsto\varphi^\delta(t;s,\omega)$. Another consequence of using these truncated coefficients is that the notion of time and path derivatives must be understood as limits when the truncation parameter δ tends to 0. Specifically, we prove the following functional Itô formula, where in particular, the said limits exist.

THEOREM 3.17. *Let Assumptions 3.1 and 3.15 hold. Assume $u\in C^{1,2}_{+,\alpha}(\Lambda)$ with $\beta:=\alpha+H-\frac12>0$. Then the functional Itô formula (3.16) still holds true, where*

$$\begin{aligned}&\big\langle\partial_\omega u(t,\omega),\varphi^{t,\omega}\big\rangle:=\lim_{\delta\downarrow0}\big\langle\partial_\omega u(t,\omega),\varphi^{\delta,t,\omega}\big\rangle,\quad\varphi=b,\sigma,\\&\big\langle\partial^2_{\omega\omega}u(t,\omega),\big(\sigma^{t,\omega},\sigma^{t,\omega}\big)\big\rangle:=\lim_{\delta\downarrow0}\big\langle\partial^2_{\omega\omega}u(t,\omega),\big(\sigma^{\delta,t,\omega},\sigma^{\delta,t,\omega}\big)\big\rangle.\end{aligned}\tag{3.41}$$

PROOF. We proceed in three steps.

Step 1. We first show that the limits in (3.41) exist. We shall only prove it for σ, and the result for b follow the same arguments. Denote

$$\psi_n(s) := \frac{s - t_{n+1}}{\delta_n - \delta_{n+1}} \mathbf{1}_{(t_{n+1}, t_n]}(s) + \frac{t_{n-1} - s}{\delta_{n-1} - \delta_n} \mathbf{1}_{(t_n, t_{n-1})}(s), \tag{3.42}$$

where $\delta_n := \frac{1}{2^n}$, $t_n := t + \delta_n$. Then ψ_n is continuous, with support (t_{n+1}, t_{n-1}), and $\psi_n + \psi_{n+1} = 1$ on $[t_{n+1}, t_n]$. Now for any $\delta' < \delta$, assume $\delta' \in [\delta_{n+1}, \delta_n)$ and $\delta \in [\delta_{m+1}, \delta_m)$ for some $m \le n$. Consider the following decomposition into continuous functions for the constant 1:

$$\mathbf{1}_{[t,T]} = [1 - \psi_m]\mathbf{1}_{[t_m,T]} + \sum_{k=m}^{n} \psi_k + [1 - \psi_n]\mathbf{1}_{[t,t_n]}. \tag{3.43}$$

Note that $\sigma_s^{\delta,t,\omega} = \sigma_s^{\delta',t,\omega}$ for $s \in [t_m, T]$. Then, for $s \ge t$,

$$\begin{aligned} \sigma_s^{\delta,t,\omega} - \sigma_s^{\delta',t,\omega} &= \sum_{k=m}^{n} [\sigma_s^{\delta,t,\omega} - \sigma_s^{\delta',t,\omega}]\psi_k(s) \\ &\quad + [\sigma_s^{\delta,t,\omega} - \sigma_s^{\delta',t,\omega}][1 - \psi_n]\mathbf{1}_{[t,t_n]}. \end{aligned} \tag{3.44}$$

Thus, by the first inequalities of (3.39) and (3.38),

$$\begin{aligned} &|\langle \partial_\omega u(t,\omega), \sigma^{\delta,t,\omega}\rangle - \langle \partial_\omega u(t,\omega), \sigma^{\delta',t,\omega}\rangle| \\ &\quad \le \sum_{k=m}^{n} |\langle \partial_\omega u(t,\omega), \psi_k[\sigma^{\delta,t,\omega} - \sigma^{\delta',t,\omega}]\rangle| \\ &\qquad + |\langle \partial_\omega u(t,\omega), [1 - \psi_n]\mathbf{1}_{[t,t_n]}[\sigma^{\delta,t,\omega} - \sigma^{\delta',t,\omega}]\rangle| \\ &\quad \le C[1 + \|\omega\|_T^\kappa]\Bigg[\sum_{k=m}^{n} \sup_{t_{k+1} \le s \le t_{k-1}} [|\sigma_s^{\delta,t,\omega}| + |\sigma_s^{\delta',t,\omega}|]\delta_{k-1}^\alpha \\ &\qquad + \sup_{t \le s \le t_{n-1}} [|\sigma_s^{\delta,t,\omega}| + |\sigma_s^{\delta',t,\omega}|]\delta_{n-1}^\alpha\Bigg] \\ &\quad \le C[1 + \|\omega\|_T^\kappa]\Bigg[\sum_{k=m}^{n} \delta_{k+1}^{H-\frac{1}{2}}\delta_{k-1}^\alpha + \delta_{n+1}^{H-\frac{1}{2}}\delta_{n-1}^\alpha\Bigg] \\ &\quad \le C[1 + \|\omega\|_T^\kappa]\sum_{k=m}^{\infty} 2^{-\beta k} \le C[1 + \|\omega\|_T^\kappa]2^{-\beta m} \\ &\quad \le C[1 + \|\omega\|_T^\kappa]\delta^\beta \to 0 \quad \text{as } \delta \to 0. \end{aligned} \tag{3.45}$$

Similarly, by the second inequality of (3.39) and the first inequality of (3.38), we have

$$
\begin{aligned}
&\big|\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{\delta,t,\omega},\sigma^{\delta,t,\omega})\rangle-\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{\delta',t,\omega},\sigma^{\delta',t,\omega})\rangle\big| \\
&\quad\le\big|\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{\delta,t,\omega},\sigma^{\delta,t,\omega}-\sigma^{\delta',t,\omega})\rangle\big| \\
&\qquad+\big|\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{\delta,t,\omega}-\sigma^{\delta',t,\omega},\sigma^{\delta',t,\omega})\rangle\big| \\
&\quad\le C\big[1+\|\omega\|_T^\kappa\big]\Bigg[\sum_{k=m}^{n}\sup_{t_{k+1}\le s\le t_{k-1}}\big[|\sigma_s^{\delta,t,\omega}|+|\sigma_s^{\delta',t,\omega}|\big]^2\delta_{k-1}^{2\alpha} \\
&\qquad+\sup_{t\le s\le t_{n-1}}\big[|\sigma_s^{\delta,t,\omega}|+|\sigma_s^{\delta',t,\omega}|\big]^2\delta_{n-1}^{2\alpha}\Bigg] \qquad (3.46)\\
&\quad\le C\big[1+\|\omega\|_T^\kappa\big]\Bigg[\sum_{k=m}^{n}\delta_{k+1}^{2H-1}\delta_{k-1}^{2\alpha}+\delta_{n+1}^{2H-1}\delta_{n-1}^{2\alpha}\Bigg] \\
&\quad\le C\big[1+\|\omega\|_T^\kappa\big]\sum_{k=m}^{\infty}2^{-2\beta k}\le C\big[1+\|\omega\|_T^\kappa\big]\delta^{2\beta}\to 0\quad\text{as }\delta\to 0.
\end{aligned}
$$

This, together with (3.45), implies (3.41). Moreover, by sending $\delta'\to 0$, we obtain the following estimates which are stronger than (3.41):

$$
\begin{aligned}
&\big|\langle\partial_\omega u(t,\omega),\varphi^{\delta,t,\omega}\rangle-\langle\partial_\omega u(t,\omega),\varphi^{t,\omega}\rangle\big|\le C\big[1+\|\omega\|_T^\kappa\big]\delta^\beta,\qquad\varphi=b,\sigma,\\
&\big|\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{\delta,t,\omega},\sigma^{\delta,t,\omega})\rangle-\langle\partial^2_{\omega\omega}u(t,\omega),(\sigma^{t,\omega},\sigma^{t,\omega})\rangle\big| \qquad (3.47)\\
&\quad\le C\big[1+\|\omega\|_T^\kappa\big]\delta^{2\beta}.
\end{aligned}
$$

Step 2. Denote

$$
\begin{aligned}
X_t^\delta&:=x+\int_0^t b^\delta(t;r,X)\,dr+\int_0^t\sigma^\delta(t;r,X)\,dW_r,\\
\Theta_s^{\delta,t}&:=x+\int_0^t b^\delta(s;r,X)\,dr+\int_0^t\sigma^\delta(s;r,X)\,dW_r.
\end{aligned}
\qquad (3.48)
$$

We emphasize that in the above definitions, b^δ, σ^δ depend on X, not X^δ, since the truncation occurs in the first two parameters of b, σ only. In particular, X^δ is explicit given by X, and does not solve a SDE. For notational simplicity at below, we shall still omit X in the coefficients b, σ. Recall (3.4) and denote $\check{X}^{\delta,t}:=X^\delta\otimes_t\Theta^{\delta,t}$. In this step, we prove

$$
\mathbb{E}\big[\|\check{X}^{\delta,t}-\check{X}^t\|_T^{2p}\big]\le C_p\delta^{pH}\quad\text{for any }0\le t\le T,\ p\ge 1,\text{ and }\delta>0. \qquad (3.49)
$$

We first estimate the difference of Θ. By stochastic Fubini theorem, we have

$$\begin{aligned}
&\sup_{t\le s\le T}\left|\Theta_s^{\delta,t}-\Theta_s^t\right| \\
&\quad=\sup_{t\le s\le t+\delta}\left|\int_{s-\delta}^{t}\left[\left[b(r+\delta;r)-b(s;r)\right]dr+\left[\sigma(r+\delta;r)-\sigma(s;r)\right]dW_r\right]\right| \\
&\quad=\sup_{t\le s\le t+\delta}\left|\int_{s-\delta}^{t}\int_{s}^{r+\delta}\left[\partial_t b(\lambda;r)\,d\lambda\,dr+\partial_t\sigma(\lambda;r)\,d\lambda\,dW_r\right]\right| \\
&\quad=\sup_{t\le s\le t+\delta}\left|\int_{s}^{t+\delta}\int_{s-\delta}^{\lambda-\delta}\left[\partial_t b(\lambda;r)\,dr+\partial_t\sigma(\lambda;r)\,dW_r\right]d\lambda\right| \\
&\quad=\int_{t}^{t+\delta}\sup_{t-\delta\le l\le\lambda-\delta}\left|\int_{l}^{\lambda-\delta}\left[\partial_t b(\lambda;r)\,dr+\partial_t\sigma(\lambda;r)\,dW_r\right]\right|d\lambda.
\end{aligned}$$

