



UNIVERSITY OF LEEDS

This is a repository copy of *Forward–backward SDEs with distributional coefficients*.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/140760/>

Version: Accepted Version

Article:

Issoglio, E orcid.org/0000-0003-3035-2712 and Jing, S (2020) Forward–backward SDEs with distributional coefficients. *Stochastic Processes and their Applications*, 130 (1). pp. 47-78. ISSN 0304-4149

<https://doi.org/10.1016/j.spa.2019.01.001>

© 2019 Elsevier B.V. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

FORWARD-BACKWARD SDES WITH DISTRIBUTIONAL COEFFICIENTS

ELENA ISSOGLIO^{1,*} AND SHUAI JING^{2,°}

ABSTRACT. Forward-backward stochastic differential equations (FB-SDEs) have attracted significant attention since they were introduced, due to their wide range of applications, from solving non-linear PDEs to pricing American-type options. Here, we consider two new classes of multidimensional FBSDEs with distributional coefficients (elements of a Sobolev space with negative order). We introduce a suitable notion of solution and show its existence and in certain cases its uniqueness. Moreover we establish a link with PDE theory via a non-linear Feynman-Kac formula. The associated semi-linear parabolic PDE is the same for both FBSDEs, also involves distributional coefficients and has not previously been investigated.

1. INTRODUCTION

In this paper we study systems of multidimensional forward-backward stochastic differential equations (forward-backward SDEs or FBSDEs for shortness) with generalized coefficients. In particular, we consider a class of coefficients b which are elements of the space $L^\infty([0, T], H_q^{-\beta})$ for some $\beta \in (0, 1/2)$, where $H_q^{-\beta}$ is a fractional Sobolev space of negative derivation order, hence its elements are distributions (see Section 2 for its definition). We consider two different systems of *FBSDEs with distributional coefficients*, both decoupled so that the forward equation can be solved first and the solution plugged into the backward equation.

¹DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LEEDS, LEEDS, LS2 9JT, UK. E.ISSOGLIO@LEEDS.AC.UK

²DEPARTMENT OF MANAGEMENT SCIENCE, CENTRAL UNIVERSITY OF FINANCE AND ECONOMICS, BEIJING, 100081, CHINA. JING@CUFE.EDU.CN

[°]PARTIALLY FUNDED BY NSFC NO.71772195, 51609270.

*CORRESPONDING AUTHOR

2010 *Mathematics Subject Classification*. Primary 60H10; Secondary 35K55, 60H30, 35K10.

Key words and phrases. Forward-backward stochastic differential equations, distributional coefficients, non-linear Feynman-Kac formula, weak solutions, virtual solutions, mild solutions, Sobolev spaces, singular FBSDEs, singular PDEs.

In the first system, the distribution b appears in the driver of the backward equation as follows

$$(1) \quad \begin{cases} X_s^{t,x} = x + \int_t^s dW_r, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) - \int_s^T Z_r^{t,x} dW_r + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr \\ \quad + \int_s^T Z_r^{t,x} b(r, X_r^{t,x}) dr, \\ \forall s \in [t, T], \end{cases}$$

where W is a d -dimensional Brownian motion, Φ and f are functions with standard regularity properties which will be specified later, and the processes X, Y, Z are d, m and $m \times d$ -dimensional, respectively.

In the second system, the distribution appears in the forward equation as follows

$$(2) \quad \begin{cases} X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x}) dr + \int_t^s dW_r, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) - \int_s^T Z_r^{t,x} dW_r + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr, \\ \forall s \in [t, T]. \end{cases}$$

The two systems are studied independently. We give a meaning to the integral terms $\int_s^T Z_r^{t,x} b(r, X_r^{t,x}) dr$ and $\int_t^s b(r, X_r^{t,x}) dr$ by introducing a suitable notion of solution for the systems (1) and (2), and then investigate their existence and uniqueness. Moreover we look at the associated PDE and show its link with the FBSDEs (the well known *non-linear Feynman-Kac formula*). As one might expect, it turns out that the PDE associated to both systems (1) and (2) is the same, and it is a semi-linear equation of the form

$$(3) \quad \begin{cases} u_t(t, x) + L^b u(t, x) + f(t, x, u(t, x), \nabla u(t, x)) = 0, \\ u(T, x) = \Phi(x), \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d, \end{cases}$$

where the operator $L^b u := \frac{1}{2} \Delta u + \nabla u b$ is defined component by component (see Section 3). This PDE also involves distributional coefficients, in particular the drift b which is multiplied by ∇u . A thorough investigation of the partial differential equation is carried out.

Literature review. The history of FBSDEs dates back to 1990, when the foundational paper of Pardoux and Peng [27] appeared. In 1992 the same authors established the link between (decoupled) FBSDEs and quasi-linear PDEs, well-known as the non-linear Feynman-Kac formula [28]. A year later, Antonelli [1] studied for the first time fully coupled FBSDEs in a small time interval. Since then, the theory of BSDEs and of FBSDEs received a lot of attention by the mathematical community and found many applications in different fields, especially in finance. For more details on the latter we refer to the paper of El Karoui et al. [9] and references therein.