Then, for any $p\ge 1$, by Burkholder–Davis–Gundy inequality and the second inequality of (3.38), we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\le s\le T}\left|\Theta_s^{\delta,t}-\Theta_s^t\right|^{2p}\Big] \\
&\quad\le C_p\delta^{2p-1}\int_t^{t+\delta}\mathbb{E}\Big[\sup_{t-\delta\le l\le\lambda-\delta}\left|\int_{l}^{\lambda-\delta}\left[\partial_t b(\lambda;r)\,dr+\partial_t\sigma(\lambda;r)\,dW_r\right]\right|^{2p}\Big]d\lambda \\
&\quad\le C_p\delta^{2p-1}\int_t^{t+\delta}\mathbb{E}\Big[\Big(\int_{t-\delta}^{\lambda-\delta}\left[\left|\partial_t b(\lambda;r)\right|^2+\left|\partial_t\sigma(\lambda;r)\right|^2\right]dr\Big)^p\Big]d\lambda \\
&\quad\le C_p\delta^{2p-1}\int_t^{t+\delta}\Big(\int_{t-\delta}^{\lambda-\delta}(\lambda-r)^{2H-3}\,dr\Big)^p d\lambda \\
&\quad= C_p\delta^{2p-1}\int_t^{t+\delta}\left(\delta^{2H-2}-(\delta+\lambda-t)^{2H-2}\right)^p d\lambda.
\end{aligned}$$

By changing variable, this implies

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\le s\le T}\left|\Theta_s^{\delta,t}-\Theta_s^t\right|^{2p}\Big] \\
&\quad\le C_p\delta^{2pH}\int_0^1\left(1-(1+\lambda)^{2H-2}\right)^p d\lambda=C_p\delta^{2pH}.
\end{aligned}\tag{3.50}$$

We next estimate the difference of X. For any $t<t'$, note that

$$\left|\left(X_t^\delta-X_t\right)-\left(X_{t'}^\delta-X_{t'}\right)\right|\le I_1+I_2,\tag{3.51}$$

where

$$I_1:=\left|\int_t^{t'}\left[b^\delta(t';r)-b(t';r)\right]dr+\int_t^{t'}\left[\sigma^\delta(t';r)-\sigma(t';r)\right]dW_r\right|,$$

$$I_2 := \left| \int_0^t [b^\delta(t;r) - b(t;r) - b^\delta(t';r) + b(t';r)]\,dr \right.$$
$$\left. + \int_0^t [\sigma^\delta(t;r) - \sigma(t;r) - \sigma^\delta(t';r) + \sigma(t';r)]\,dW_r \right|.$$

Denote $\delta' := t' - t$. For any $p \geq 1$, by the Burkholder–Davis–Gundy inequality and (3.38),

$$\begin{aligned}
\mathbb{E}[I_1^{2p}] &\leq C_p \mathbb{E}\left[\left(\int_t^{t'} [|b^\delta(t';s) - b(t';s)|^2 + |\sigma^\delta(t';s) - \sigma(t';s)|^2]\,ds\right)^p\right] \\
&= C_p \mathbb{E}\left[\left(\int_{t\vee(t'-\delta)}^{t'} [|b(s+\delta;s) - b(t';s)|^2 \right.\right. \\
&\qquad \left.\left. + |\sigma(s+\delta;s) - \sigma(t';s)|^2]\,ds\right)^p\right] \\
&\leq C_p \mathbb{E}\left[\left(\int_{t\vee(t'-\delta)}^{t'} \left[\int_{t'}^{s+\delta} [|\partial_t b(r;s)| + |\partial_t \sigma(r;s)|]\,dr\right]^2 ds\right)^p\right] \qquad (3.52)\\
&\leq C_p \left(\int_{t\vee(t'-\delta)}^{t'} \left|\int_{t'}^{s+\delta} (r-s)^{H-\frac{3}{2}}\,dr\right|^2 ds\right)^p \\
&= C_p \left(\int_0^{\delta\wedge\delta'} [r^{H-\frac{1}{2}} - \delta^{H-\frac{1}{2}}]^2\,dr\right)^p \\
&\leq C_p \left(\int_0^{\delta\wedge\delta'} r^{2H-1}\,dr\right)^p \leq C_p(\delta\wedge\delta')^{2pH}.
\end{aligned}$$

To estimate I_2, when $\delta' \geq \delta$, by (3.50) we have

$$\begin{aligned}
\mathbb{E}[|I_2|^{2p}] &= \mathbb{E}\left[\left|\int_0^t [b^\delta(t;s) - b(t;s)]\,ds \right.\right. \\
&\qquad \left.\left. + \int_0^t [\sigma^\delta(t;s) - \sigma(t;s)]\,dW_s\right|^{2p}\right] \qquad (3.53)\\
&= \mathbb{E}[|\Theta_t^{\delta,t} - \Theta_t|^{2p}] \leq C_p \delta^{2pH}.
\end{aligned}$$

When $\delta' < \delta$, one can check straightforwardly that

$$I_2 = \left|\int_{t-\delta}^t [b(t'\wedge(s+\delta);s) - b(t;s)]\,ds + \int_{t-\delta}^t [\sigma(t'\wedge(s+\delta);s) - \sigma(t;s)]\,dW_s\right|.$$

Then, again by the Burkholder–Davis–Gundy inequality and (3.38),

$$\begin{aligned}
\mathbb{E}[I_2^{2p}] &\leq C_p \left(\int_{t-\delta}^t \left[\int_t^{t'\wedge(s+\delta)} (r-s)^{H-\frac{3}{2}}\,dr\right]^2 ds\right)^p \\
&\leq C_p \left(\int_0^\delta [r^{H-\frac{1}{2}} - [(r+\delta')\wedge\delta]^{H-\frac{1}{2}}]^2\,dr\right)^p
\end{aligned}$$

$$
\begin{aligned}
&= C_p\Big(\int_0^{\delta-\delta'}\big[r^{H-\frac12}-(r+\delta')^{H-\frac12}\big]^2\,dr\Big)^p\\
&\quad + C_p\Big(\int_{\delta-\delta'}^{\delta}\big[r^{H-\frac12}-\delta^{H-\frac12}\big]^2\,dr\Big)^p \qquad (3.54)\\
&\le C_p(\delta')^{2pH}\Big(\int_0^{\infty}\big[r^{H-\frac12}-(r+1)^{H-\frac12}\big]^2\,dr\Big)^p\\
&\quad + C_p\Big(\int_{\delta-\delta'}^{\delta} r^{2H-1}\,dr\Big)^p.\\
&\le C_p(\delta')^{2pH} + C_p\big[\delta^{2H}-(\delta-\delta')^{2H}\big]^p \le C_p(\delta')^{2pH},
\end{aligned}
$$

where the last inequality thanks to the assumption that $H<\frac12$. Plug (3.52), (3.53), and (3.54) into (3.51), we get

$$
\begin{aligned}
&\mathbb{E}\big[\big|(X_t^\delta - X_t) - (X_{t'}^\delta - X_{t'})\big|^{2p}\big]\\
&\quad \le C_p(\delta\wedge\delta')^{2pH} = C_p\delta^{pH}(\delta')^{pH} = C_p\delta^{pH}|t'-t|^{pH}.
\end{aligned}
$$

Then by Lemma 3.11, we see that $\mathbb{E}[\|X^\delta - X\|_T^{2p}]\le C_p\delta^{pH}$. This, together with (3.50), proves (3.49).

Step 3. We now prove (3.16). We first note that $u(t,\check{X}^{\delta,t})$ falls short of satisfying the conditions in Theorem 3.10. In fact, X^δ is not a solution to the SDE (3.1) with coefficients (b^δ,σ^δ), and b^δ, σ^δ are not differentiable at $t=s+\delta$. However, by checking the arguments of the proof of Theorem 3.10 one can see that these differences do not cause any trouble, and thus the conclusion still holds true. Therefore, $u(t,\check{X}^{\delta,t})$ satisfies (3.16):

$$
\begin{aligned}
du(t,\check{X}^{\delta,t}) &= \partial_t u(t,\check{X}^{\delta,t})\,dt + \frac12\big\langle\partial^2_{\omega\omega}u(t,\check{X}^{\delta,t}),(\sigma^{\delta,t,X},\sigma^{\delta,t,X})\big\rangle\,dt\\
&\quad + \big\langle\partial_\omega u(t,\check{X}^{\delta,t}),b^{\delta,t,X}\big\rangle\,dt + \big\langle\partial_\omega u(t,\check{X}^{\delta,t}),\sigma^{\delta,t,X}\big\rangle\,dW_t,\quad \mathbb{P}\text{-a.s.}
\end{aligned}
$$

Then, by (3.49) and by the continuity of u and $\partial_t u$, we have

$$
\lim_{\delta\to 0}\mathbb{E}\big[\big|u(t,\check{X}^{\delta,t})-u(t,\check{X}^t)\big| + \big|\partial_t u(t,\check{X}^{\delta,t})-\partial_t u(t,\check{X}^t)\big|\big]=0. \tag{3.55}
$$

Moreover, recall the notation in (3.42) and assume $\delta_{n+1}\le\delta<\delta_n$. Set $m=1$ in (3.43):

$$
\mathbf{1}_{[t_1,T]} = [1-\psi_1]\mathbf{1}_{[t_1,T]} + \sum_{k=1}^n \psi_k + [1-\psi_n]\mathbf{1}_{[t,t_n]}.
$$

Then by (3.47) and (3.40) we have, for $\varphi=b,\sigma$,

$$
\begin{aligned}
&\big|\big\langle\partial_\omega u(t,\check{X}^{\delta,t}),\varphi^{\delta,t,X}\big\rangle - \big\langle\partial_\omega u(t,\check{X}^t),\varphi^{t,X}\big\rangle\big|\\
&\quad\le \big|\big\langle\partial_\omega u(t,\check{X}^{\delta,t})-\partial_\omega u(t,\check{X}^t),\varphi^{\delta,t,X}\big\rangle\big| + \big|\big\langle\partial_\omega u(t,\check{X}^t),\varphi^{\delta,t,X}-\varphi^{t,X}\big\rangle\big|
\end{aligned}
$$

$$
\begin{aligned}
&\le C[1+\|\check{X}^t\|_T^\kappa]\delta^\beta + \bigg|\bigg\langle \partial_\omega u(t,\check{X}^{\delta,t}) - \partial_\omega u(t,\check{X}^t), \\
&\quad \bigg[[1-\psi_1]\mathbf{1}_{[t_1,T]} + \sum_{k=1}^n \psi_k + [1-\psi_n]\mathbf{1}_{[t,t_n]}\bigg]\varphi^{\delta,t,X}\bigg\rangle\bigg| \\
&\le C[1+\|\check{X}^t\|_T^\kappa + \|\check{X}^{\delta,t}\|_T^\kappa]\bigg[\rho(\|\check{X}^{\delta,t}-\check{X}^t\|_T)\bigg[1+\sum_{k=1}^n \delta_k^\beta + \delta_n^\beta\bigg] + \delta^\beta\bigg] \\
&\le C[1+\|\check{X}^t\|_T^\kappa + \|\check{X}^{\delta,t}\|_T^\kappa][\rho(\|\check{X}^{\delta,t}-\check{X}^t\|_T) + \delta^\beta].
\end{aligned}
$$

Similarly, by (3.47) and (3.40) we have

$$
\begin{aligned}
&|\langle \partial^2_{\omega\omega} u(t,\check{X}^{\delta,t}), (\sigma^{\delta,t,X},\sigma^{\delta,t,X})\rangle - \langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X},\sigma^{t,X})\rangle| \\
&\quad \le |\langle \partial^2_{\omega\omega} u(t,\check{X}^{\delta,t}) - \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{\delta,t,X},\sigma^{\delta,t,X})\rangle| \\
&\qquad + |\langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{\delta,t,X},\sigma^{\delta,t,X})\rangle - \langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X},\sigma^{t,X})\rangle| \\
&\quad \le C[1+\|X\|_T^\kappa]\delta^{2\beta} + \bigg|\bigg\langle \partial^2_{\omega\omega} u(t,\check{X}^{\delta,t}) - \partial^2_{\omega\omega} u(t,\check{X}^t), \\
&\qquad \bigg(\sigma^{\delta,t,X}, \bigg[[1-\psi_1]\mathbf{1}_{[t_1,T]} + \sum_{k=1}^n \psi_k + [1-\psi_n]\mathbf{1}_{[t,t_n]}\bigg]\sigma^{\delta,t,X}\bigg)\bigg\rangle\bigg| \\
&\quad \le C[1+\|\check{X}^t\|_T^\kappa + \|\check{X}^{\delta,t}\|_T^\kappa]\bigg[\rho(\|\check{X}^{\delta,t}-\check{X}^t\|_T)\bigg[1+\sum_{k=1}^n \delta_k^{2\beta} + \delta_n^{2\beta)}\bigg] + \delta^{2\beta}\bigg] \\
&\quad \le C[1+\|\check{X}^t\|_T^\kappa + \|\check{X}^{\delta,t}\|_T^\kappa][\rho(\|\check{X}^{\delta,t}-\check{X}^t\|_T) + \delta^{2\beta}].
\end{aligned}
$$

Put together, we derive from (3.49) that

$$
\lim_{\delta\to 0}\mathbb{E}[|\langle \partial_\omega u(t,\check{X}^{\delta,t}), \varphi^{\delta,t,X}\rangle - \langle \partial_\omega u(t,\check{X}^t), \varphi^{t,X}\rangle|^2] = 0,
$$

$$
\lim_{\delta\to 0}\mathbb{E}[|\langle \partial^2_{\omega\omega} u(t,\check{X}^{\delta,t}), (\sigma^{\delta,t,X},\sigma^{\delta,t,X})\rangle - \langle \partial^2_{\omega\omega} u(t,\check{X}^t), (\sigma^{t,X},\sigma^{t,X})\rangle|] = 0.
$$

Plug this and (3.55) into (3.55), we obtain (3.16) in this singular case. □

4. Path dependent PDEs and Feynman–Kac formulae.

4.1. *The linear case.* We first apply Theorem 3.17 to the path-dependent example of Section 2.3. This works because the kernel of fBm satisfies all the hypotheses on b and σ in Theorem 3.17, including the ones where u is weakly dependent on paths near its time diagonal. In fact, we find that we may choose $\alpha = 1/2$ in Definition 3.16.

THEOREM 4.1. *Consider the setting in Section 2.3, and denote*

$$u(t,\omega) := \mathbb{E}\Big[g\Big(\omega_T + \int_t^T K(T,r)\,dW_r\Big) + \int_t^T f\Big(s, \omega_s + \int_t^s K(s,r)\,dW_r\Big)\,ds\Big]. \tag{4.1}$$

Assume f is continuous in t; and for $\varphi = g, f(t,\cdot)$, $\varphi \in C^2(\mathbb{R})$ such that all the derivatives have polynomial growth with $|\varphi''(t,x) - \varphi''(t,\tilde{x})| \le C[1 + |x|^\kappa + |\tilde{x}|^\kappa]\rho(|x-\tilde{x}|)$. Then:

(i) *u evaluated at $B^H \otimes_t \Theta^t$ coincides with the conditional expectation*:

$$Y_t := \mathbb{E}\Big[g(B_T^H) + \int_t^T f(s, B_s^H)\,ds \Big| \mathcal{F}_t\Big] = u(t, B^H \otimes_t \Theta^t). \tag{4.2}$$

(ii) *$u \in C_+^{1,2}(\overline{\Lambda})$ with path derivatives*:

$$\begin{aligned}
&\langle \partial_\omega u(t,\omega), \eta\rangle \\
&\quad = \mathbb{E}\Big[g'\Big(\omega_T + \int_t^T K(T,r)\,dW_r\Big)\eta_T \\
&\qquad + \int_t^T f'\Big(s, \omega_s + \int_t^s K(s,r)\,dW_r\Big)\eta_s\,ds\Big], \\
&\langle \partial^2_{\omega\omega} u(t,\omega), (\eta_1,\eta_2)\rangle \\
&\quad = \mathbb{E}\Big[g''\Big(\omega_T + \int_t^T K(T,r)\,dW_r\Big)\eta_1(T)\eta_2(T) \\
&\qquad + \int_t^T f''\Big(s, \omega_s + \int_t^s K(s,r)\,dW_r\Big)\eta_1(s)\eta_2(s)\,ds\Big].
\end{aligned} \tag{4.3}$$

(iii) *u vanishes diagonally with rate $\alpha = \frac{1}{2}$, in the sense of Definition 3.16. Consequently, the functional Itô formula (3.16) holds true for all $H \in (0,1)$.*

(iv) *u is a classical solution to the following linear PPDE which, together with (4.3), provides a representation for $\partial_t u$*:

$$\begin{aligned}
&\partial_t u(t,\omega) + \frac{1}{2}\langle \partial^2_{\omega\omega} u(t,\omega), (K^t, K^t)\rangle + f(t,\omega_t) = 0, \quad (t,\omega) \in \Lambda; \\
&\qquad u(T,\omega) = g(\omega_T).
\end{aligned} \tag{4.4}$$

PROOF. We shall only prove the irregular case $H < \frac{1}{2}$. The regular case $H \ge \frac{1}{2}$ follows by similar but easier arguments.

First, (4.2) follows directly from the arguments in Section 2.2.

Next, applying (3.6) and (3.11) on (4.1), one may easily verify (4.3), and that $\partial_\omega u$, $\partial^2_{\omega\omega} u$ are continuous, have polynomial growth, and satisfy (3.12), (3.39) and (3.40) with $\alpha = \frac{1}{2}$.

Moreover, denote $\overline{\sigma}^2(s,t) := \int_t^s K^2(s,r)\,dr < \infty$, $0 \le t < s \le T$. Then $\int_t^s K(s,r)\,dW_r$ has distribution Normal $(0, \overline{\sigma}^2(s,t))$. Thus, denoting by $\mathcal{N}$ a standard normal distribution,

$$u(t,\omega) = \mathbb{E}\Big[g\big(\omega_T + \overline{\sigma}(T,t)\mathcal{N}\big)) + \int_t^T f\big(s, \omega_s + \overline{\sigma}(s,t)\mathcal{N}\big)\,ds\Big].$$

Note that $\partial_t \overline{\sigma}(s,t) = -\frac{K^2(s,t)}{2\overline{\sigma}(s,t)}$. Then one can easily see that

$$\begin{aligned}\partial_t u(t,\omega) &= -\frac{1}{2}\mathbb{E}\Big[\big[g'\big(\omega_T + \overline{\sigma}(T,t)\mathcal{N}\big) - g'(\omega_T)\big]\frac{K^2(T,t)}{\overline{\sigma}(T,t)}\mathcal{N} \\ &\qquad + \int_t^T \big[f'\big(s,\omega_s + \overline{\sigma}(s,t)\mathcal{N}\big) - f'(s,\omega_s)\big]\frac{K^2(s,t)}{\overline{\sigma}(s,t)}\mathcal{N}\,ds\Big] \\ &= -\frac{1}{2}\int_0^1 \mathbb{E}\Big[g''\big(\omega_T + \lambda\overline{\sigma}(T,t)\mathcal{N}\big)K^2(T,t)\mathcal{N}^2 \\ &\qquad + \int_t^T f''\big(s,\omega_s + \lambda\overline{\sigma}(s,t)\mathcal{N}\big)K^2(s,t)\mathcal{N}^2\,ds\Big]\,d\lambda,\end{aligned} \tag{4.5}$$

where, to justify the integrability in the right-hand side above, we note that

$$\big|\partial_t u(t,\omega)\big| \le C\Big[K^2(T,t) + \int_t^T K^2(s,t)\,ds\Big]\big[1 + \|\omega\|_T^\kappa\big] \le C\big[1 + \|\omega\|_T^\kappa\big].$$

This means that $\partial_t u$ exists and has polynomial growth. By (4.5), one can also see that $\partial_t u$ is continuous. Then $u \in C_+^{1,2}(\overline{\Lambda})$.

Finally, note that

$$u\big(t, B^H \otimes_t \Theta^t\big) + \int_0^t f\big(s, B_s^H\big)\,ds = \mathbb{E}\Big[g\big(B_T^H\big) + \int_0^T f\big(s, B_s^H\big)\,ds\Big|\mathcal{F}_t\Big]$$

is a martingale. Applying the functional Itô formula (3.16) on $u(t, B^H \otimes_t \Theta^t)$, we see that (4.4) holds on $B^H \otimes_t \Theta^t$, $\mathbb{P}$-a.s. In particular, (4.4) holds at $(0,0)$. Given $(t,\omega) \in \overline{\Lambda}$, apply the same arguments on the system starting with (t,ω), in the spirit of the proof of Proposition 3.7 (Step 2 of that joint proof), we can see that (4.4) holds at (t,ω) as well. □

4.2. *The semilinear case.* In this subsection, we consider the following BSDE:

$$Y_t = g(X_\cdot) + \int_t^T f(s, X_\cdot, Y_s, Z_s)\,ds - \int_t^T Z_s\,dW_s, \quad 0 \le t \le T, \tag{4.6}$$

where $f = f(s, X_{s\wedge\cdot}, y, z)$ is adapted. We emphasize again that, as we saw in the previous section, even if the coefficients b, σ, f, g are state dependent (namely depending only on X_s at time s), the BSDE is not Markovian as soon as X is not a Markov process. When X is a strong solution to SDE (3.1), namely X is

$\mathbb{F}^W$-progressively measurable, then it follows from the seminal work Pardoux and Peng [32] that the above BSDE is well-posed, provided f is uniformly Lipschitz-continuous in (y,z). When X is a weak solution and no strong solution exists, then typically one needs to introduce an orthogonal martingale term in the BSDE (4.6). We avoid this situation in the sequel, though the uniqueness of a strong solution to (4.6) is not a requirement.

This BSDE is closely related to the following semilinear PPDE, where the notation is that of Section 3.1:

$$\begin{aligned}
&\partial_t u(t,\omega) + \frac{1}{2}\langle \partial^2_{\omega\omega} u(t,\omega), (\sigma^{t,\omega}, \sigma^{t,\omega})\rangle + \langle \partial_\omega u(t,\omega), b^{t,\omega}\rangle \\
&\qquad + f\bigl(t,\omega,u(t,\omega),\langle \partial_\omega u(t,\omega), \sigma^{t,\omega}\rangle\bigr) = 0, \quad (t,\omega)\in\Lambda, \qquad (4.7)\\
&u(T,\omega) = g(\omega).
\end{aligned}$$

We have the following Feynman–Kac formula.

THEOREM 4.2. *Let Assumption 3.1 hold and assume the semilinear PPDE (4.7) has a classical solution $u\in C^{1,2}_+(\Lambda)$. Assume further that either Assumption 3.9 holds or Assumption 3.15 holds and $u\in C^{1,2}_{+,\alpha}(\Lambda)$ for some $\alpha>\frac{1}{2}-H$. Then the BSDE (4.6) has a strong solution*:

$$Y_t := u\bigl(t,\check{X}^t\bigr), \qquad Z_t := \bigl\langle \partial_\omega u\bigl(t,\check{X}^t\bigr), \sigma^{t,X}\bigr\rangle. \qquad (4.8)$$

PROOF. By our assumptions, $u(t,\check{X}^t)$ satisfies the functional Itô formula (3.16). Then it is straightforward to verify that the process (Y,Z) defined by (4.8) satisfies (4.6). $\square$

We note that when the PPDE has a classical solution, (4.8) provides a solution to the BSDE (4.6) even if X is a weak solution to (3.1). However, in this case typically (Y,Z) are also not $\mathbb{F}^W$-progressively measurable.

We now proceed in the opposite direction, namely to provide a representation for the solution of PPDE (4.7) through the BSDE (4.6). For each $(t,\omega)\in\overline{\Lambda}$, define

$$u(t,\omega) := Y^{t,\omega}_t, \qquad (4.9)$$

where, for $t\le s\le T$,

$$X^{t,\omega}_s = \omega_s + \int_t^s b\bigl(s;r,\omega\otimes_t X^{t,\omega}\bigr)\,dr + \int_t^s \sigma\bigl(s;r,\omega\otimes_t X^{t,\omega}\bigr)\,dW_r,$$

$$Y^{t,\omega}_s = g\bigl(\omega\otimes_t X^{t,\omega}\bigr) + \int_s^T f\bigl(r,\omega\otimes_t X^{t,\omega}, Y^{t,\omega}_r, Z^{t,\omega}_r\bigr)\,dr - \int_s^T Z^{t,\omega}_r\,dW_r.$$

Here, we are assuming that the FBSDE in (4.9) has a unique strong solution for all $(t,\omega)\in\overline{\Lambda}$. With that assumption, for fixed $(t,\omega)\in\overline{\Lambda}$, the pair of processes $(Y^{t,\omega},Z^{t,\omega})$ is given unambiguously by the FBSDE, and u in (4.9) is well defined on $\overline{\Lambda}$. We avoid further technical discussion, stating the representation result with comments only.

REMARK 4.3. (i) Provided appropriate conditions on the coefficients of (3.1) and (4.6), one can show that the u defined by (4.9) is indeed smooth and is the classical solution to PPDE (4.7). See Peng and Wang [35] for a related result in the Brownian motion framework.

(ii) Our BSDE (4.6) is time consistent: the coefficient f depends only on one time variable and g is independent of time. We refer to Yong [39] and the references therein for Volterra-type BSDEs.

4.3. *A strategy for control problems*. The framework of the previous subsection applies directly, as a slight extension, if there is control involved. Formally, one can easily write down the path-dependent Hamilton–Jacobi–Bellman (HJB) equation. More precisely, let $\mathcal{A}$ be an appropriate set of admissible controls taking values in certain set A; X solve a controlled Volterra SDE; and Y solve a controlled BSDE. We define the value function u for the control problem as follows:

$$u(t,\omega) := \sup_{a\in\mathcal{A}} Y_t^{t,\omega,a}, \tag{4.10}$$

where, for $t \le s \le T$,

$$X_s^{t,\omega,a} = \omega_s + \int_t^s b(s; r, \omega \otimes_t X^{t,\omega,a}, a_s)\,dr + \int_t^s \sigma(s; r, \omega \otimes_t X^{t,\omega,a}, a_s)\,dW_r,$$

$$Y_s^{t,\omega,a} = g(\omega \otimes_t X^{t,\omega,a}) + \int_s^T f(r, \omega \otimes_t X^{t,\omega,a}, Y_r^{t,\omega,a}, Z_r^{t,\omega,a}, a_s)\,dr - \int_s^T Z_r^{t,\omega,a,}\,dW_r.$$

Then formally u should satisfy the following path dependent HJB equation:

$$\begin{aligned} &\partial_t u(t,\omega) + \sup_{a\in A}\Big[\frac{1}{2}\langle \partial^2_{\omega\omega} u(t,\omega), (\sigma^{t,\omega,a}, \sigma^{t,\omega,a})\rangle + \langle \partial_\omega u(t,\omega), b^{t,\omega,a}\rangle \\ &\quad + f(t,\omega,u(t,\omega), \langle \partial_\omega u(t,\omega), \sigma^{t,\omega,a}\rangle, a_t)\Big] = 0, \quad (t,\omega)\in\Lambda, \end{aligned} \tag{4.11}$$

with terminal condition $u(T,\omega) = g(\omega)$. Here, for $\varphi = b, \sigma$, $\varphi_s^{t,\omega,a} := \varphi(s; t, \omega, a)$. See Fouque and Hu [22] for an application in this direction.

When the path dependent HJB equation (4.11) has a classical solution $u \in C_+^{1,2}(\Lambda)$ or when the value function u defined by (4.10) is indeed in $C_+^{1,2}(\Lambda)$ (or $u \in C_{+,\alpha}^{1,2}(\Lambda)$ for some appropriate α in the singular case), it is not difficult to prove that they are equal, as in the standard verification theorem. However, in general it is difficult to expect a classical solution for such a control problem, because of the path dependence. We shall study viscosity solutions, in the spirit of Ekren et al. [18, 19], for these fully nonlinear PPDEs in our future research.

5. An application to finance. In the reference El Euch and Rosenbaum [20], the authors work with the so-called rough Heston model, whose well-posedness was established in their previous publication [21]. In [20], they show that options on equities given by this model can be hedged if one assumes that the volatility is observed. In fact, for an option on a given equity, they argue that, since the forward variance can be replicated in the market using liquid instruments, then all that is required for hedging purposes is observation of that forward variance and the equity's spot price. We will describe their model more precisely, how it fits in ours and what more pricing and hedging questions can be reached in ours.

Recall Section 1.2 and in particular (1.5). Consider the following rough Heston model:

$$\begin{aligned} S_t &= S_0 + \int_0^t S_r\sqrt{V_r}\big[\sqrt{1-\rho^2}\,dW_r^1 + \rho\,dW_r^2\big], \\ V_t &= V_0 + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-r)^{H-\frac{1}{2}}\big[\lambda[\theta - V_r]\,dr + \nu\sqrt{V_r}\,dW_r^2\big]. \end{aligned} \tag{5.1}$$

Here, $W = (W^1, W^2)^\top$ is a two-dimensional Brownian motion, $\rho \in [0, 1]$ is a correlation parameter, θ is a mean-reversion level, λ is a mean-reversion rate, ν is a noise intensity and the roughness parameter H is typically in $(0, \frac{1}{2})$. We leave aside the question of whether any of these parameters can be estimated or calibrated from the data. We note instead that the term $(t-r)^{H-\frac{1}{2}}$ is similar to the kernel K of fBm (it is in fact identical to the kernel of the so-called Riemann–Liouville fBm). By [21], the SDE (5.1) has a unique weak solution $X := (S, V)^\top$. The paper [20] asks the question of how to compute the conditional expectation of a nonpath-dependent contingent claim at any time prior to maturity:

$$Y_t := C_t := \mathbb{E}\big[g(S_T)|\mathcal{F}_t\big] \tag{5.2}$$

for some deterministic contract function g. They express C_t as a function of S_t as well as the so-called forward variance:

$$\hat{\Theta}_s^t := \mathbb{E}[V_s|\mathcal{F}_t], \quad 0 \le t \le s \le T. \tag{5.3}$$

Note that both the forward variance $\hat{\Theta}_s^t$ and the forward volatility $\mathbb{E}[\sqrt{V_s}|\mathcal{F}_t]$ introduced in Section 1.2 are financial indexes available in the market. A PPDE is derived for C_t in this special case. Moreover, the forward variance can be approximated by using liquid variance swaps or vanilla options, and in this sense one may view the forward variance as a set of additional tradable assets. The main contribution of [20] is to provide a perfect hedge for the derivative $g(S_T)$ by using the stock S and the forward variance $\hat{\Theta}$. The hedging portfolio relies on the Frechet derivative of C_t and certain characteristic functions, which requires the special structure of (5.1) and that $C_T = g(S_T)$ is state dependent.

We now explain how our framework covers the above example and beyond. First note that, for $X = (S, V)^\top$, (5.1) is a Volterra SDE (3.1) with

$$
\begin{aligned}
b(t; r, x_1, x_2) &= \begin{bmatrix} 0 \\ \dfrac{\lambda (t-r)^{H-\frac{1}{2}}[\theta - x_2]}{\Gamma(H+\frac{1}{2})} \end{bmatrix}, \\
\sigma(t; r, x_1, x_2) &= \begin{bmatrix} \sqrt{1-\rho^2} x_1 \sqrt{x_2} & \rho x_1 \sqrt{x_2} \\ 0 & \dfrac{\nu (t-r)^{H-\frac{1}{2}} \sqrt{x_2}}{\Gamma(H+\frac{1}{2})} \end{bmatrix}.
\end{aligned} \tag{5.4}
$$

One may easily check that (5.1) satisfies all the properties in Assumptions 3.1 and 3.15, needed in Section 3.3 for $H \in (0, 1/2)$; see Remark A.2 below. Note that the dynamics of S is standard, without involving a two-time-variable kernel. While we may apply the results in previous sections directly on the two-dimensional SDE (5.1), for simplicity we restrict the path dependence only to the dynamics of V. Therefore, recall (2.6), for $t < s$ we denote

$$
\Theta^t_s := V_0 + \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (s-r)^{H-\frac{1}{2}} \big[\lambda[\theta - V_r]\, dr + \nu \sqrt{V_r}\, dW^2_r\big]. \tag{5.5}
$$

By the special structure of the rough Heston model, we can actually see that

$$
C_t = u\big(t, S_t, \Theta^t_{[t,T]}\big). \tag{5.6}
$$

In particular, the dependence of C_t on S is only via S_t and its dependence on V does not involve $V_{[0,t)}$. Denote u as $u(t, x, \omega)$ and we shall assume g is smooth which would imply the smoothness of u. Now following the arguments in Section 4.1, in particular noting that C is a martingale, we see that u satisfies the following PPDE:

$$
\begin{aligned}
&\partial_t u + \frac{\lambda[\theta - \omega_t]}{\Gamma(H+\frac{1}{2})} \langle \partial_\omega u, a^t \rangle + \frac{x^2 \omega_t}{2} \partial^2_{xx} u + \frac{\rho \nu x \omega_t}{\Gamma(H+\frac{1}{2})} \langle \partial_\omega (\partial_x u), a^t \rangle \\
&\quad + \frac{\nu^2 \omega_t}{2\Gamma(H+\frac{1}{2})} \langle \partial^2_{\omega\omega} u, (a^t, a^t) \rangle = 0 \quad \text{where } a^t_s := (s-t)^{H-\frac{1}{2}}.
\end{aligned} \tag{5.7}
$$

Moreover, by Theorem 3.17, we have (recalling $V_t = \Theta^t_t$)

$$
dC_t = \partial_x u\big(t, S_t, \Theta^t_{[t,T]}\big)\, dS_t + \frac{\nu \sqrt{V_t}}{\Gamma(H+\frac{1}{2})} \big\langle \partial_\omega u\big(t, S_t, \Theta^t_{[t,T]}\big), a^t \big\rangle\, dW^2_t. \tag{5.8}
$$

The first term in the right-hand side above obviously provides the Δ-hedging in terms of the stock S. Note further that $t \mapsto \Theta^t_T$ is a semimartingale, and we have

$$
\frac{\nu \sqrt{V_t}}{\Gamma(H+\frac{1}{2})}\, dW^2_t = (T-t)^{\frac{1}{2}-H}\, d\Theta^t_T - \frac{\lambda[\theta - V_t]}{\Gamma(H+\frac{1}{2})}\, dt.
$$

Then

$$(5.9)\quad \begin{aligned} dC_t &= \partial_x u(t, S_t, \Theta^t_{[t,T]})\, dS_t + (T-t)^{\frac{1}{2}-H}\langle \partial_\omega u(t, S_t, \Theta^t_{[t,T]}), a^t\rangle\, d\Theta^t_T \\ &\quad - \frac{\lambda[\theta - V_t]}{\Gamma(H+\frac{1}{2})}\langle \partial_\omega u(t, S_t, \Theta^t_{[t,T]}), a^t\rangle\, dt. \end{aligned}$$

That is, provided that we could replicate Θ^t_T using market instruments, which we will discuss in details below, then we may (perfectly) hedge $g(S_T)$ as claimed in [20].

We note that our Θ^t in (5.5) is different from the forward variance $\hat{\Theta}^t$ in (5.3). However, it can easily be replicated by using $\hat{\Theta}^t$, which can further be replicated (approximately) by variance swaps. Indeed, by (5.5) and taking conditional expectation on the dynamics of V in (5.1), we see that

$$(5.10)\qquad \hat{\Theta}^t_s = \Theta^t_s + \frac{1}{\Gamma(H+\frac{1}{2})}\int_t^s (s-r)^{H-\frac{1}{2}}\lambda[\theta - \hat{\Theta}^t_r]\, dr, \quad t \le s \le T.$$

For any fixed t, clearly Θ^t_s is uniquely determined by $\{\hat{\Theta}^t_r\}_{t\le r\le s}$:

$$(5.11)\qquad \Theta^t_s = \hat{\Theta}^t_s - \frac{1}{\Gamma(H+\frac{1}{2})}\int_t^s (s-r)^{H-\frac{1}{2}}\lambda[\theta - \hat{\Theta}^t_r]\, dr.$$

In particular, this implies that, provided we observe the forward variance $\hat{\Theta}^t_s$, the process Θ^t_s is also observable at t. Moreover, as a function of t,

$$d\Theta^t_T = d\hat{\Theta}^t_T + \frac{1}{\Gamma(H+\frac{1}{2})}(T-t)^{H-\frac{1}{2}}\lambda[\theta - \hat{\Theta}^t_t]\, dt.$$

Plug this into (5.9) and note that $\hat{\Theta}^t_t = V_t$, we obtain

$$(5.12)\quad dC_t = \partial_x u(t, S_t, \Theta^t_{[t,T]})\, dS_t + (T-t)^{\frac{1}{2}-H}\langle \partial_\omega u(t, S_t, \Theta^t_{[t,T]}), a^t\rangle\, d\hat{\Theta}^t_T.$$

That is, C_T can be replicated by using S_t and $\hat{\Theta}^t_T$, with the corresponding hedging portfolios $\partial_x u$ and $(T-t)^{\frac{1}{2}-H}\langle \partial_\omega u, a^t\rangle$, respectively.

REMARK 5.1. We notice that, to hedge $C_T = g(S_T)$ in the rough Heston model, it is sufficient to use S and $\hat{\Theta}^{\cdot}_T$. However, if we want to hedge $C_T = g(S_T) + \int_0^T f(t, S_t)\, dt$ (or even more general path dependent contingent claims, which is not covered by [20]), then we shall need S and $\{\hat{\Theta}^{\cdot}_s\}_{0\le s\le T}$. Indeed, in this case we will have

$$C_t := \mathbb{E}\Big[g(S_T) + \int_t^T f(s, S_s)\, ds \Big| \mathcal{F}_t\Big] = u(t, S_t, \Theta^t_{[t,T]}),$$

and, provided u is smooth,

$$dC_t = \partial_x u\big(t, S_t, \Theta^t_{[t,T]}\big)\,dS_t + (T-t)^{\frac{1}{2}-H}\big\langle \partial_\omega u\big(t, S_t, \Theta^t_{[t,T]}\big), a^t\big\rangle\, d\hat{\Theta}^t_T + \int_t^T (s-t)^{\frac{1}{2}-H}\big\langle \partial_\omega u\big(t, S_t, \Theta^t_{[t,T]}\big), a^t\big\rangle\, d\hat{\Theta}^t_s\,ds - f(t, S_t)\,dt.$$

Thus, the portfolio of $\hat{\Theta}^{\cdot}_s$ at time t is $(s-t)^{\frac{1}{2}-H}\langle \partial_\omega u(t, S_t, \Theta^t_{[t,T]}), a^t\rangle\, ds$.

We would like to comment further on how to replicate Θ^t by using $\hat{\Theta}^t$ in more general cases. Mathematically, as we saw in previous sections, Θ^t is intrinsically more appropriate for this framework. In fact, recall (2.4) and the discussion afterwards. In the general model (3.1), if we use $\hat{\Theta}^t_s := \mathbb{E}[V_s|\mathcal{F}_t]$ as our "state variable," it is not clear if one would be able to derive a sensible PPDE. However, it is clear that $\hat{\Theta} = \Theta$ when $b = 0$. For the rough Heston model (5.1), thanks to the fact that its drift b is linear in V, Θ^t and $\hat{\Theta}^t$ are still equivalent in the following sense. Given $\Theta^t_{[t,T]}$, (5.10) is a linear convolution ODE which, by Laplace transformation, has a unique solution $\hat{\Theta}^t$: denoting $\alpha := H + \frac{1}{2}$,

$$\hat{\Theta}^t_s = \Theta^t_s + \frac{\lambda(s-t)^\alpha}{\Gamma(\alpha+1)} + \int_t^s \left[\sum_{n=1}^{\infty} \frac{(-\lambda)^n}{\Gamma(n\alpha)}(s-r)^{n\alpha-\frac{1}{2}}\right]\left[\Theta^t_r + \frac{\lambda(r-t)^\alpha}{\Gamma(\alpha+1)}\right]dr. \tag{5.13}$$

Together with (5.11), we have a one-to-one mapping between the paths $\Theta^t_{[t,T]}$ and $\hat{\Theta}^t_{[t,T]}$. In this sense, it is conceivable to write C_t as a function of $\hat{\Theta}^t$ in the rough Heston model, and we believe this is the underlying reason that a PPDE could be derived in [20]. The same arguments would work for the affine Volterra process in Abi Jaber, Larsson and Pulido [1], where V satisfies the following convolution type of Volterral SDE:

$$V_t = V_0 + \int_0^t K(t-r)\big[[b_0 + b_1 V_r]\,dr + \sqrt{a_0 + a_1 V_r}\,dW^2_r\big]. \tag{5.14}$$

However, we emphasize that one cannot extend (5.10) when the volatility V satisfies the following general model with nonlinear b:

$$V_t = V_0 + \int_0^t b(t; r, V_r)\,dr + \int_0^t \sigma(t; r, V_r)\,dW^2_r. \tag{5.15}$$

In this case, as before we denote

$$\Theta^t_s := V_0 + \int_0^t b(s; r, V_r)\,dr + \int_0^t \sigma(s; r, V_r)\,dW^2_r. \tag{5.16}$$

Then we have

$$\Theta^t_s = \mathbb{E}[V_s|\mathcal{F}_t] - \int_t^s \mathbb{E}\big[b(s; r, V_r)|\mathcal{F}_t\big]\,dr. \tag{5.17}$$

As we mentioned above, the forward variance $\mathbb{E}[V_s|\mathcal{F}_t]$ can be replicated by using variance swaps. For nonlinear b, under technical conditions, by Carr and Madan [7] one may replicate $\mathbb{E}[b(s;r,V_r)|\mathcal{F}_t]$, and hence Θ_s^t provided one can replicate the variance options $\mathbb{E}[(V_s-K)^+|\mathcal{F}_t]$, or the volatility options $\mathbb{E}[(\sqrt{V_s}-K)^+|\mathcal{F}_t]$, for all K. We note again that a wide range of volatility options are available in the financial market, at least for the S&P 500; see the VIX options in Gatheral [23].

To conclude this article, we point out that our framework can cover much more general models than the rough Heston model (5.1). As already mentioned, we allow for nonlinear b (and σ) in (5.15) and we can still derive the PPDE and provide a perfect hedge for $g(S_T)$, as long as the PPDE has a classical solution and Θ can be replicated as we discussed above. In addition, our framework also covers the fractional Stein–Stein model, where $\sqrt{V}$ is Gaussian and is the fractional Ornstein–Uhlenbeck process; see Comte and Renault [9], Chronopoulou and Viens [8] and Gulisashvili, Viens and Zhang [26] and references therein. Besides the generality of the underlying model, we also allow for more general derivatives. On the one hand, the derivatives can be path dependent in our framework; for instance, we discussed the special case $C_T=g(S_T)+\int_0^T f(t,S_t)\,dt$ in Section 2.3 and Remark 5.1. On the other hand, we can allow for nonlinear pricing (e.g., when the borrowing and lending interest rates are different) as a solution of the BSDE (4.6). We leave these details to the interested readers and further investigations.

REMARK 5.2. In this remark, we provide more details concerning the rough Bergomi model considered in Bayer, Friz and Gatheral [3]. Here, we shall only formally discuss the hedging issues, and leave some technical issues in Remark A.4 below. Let S be as in (5.1), but the variance V is replaced with

$$V_t := V_0\exp\left(M_t-\frac{1}{2}\lambda^2t^{2H}\right),\qquad M_t=\lambda\sqrt{2H}\int_0^t(t-r)^{H-\frac{1}{2}}\,dW_r^2. \tag{5.18}$$

We note that the variance V is not in the form (5.15), so the situation here is slightly different from above. However, clearly the dynamics of $X=(S,M)$ is in the form of Volterra SDE (3.1) with two-dimensional W and $b=0$,

$$\sigma(t;r,x_1,x_2)=\begin{bmatrix}\sqrt{1-\rho^2}x_1\sqrt{V_0e^{x_2-\frac{1}{2}\lambda^2t^{2H}}} & \rho x_1\sqrt{V_0e^{x_2-\frac{1}{2}\lambda^2t^{2H}}}\\ 0 & \lambda\sqrt{2H}(t-r)^{H-\frac{1}{2}}\end{bmatrix}.$$

As in (5.1), the dynamics of S is linear, and thus has explicit representation:

$$S_t=S_0\exp\left(\int_0^t\sqrt{V_s}d\tilde{W}_s-\frac{1}{2}\int_0^tV_s\,ds\right), \tag{5.19}$$

where $\tilde{W}_t:=\sqrt{1-\rho^2}W_t^1+\rho\,dW_t^2$.

In this case, $\Theta_s^t=\mathbb{E}[M_s|\mathcal{F}_t]=\lambda\sqrt{2H}\int_0^t(s-r)^{H-\frac{1}{2}}\,dW_r^2$ (again there is no need to introduce another component corresponding to S), which in particular is a martingale in this case. The option price C_t in (5.2) still takes the form (5.6), while (5.8) becomes

$$(5.20)\qquad dC_t = \partial_x u\big(t, S_t, \Theta^t_{[t,T]}\big)\, dS_t + \lambda\sqrt{2H}\big\langle \partial_\omega u\big(t, S_t, \Theta^t_{[t,T]}\big), a^t\big\rangle\, dW^2_t,$$

for the same a^t as in (5.7). Define the forward variance $\hat{\Theta}^t_s$ as in (5.3). By using the orthogonal decomposition of M as in Section 2, by (5.18) we can easily have

$$\hat{\Theta}^t_T = V_0 \exp\bigg(\Theta^t_T + \frac{1}{2}\lambda^2\big[(T-t)^{2H} - T^{2H}\big]\bigg).$$

By straightforward computation, we obtain

$$dW^2_t = \frac{(T-t)^{\frac{1}{2}-H}}{\lambda\sqrt{2H}}\, d\Theta^t_T = \frac{(T-t)^{\frac{1}{2}-H}}{\lambda\sqrt{2H}\hat{\Theta}^t_T}\, d\hat{\Theta}^t_T.$$

Plug this into (5.20), we have

$$(5.21)\quad dC_t = \partial_x u\big(t, S_t, \Theta^t_{[t,T]}\big)\, dS_t + \frac{(T-t)^{\frac{1}{2}-H}}{\hat{\Theta}^t_T}\big\langle \partial_\omega u\big(t, S_t, \Theta^t_{[t,T]}\big), a^t\big\rangle\, d\hat{\Theta}^t_T.$$

That is, we can replicate C_T by using S_t and $\hat{\Theta}^t_T$, with the corresponding hedging portfolios $\partial_x u$ and $\frac{(T-t)^{\frac{1}{2}-H}}{\hat{\Theta}^t_T}\langle \partial_\omega u, a^t\rangle$, respectively.

APPENDIX: INTEGRABILITY OF THE STATE PROCESS X

In this Appendix, we provide some sufficient conditions for Assumption 3.1(ii). We first remark that, by examining our proofs carefully, it is sufficient to assume that X has the p^*th moment for some finite p^* large enough; however, in that case we need to put corresponding constraints on the polynomial growth order κ in Definitions 3.3, 3.4 and 3.16, as well as the κ_0 in Assumptions 3.9 and 3.15. We also remark that, for many financial models like those we saw in the previous section, the dynamics of S is typically a semimartingale and is linear in S, and thus we have a representation like (5.19). Then we need much lower integrability for the S part, as in the standard literature.

The following result extends Abi Jaber, Larsson and Pulido [1], Lemma 3.1.

THEOREM A.1. *Let* (X, W) *be a weak solution to Volterra SDE* (3.1). *Assume, for* $\varphi = b, \sigma$, $\partial_t\varphi(t; s, \cdot)$ *exists for* $t \in (s, T]$, *and there exists* $0 < H < \frac{1}{2}$ *such that, for any* $0 \le s < t \le T$,

$$(A.1)\qquad \begin{aligned} \big|\varphi(t; s, \omega)\big| &\le C_0\big[1 + \|\omega\|_T\big](t-s)^{H-\frac{1}{2}},\\ \big|\partial_t\varphi(t; s, \omega)\big| &\le C_0\big[1 + \|\omega\|_T\big](t-s)^{H-\frac{3}{2}}. \end{aligned}$$

Then Assumption 3.1(ii) *holds true.*

PROOF. Fix $p \geq 2$. We first show that it suffices to prove a priori estimates by using the standard truncation arguments. Indeed, for any n, denote

$$\tau_n := \inf\{t : |X_T| \geq n\} \wedge T, \qquad X_t^n := X_{\tau_n \wedge t},$$
$$b^n(t; s, \omega) := b(t; s, \omega)\mathbf{1}_{\{\tau(\omega) \geq s\}}, \qquad \sigma^n(t; s, \omega) := \sigma(t; s, \omega)\mathbf{1}_{\{\tau(\omega) \geq s\}}.$$

Then (X^n, W) satisfies (3.1) with coefficients (b^n, σ^n) and (b^n, σ^n) satisfies (A.1) with the same constants H and C_0. Note that X^n is bounded, and we shall prove that there exists a constant C_p, independent of n, such that $\mathbb{E}[\|X^n\|_T^p] \leq C_p[1 + |x|^p]$. Then by sending $n \to \infty$, we prove the theorem.

We now assume X is bounded and prove in two steps that

$$\mathbb{E}[\|X\|^p] \leq C_p[1 + |x|^p]. \tag{A.2}$$

Step 1. Assume $T \leq \delta_0$, for some small $\delta_0 > 0$ which will be specified later. Then

$$\begin{aligned}
\mathbb{E}[|X_t - x|^p] &\leq C_p \mathbb{E}\left[\left|\int_0^t b(t; s, X.)\, ds\right|^p + \left|\int_0^t \sigma(t; s, X.)\, dW_s\right|^p\right] \\
&\leq C_p \mathbb{E}\left[\left[\int_0^t |b(t; s, X.)|\, ds\right]^p + \left[\int_0^t |\sigma(t; s, X.)|^2\, ds\right]^{\frac{p}{2}}\right] \\
&\leq C_p \mathbb{E}\left[\left[\int_0^t (t-s)^{2H-1} \|X\|_T^2\, ds\right]^{\frac{p}{2}}\right] \\
&\leq C_p \mathbb{E}[[t^{2H} \|X\|_T^2]^{\frac{p}{2}}] = C_p t^{pH} \mathbb{E}[\|X\|_T^p].
\end{aligned}$$

Next, for $0 \leq t_1 < t_2 \leq T$, denoting $\delta := t_2 - t_1$. When $t_1 \leq \delta$, we have

$$\begin{aligned}
\mathbb{E}[|X_{t_2} - X_{t_1}|^p] &\leq C_p \mathbb{E}[|X_{t_2} - x|^p + |X_{t_1} - x|^p] \\
&\leq C_p \mathbb{E}[\|X\|_T^p][t_2^{pH} + t_1^{pH}] \leq C_p \mathbb{E}[\|X\|_T^p]\delta^{pH}.
\end{aligned}$$

When $t_1 > \delta$, we have

$$X_{t_2} - X_{t_1} = I_1 + I_2,$$

where

$$\begin{aligned}
I_1 &:= \int_0^{t_1-\delta}\left[\int_{t_1}^{t_2} \partial_t b(t; s, X.)\, dt\, ds + \int_{t_1}^{t_2} \partial_t \sigma(t; s, X.)\, dt\, dW_s\right], \\
I_2 &:= \int_{t_1-\delta}^{t_2} [b(t_2; s, X.)\, ds + \sigma(t_2; s, X.)\, dW_s] \\
&\quad + \int_{t_1-\delta}^{t_1} [b(t_2; s, X.)\, ds + \sigma(t_2; s, X.)\, dW_s].
\end{aligned}$$

Note that

$$
\begin{aligned}
\mathbb{E}\big[|I_1|^p\big] &\le C_p\mathbb{E}\bigg[\bigg(\int_0^{t_1-\delta}\int_{t_1}^{t_2}\big|\partial_t b(t;s,X.)\big|\,dt\,ds\bigg)^p \\
&\qquad + \bigg(\int_0^{t_1-\delta}\bigg(\int_{t_1}^{t_2}\big|\partial_t\sigma(t;s,X.)\big|\,dt\bigg)^2 ds\bigg)^{\frac{p}{2}}\bigg] \\
&\le C_p\mathbb{E}\bigg[\bigg(\int_0^{t_1-\delta}\big(\delta(t_1-s)^{H-\frac{3}{2}}\big[1+\|X\|_T\big]\big)^2 ds\bigg)^{\frac{p}{2}}\bigg] \\
&\le C_p\mathbb{E}\big[1+\|X\|_T^p\big]\delta^{pH}; \\
\mathbb{E}\big[|I_2|^p\big] &\le C_p\sum_{i=1}^2\mathbb{E}\bigg[\bigg(\int_{t_1-\delta}^{t_i}\big|b(t_i;s,X.)\big|\,ds\bigg)^p + \bigg(\int_{t_1-\delta}^{t_i}\big|\sigma(t_i;s,X.)\big|^2 ds\bigg)^{\frac{p}{2}}\bigg] \\
&\le C_p\sum_{i=1}^2\mathbb{E}\bigg[\bigg(\int_{t_1-\delta}^{t_i}\big((t_i-s)^{H-\frac{1}{2}}\big[1+\|X\|_T\big]\big)^2 ds\bigg)^{\frac{p}{2}}\bigg] \\
&= C_p\sum_{i=1}^2\mathbb{E}\big[1+\|X\|_T^p\big]\big[t_i-(t_1-\delta)\big]^{pH} \le C_p\mathbb{E}\big[1+\|X\|_T^p\big]\delta^{pH}.
\end{aligned}
$$

Then

$$
\mathbb{E}\big[|X_{t_2}-X_{t_1}|^p\big] \le C_p\mathbb{E}\big[1+\|X\|_T^p\big]\delta^{pH}.
$$

This implies that

$$
\mathbb{E}\big[|X_{t_2}-X_{t_1}|^p\big] \le C_p\mathbb{E}\big[1+\|X\|_T^p\big]\delta_0^{\frac{pH}{2}}\delta^{\frac{pH}{2}}.
$$

By Lemma 3.11, we obtain

$$
\mathbb{E}\big[\|X\|_T^p\big] \le C_p\mathbb{E}\Big[|x|^p + \sup_{0\le t\le T}|X_t-x|^p\Big] \le C_p\mathbb{E}\big[|x|^p + \big[1+\|X\|_T^p\big]\delta_0^{\frac{pH}{2}}\big].
$$

By choosing δ_0 such that $C_p\delta_0^{\frac{pH}{2}} = \frac{1}{2}$, we obtain (A.2). We emphasize that δ_0 depends on p, H, but not on x.

Step 2. Now consider arbitrary T. Let δ_0 be as in Step 1, and denote $0 = t_0 < t_1 < \cdots < t_m = T$ be such that $\frac{\delta_0}{2} < t_{i+1} - t_i \le \delta_0$. Note that

$$
X_t = \Theta_t^{t_i} + \int_{t_i}^t b(t;s,X.)\,ds + \int_{t_i}^t \sigma(t;s,X.)\,dW_s,
$$

for $t_i \le t \le t_{i+1}$. Following the same arguments as in Step 1, we obtain

$$
\mathbb{E}\Big[\sup_{t_i\le t\le t_{i+1}}\big|X_t-\Theta_t^{t_i}\big|^p\Big] \le C_p\mathbb{E}\Big[1+\sup_{0\le t\le t_i}|X_t|^p\Big].
$$

Notice further that, for $t_i \le t < t+\delta \le t_{i+1}$,

$$\Theta^{t_i}_{t+\delta} - \Theta^{t_i}_t = \int_0^{t_i} \int_t^{t+\delta} \partial_t b(r; s, X.)\, dr\, ds + \int_0^{t_i} \int_t^{t+\delta} \partial_t \sigma(r; s, X.)\, dr\, dW_s.$$

Then

$$\begin{aligned} \mathbb{E}\big[|\Theta^{t_i}_{t+\delta} - \Theta^{t_i}_t|^p\big] &\le C_p \mathbb{E}\bigg[\bigg(\int_0^{t_i} \int_t^{t+\delta} |\partial_t b(r; s, X.)|\, dr\, ds\bigg)^p \\ &\qquad + \bigg(\int_0^{t_i} \bigg(\int_t^{t+\delta} |\partial_t \sigma(r; s, X.)|\, dr\bigg)^2 ds\bigg)^{\frac{p}{2}}\bigg] \\ &\le C_p \mathbb{E}\Big[1 + \sup_{0\le t\le t_i} |X_t|^p\Big] \bigg(\int_0^{t_i} \bigg(\int_t^{t+\delta} (r-s)^{H-\frac{3}{2}}\, dr\bigg)^2 ds\bigg)^{\frac{p}{2}}. \end{aligned}$$

Note that

$$\begin{aligned} &\int_0^{t_i} \bigg(\int_t^{t+\delta} (r-s)^{H-\frac{3}{2}}\, dr\bigg)^2 ds \\ &\quad \le \int_0^{t_i} \bigg(\int_{t_i}^{t_i+\delta} (r-s)^{H-\frac{3}{2}}\, dr\bigg)^2 ds \\ &\quad = \bigg[\int_0^{t_i - \delta^{\frac{2}{3}}} + \int_{t_i-\delta^{\frac{2}{3}}}^{t_i}\bigg] \bigg(\int_{t_i}^{t_i+\delta} (r-s)^{H-\frac{3}{2}}\, dr\bigg)^2 ds \\ &\quad \le \int_0^{t_i-\delta^{\frac{2}{3}}} \big[\delta(\delta^{\frac{2}{3}})^{H-\frac{3}{2}}\big]^2 ds + C \int_{t_i-\delta^{\frac{2}{3}}}^{t_i} (t_i - s)^{2H-1}\, ds \le C\delta^{\frac{4H}{3}}. \end{aligned}$$

Then

$$\mathbb{E}\big[|\Theta^{t_i}_{t+\delta} - \Theta^{t_i}_t|^p\big] \le C_p \mathbb{E}\Big[1 + \sup_{0\le t\le t_i} |X_t|^p\Big] \delta^{\frac{2pH}{3}}.$$

By Lemma 3.11, we have

$$\mathbb{E}\Big[\sup_{t_i\le t\le t_{i+1}} |\Theta^{t_i}_t|^p\Big] \le C_p \mathbb{E}\Big[1 + \sup_{0\le t\le t_i} |X_t|^p\Big].$$

Then

$$\mathbb{E}\Big[\sup_{t_i\le t\le t_{i+1}} |X_t|^p\Big] \le C_p \mathbb{E}\Big[1 + \sup_{0\le t\le t_i} |X_t|^p\Big].$$

Now by induction, one obtains (A.2) immediately. □

REMARK A.2. Consider the rough Heston model (5.1). By Theorem A.1, it is clear that $\mathbb{E}[\sup_{0\le t\le T} V_t^p] < \infty$ for all $p \ge 1$. However, we note that the coefficient of S does not grow linearly, and thus we cannot apply Theorem A.1 on

S directly. We shall instead utilize the representation formula (5.19). Note that $V \geq 0$, then

$$V_t \leq C + c \int_0^t (t-r)^{H-\frac{1}{2}} \sqrt{V_r}\, dW_r^2,$$

for some generic constants C, c. Thus

$$\begin{aligned} 0 \leq \int_0^t V_s\, ds &\leq C + c \int_0^t \int_0^s (s-r)^{H-\frac{1}{2}} \sqrt{V_r}\, dW_r^2\, ds \\ &= C + c \int_0^t (t-r)^{H+\frac{1}{2}} \sqrt{V_r}\, dW_r^2. \end{aligned}$$

Therefore, for any $n \geq 1$,

$$\begin{aligned} \mathbb{E}\bigg[\bigg(\int_0^t V_s\, ds\bigg)^n\bigg] &\leq C^n + C^n \mathbb{E}\bigg[\bigg(\int_0^t (t-r)^{2H+1} V_r dr\bigg)^{\frac{n}{2}}\bigg] \\ &\leq C^n + C^n \bigg(\mathbb{E}\bigg[\bigg(\int_0^t V_r\, dr\bigg)^n\bigg]\bigg)^{\frac{1}{2}}. \end{aligned}$$

This implies that

$$\mathbb{E}\bigg[\bigg(\int_0^t V_s\, ds\bigg)^n\bigg] \leq C^n \quad \text{and hence} \quad \mathbb{E}\bigg[\exp\bigg(p\int_0^t V_s\, ds\bigg)\bigg] \leq C_p < \infty.$$

Now by (5.19) we obtain immediately that $\mathbb{E}[S_t^p] \leq C_p < \infty$. Finally, again since S is a standard diffusion, applying Burkholder–Davis–Gundy inequality on the first equation in (5.1), we see that $\mathbb{E}[\sup_{0 \leq t \leq T} S_t^p] \leq C_p < \infty$.

REMARK A.3. For a rough volatility model, we may also denote the state process as $X = (\hat{S}, V)$ where $\hat{S} := \ln S$. Then, in the case of (5.1), we have

$$\hat{S}_t = \hat{S}_0 + \int_0^t \sqrt{V_r}\big[\sqrt{1-\rho^2}\, dW_r^1 + \rho\, dW_r^2\big] - \frac{1}{2}\int_0^t V_r\, dr.$$

This clearly satisfies (A.1), and thus $(\hat{S}, V)$ satisfies Assumption 3.1(ii). However, in this case we shall write $C_t = \hat{u}(t, \hat{S}_t, \Theta^t_{[t,T]})$, where $\hat{u}(t, x, \theta) := u(t, e^x, \theta)$. If it turns out that $\hat{u}$ has the desired polynomial growth in x (which in particular requires $\hat{g}(x) := g(e^x)$ has polynomial growth in x), then we may derive the required results by using $(\hat{S}, V)$. However, when g has linear growth, $\hat{g}$ would grow exponentially and then the integrability in Assumption 3.1(ii) will not be enough.

REMARK A.4. In this remark, we discuss the integrability for the rough Bergomi model in Remark 5.2. Let p^* denote the largest moment for S, as introduced by Lee [28]:

$$p^* := \sup\{p : \mathbb{E}[S_T^p] < \infty\}. \qquad \text{(A.3)}$$

(i) When $\rho = 0$, we have $p^* = 1$. For simplicity, let us assume $\lambda = V_0 = S_0 = 1$. Then

$$M_t = \sqrt{2H}\int_0^t (t-r)^{H-\frac{1}{2}}\,dW_r^2, \qquad V_t := e^{M_t - \frac{1}{2}t^{2H}},$$
$$S_t = e^{\int_0^t \sqrt{V_s}\,dW_s^1 - \frac{1}{2}\int_0^t V_s\,ds}.$$

In particular, V and W^1 are independent. Clearly,

$$\mathbb{E}[S_T] = \mathbb{E}\big[\mathbb{E}[S_T|\mathcal{F}_T^V]\big] = \mathbb{E}[1] = 1.$$

However, for any $p > 1$ and for some generic constant $c > 0$, denote $t_0 := \frac{T}{2}$, we have

$$\begin{aligned}
\mathbb{E}[S_T^p] &= \mathbb{E}\Big[\exp\Big(p\int_0^T \sqrt{V_s}\,dW_s^1 - \frac{p}{2}\int_0^T V_s\,ds\Big)\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(p\int_0^T \sqrt{V_s}\,dW_s^1 - \frac{p}{2}\int_0^T V_s\,ds\Big)\Big|\mathcal{F}_T^V\Big]\Big] \\
&= \mathbb{E}\Big[\exp\Big(\frac{p(p-1)}{2}\int_0^T V_s\,ds\Big)\Big] \ge \mathbb{E}\Big[\exp\Big(c\int_{t_0}^T V_s\,ds\Big)\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(c\int_{t_0}^T V_s\,ds\Big)\Big|\mathcal{F}_{t_0}^{W^2}\Big]\Big] \ge \mathbb{E}\Big[\exp\Big(c\int_{t_0}^T \mathbb{E}[V_s|\mathcal{F}_{t_0}^{W^2}]\,ds\Big)\Big],
\end{aligned}$$

where the last inequality is due to Jensen's inequality. Note that

$$\begin{aligned}
\mathbb{E}[V_s|\mathcal{F}_{t_0}^{W^2}] &\ge c\mathbb{E}[e^{M_s}|\mathcal{F}_{t_0}^{W^2}] \ge c\exp(\mathbb{E}[M_s|\mathcal{F}_{t_0}^{W^2}]) \\
&= c\exp\Big(c\int_0^{t_0}(s-r)^{H-\frac{1}{2}}\,dW_r^2\Big).
\end{aligned}$$

Then, by Jensen's inequality again,

$$\begin{aligned}
\int_{t_0}^T \mathbb{E}[V_s|\mathcal{F}_{t_0}^{W^2}]\,ds &= \frac{c}{T-t_0}\int_{t_0}^T \exp\Big(c\int_0^{t_0}(s-r)^{H-\frac{1}{2}}\,dW_r^2\Big)\,ds \\
&\ge c\exp\Big(\frac{c}{T-t_0}\int_{t_0}^T\Big(\int_0^{t_0}(s-r)^{H-\frac{1}{2}}\,dW_r^2\Big)\,ds\Big) \\
&= c\exp\Big(c\int_0^{t_0}\big[(T-r)^{H+\frac{1}{2}} - (t_0-r)^{H+\frac{1}{2}}\big]\,dW_r^2\Big).
\end{aligned}$$

Note that $\int_0^{t_0}[(T-r)^{H+\frac{1}{2}} - (t_0-r)^{H+\frac{1}{2}}]\,dW_r^2 \sim N(0,c_0)$ for some $c_0 > 0$. Then

$$\mathbb{E}[S_T^p] \ge c\mathbb{E}\big[\exp\big(ce^{cN(0,c_0)}\big)\big] = \infty.$$

(ii) The situation could be worse if $\rho \neq 0$. Assume for simplicity that $H = \frac{1}{2}$ and $\lambda = V_0 = S_0 = 1$. Then

$$V_t = \exp\left(W_t^2 - \frac{t}{2}\right),$$
$$S_t = \exp\left(\int_0^t \sqrt{V_s}\left[\sqrt{1-\rho^2}\, dW_s^1 + \rho\, dW_s^2\right] - \frac{1}{2}\int_0^t V_s\, ds\right).$$

Thus

$$\begin{aligned}
\mathbb{E}[S_T] &= \mathbb{E}[\mathbb{E}[S_T | \mathcal{F}_T^{W^2}]] \\
&= \mathbb{E}\left[\exp\left(\int_0^T \sqrt{V_s}\rho\, dW_s^2 - \frac{1}{2}\int_0^T V_s\, ds\right)\exp\left(\frac{1-\rho^2}{2}\int_0^T V_s\, ds\right)\right] \\
&= \mathbb{E}\left[\exp\left(\rho\int_0^t \sqrt{V_s}\, dW_s^2 - \frac{\rho^2}{2}\int_0^t V_s\, ds\right)\right] \\
&= \mathbb{E}\left[\exp\left(\rho\int_0^t e^{\frac{1}{2}W_s^2 - \frac{s}{4}}\, dW_s^2 - \frac{\rho^2}{2}\int_0^t e^{W_s^2 - \frac{s}{2}}\, ds\right)\right].
\end{aligned}$$

This is in the framework of the Girsanov theorem, but the drift $\rho e^{\frac{1}{2}W_s^2 - \frac{s}{4}}$ has exponential growth. While we do not have a rigorous proof, we suspect that the above integral is strictly less than 1, and then S would be a strict local martingale.

REFERENCES

[1] ABI JABER, E., LARSSON, M. and PULIDO, S. (2017). Affine Volterra processes. Preprint. Available at arXiv:1708.08796.

[2] BAUDOIN, F. and COUTIN, L. (2007). Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stochastic Process. Appl.* **117** 550–574. MR2320949

[3] BAYER, C., FRIZ, P. and GATHERAL, J. (2016). Pricing under rough volatility. *Quant. Finance* **16** 887–904. MR3494612

[4] BENNEDSEN, M., LUNDE, A. and PAKKANEN, M. (2016). Decoupling the short- and long-term behavior of stochastic volatility. Preprint. Available at arXiv:1610.00332.

[5] BERGER, M. A. and MIZEL, V. J. (1980). Volterra equations with Itô integrals. I. *J. Integral Equations* **2** 187–245. MR0581430

[6] BERGER, M. A. and MIZEL, V. J. (1980). Volterra equations with Itô integrals. II. *J. Integral Equations* **2** 319–337. MR0596459

[7] CARR, P. and MADAN, D. (2001). Towards a theory of volatility trading. In *Option Pricing, Interest Rates and Risk Management. Handb. Math. Finance* 458–476. Cambridge Univ. Press, Cambridge. MR1848559

[8] CHRONOPOULOU, A. and VIENS, F. G. (2012). Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quant. Finance* **12** 635–649. MR2909603

[9] COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8** 291–323. MR1645101

[10] CONT, R. and FOURNIÉ, D.-A. (2010). Change of variable formulas for non-anticipative functionals on path space. *J. Funct. Anal.* **259** 1043–1072. MR2652181

[11] CONT, R. and FOURNIÉ, D.-A. (2013). Functional Itô calculus and stochastic integral representation of martingales. *Ann. Probab.* **41** 109–133. MR3059194

[12] CONT, R. and FOURNIÉ, D. A. (2016). Functional Itô calculus and functional Kolmogorov equations. In *Stochastic Integration by Parts and Functional Itô Calculus. Adv. Courses Math. CRM Barcelona* 115–207. Birkhäuser/Springer, Cham. MR3497715

[13] COSSO, A. and RUSSO, F. (2019). Strong-viscosity solutions: Semilinear parabolic PDEs and path-dependent PDEs. *Osaka J. Math.* **56** 323–373. MR3934979

[14] CUCHIERO, C. and TEICHMANN, J. (2018). Generalized Feller processes and Markovian lifts of stochastic Volterra processes: The affine case. Preprint. Available at arXiv:1804.10450.

[15] DECREUSEFOND, L. and ÜSTÜNEL, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential Anal.* **10** 177–214. MR1677455

[16] DUPIRE, B. (2019). Functional Itô calculus. *Quant. Finance* **19** 721–729. MR3939653

[17] EKREN, I., KELLER, C., TOUZI, N. and ZHANG, J. (2014). On viscosity solutions of path dependent PDEs. *Ann. Probab.* **42** 204–236. MR3161485

[18] EKREN, I., TOUZI, N. and ZHANG, J. (2016). Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part I. *Ann. Probab.* **44** 1212–1253. MR3474470

[19] EKREN, I., TOUZI, N. and ZHANG, J. (2016). Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II. *Ann. Probab.* **44** 2507–2553. MR3531674

[20] EL EUCH, O. and ROSENBAUM, M. (2018). Perfect hedging in rough Heston models. *Ann. Appl. Probab.* **28** 3813–3856. MR3861827

[21] EL EUCH, O. and ROSENBAUM, M. (2019). The characteristic function of rough Heston models. *Math. Finance* **29** 3–38. MR3905737

[22] FOUQUE, J.-P. and HU, R. (2018). Optimal portfolio under fast mean-reverting fractional stochastic environment. *SIAM J. Financial Math.* **9** 564–601. MR3797731

[23] GATHERAL, J. (2008). Consistent modeling of SPX and VIX options. In *Fifth World Congress of the Bachelier Finance Society*.

[24] GATHERAL, J., JAISSON, T. and ROSENBAUM, M. (2018). Volatility is rough. *Quant. Finance* **18** 933–949. MR3805308

[25] GATHERAL, J. and KELLER-RESSEL, M. (2018). Affine forward variance models. Preprint. Available at arXiv:1801.06416.

[26] GULISASHVILI, A., VIENS, F. and ZHANG, X. (2018). Small-time asymptotics for Gaussian self-similar stochastic volatility models. *Appl. Math. Optim.* DOI:10.1007/s00245-018-9497-6.

[27] KELLER, C. (2016). Viscosity solutions of path-dependent integro-differential equations. *Stochastic Process. Appl.* **126** 2665–2718. MR3522297

[28] LEE, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Math. Finance* **14** 469–480. MR2070174

[29] LUKOYANOV, N. Y. (2007). On viscosity solution of functional Hamilton–Jacobi type equations for hereditary systems. *Proc. Steklov Inst. Math. Suppl.* **2** 190–200.

[30] MOCIOALCA, O. and VIENS, F. (2005). Skorohod integration and stochastic calculus beyond the fractional Brownian scale. *J. Funct. Anal.* **222** 385–434. MR2132395

[31] NUALART, D. (2006). Malliavin calculus and related topics. In *Stochastic Processes and Related Topics*, 2nd ed. *Math. Res.* **61** 103–127. Springer, Berlin. MR1127886

[32] PARDOUX, É. and PENG, S. G. (1990). Adapted solution of a backward stochastic differential equation. *Systems Control Lett.* **14** 55–61. MR1037747

[33] PENG, S. (2010). Backward stochastic differential equation, nonlinear expectation and their applications. In *Proceedings of the International Congress of Mathematicians, Vol. I* 393–432. Hindustan Book Agency, New Delhi. MR2827899

[34] PENG, S. and SONG, Y. (2015). G-expectation weighted Sobolev spaces, backward SDE and path dependent PDE. *J. Math. Soc. Japan* **67** 1725–1757. MR3417511

[35] PENG, S. and WANG, F. (2016). BSDE, path-dependent PDE and nonlinear Feynman–Kac formula. *Sci. China Math.* **59** 19–36. MR3436993
[36] PENG, S. G. (1991). Probabilistic interpretation for systems of quasilinear parabolic partial differential equations. *Stoch. Stoch. Rep.* **37** 61–74. MR1149116
[37] REN, Z., TOUZI, N. and ZHANG, J. (2017). Comparison of viscosity solutions of fully nonlinear degenerate parabolic path-dependent PDEs. *SIAM J. Math. Anal.* **49** 4093–4116. MR3715377
[38] REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1083357
[39] YONG, J. (2013). Backward stochastic Volterra integral equations—A brief survey. *Appl. Math. J. Chinese Univ. Ser. B* **28** 383–394. MR3143894
[40] ZHANG, J. (2017). *Backward Stochastic Differential Equations*: *From Linear to Fully Nonlinear Theory*. *Probability Theory and Stochastic Modelling* **86**. Springer, New York. MR3699487

DEPARTMENT OF STATISTICS
AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: viens@msu.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089
USA
E-MAIL: jianfenz@usc.edu