The above-mentioned literature and many subsequent papers were concerned with strong solutions, but starting from the early 2000s mathematicians introduced and studied the notion of weak solution for

FBSDEs. Weak solutions are analogous to weak solutions for SDEs, and their importance is illustrated by a series of stochastic differential equations which admit a weak solution but for which no strong solution exists. For example we mention the well-known Tsirel'son's stochastic differential equation introduced in 1975 by Tsirel'son [33], or the so-called sticky Brownian motion, which was recently studied by Engelbert and Peskir [10]. Antonelli and Ma [2] first proposed the notion of weak solutions for FBSDEs in 2003. A more general notion of weak solution was studied later by Buckdahn et al. [5] in 2004, where the equation for the forward component was implicitly given, and its existence without the uniqueness was discussed. Lejay [24] in 2004 studied existence of weak solutions by using the link between FBSDEs and weak and mild solutions of PDEs. Delarue and Guatteri [7] in 2006 were the first to establish uniqueness of weak solutions for fully coupled Markovian FBSDEs. In their paper, the coefficients for the backward equation are Lipschitz, hence the “weak” notion essentially only intervenes in the forward equation. In 2008 Ma et al. [25] also studied existence and uniqueness of weak solutions but in a more general framework, and in fact there the “weak” character appears both in the forward and in the backward equation.

The literature on FBSDEs is large but to our knowledge there is very little about (forward-)backward equations with generalized functions (Schwartz distributions). In 1997-1998, Erraoui, Ouknine and Sbi [11, 12] studied (reflected) BSDEs with distribution as terminal condition. By applying the stochastic flow method, Bally and Matoussi [3] in 2001 studied stochastic PDE with terminal values and coefficients being distributions using Backward Doubly SDEs. In 2007, Hu and Tessitore [17] studied mild solutions of elliptic PDEs in Hilbert spaces by proving the regularity properties of a bounded solution of a BSDE with infinite horizon. Recently, Russo and Wurzer [31] studied a one-dimensional BSDE indirectly involving distributional coefficients: They consider and solve a semilinear ODE with a distributional drift and study the associated one-dimensional martingale problem. The martingales are then used to construct the solution of a martingale-driven BSDE with random terminal time. We also cite the recent results of Diehl and Zhang [8] where the authors deal with BSDEs with Young integrals.

Motivation. The importance of classical results on FBSDEs and their link to PDEs through the generalized Feynman-Kac formula is well known. In our case, we relax notably the assumptions on the coefficients of Markovian systems of FBSDEs to allow for generalized functions, and investigate what kind of solutions one can expect in that case. Once a generalised Feynman-Kac formula is obtained in the irregular/distributional case, then new tools and methods can be used to investigate irregular physical phenomena described by (S)PDEs with

distributional coefficients. In particular, PDEs like (3) with irregular fields b have been considered as models of transport of passive scalars in turbulent fluids (like the Kraichnan model [22]). In recent years the Kraichnan model has been researched by physicists also when the velocity field is a stochastic process, see e.g. [26] or [14] and references therein. An example of b that we can treat in this paper is the formal gradient of the realization of some random field (like fractional Brownian noise cut at infinity, but one could consider also other fields not necessarily Gaussian so long as their realizations are α -Hölder continuous with $\alpha > 1/2$).

In this paper we are indeed able to derive a Feynman-Kac formula that links the PDE (3) with the forward-backward equations (1) and (2), but our starting point is the solution of the PDE. Hence we use our knowledge on the PDE to infer results on the FBSDE. This is only partially satisfactory if one argues that using FBSDEs to solve PDEs is more interesting than the vice versa, but nevertheless the link provides new stochastic tools to represent and study such turbulent PDEs. For example numerical methods to solve FBSDEs could be employed to find the numerical solution of the PDEs using the Feynman-Kac formula illustrated in this paper. Indeed there is a line of research that exploits this connection and uses numerical solutions of BSDEs to infer solutions of PDEs (for a recent work on this see e.g. [21]).

Novelty and main results. The present paper is the first to deal with FBSDEs like (1) with distributional coefficients appearing in the driver, both in the one-dimensional and in the multidimensional case. Because of the lack of literature on this topic, the first challenge we face is to define a suitable notion of solution for the backward component of the FBSDE (see Definition 12 of virtual-strong solution). Once this is done, the next challenge is to investigate existence and uniqueness of the solution. To do so, we introduce a transformation –which in some sense can be regarded as the analogous for BSDEs of a Zvonkin transformation for SDEs– and rewrite the original BSDE as an auxiliary backward SDE which can be treated with classical methods, see equation (26). For the auxiliary BSDE it is then possible to show *existence and uniqueness of a strong solution*, which leads to the same result for the original BSDE (1) by transforming back the equation, see Theorem 15. It is worth stressing the fact that the solution we find is a strong type of solution (and not weak, i.e. not of martingale type like in [31]). This is possible in the first place because the forward equation here is a Brownian motion and not a solution of a martingale problem.

The second main result in this paper is a *non-linear Feynman-Kac representation formula* that links the PDE (3) and the FBSDE (1) (see Theorem 18 and Theorem 19). To show this, we consider smooth approximations of b and related solutions to the FBSDE and the PDE,

and then take the limit. This requires various uniform bounds on the smoothed solutions of the PDE (3) and of auxiliary PDE (23) (see Sections 3 and 4.2). Indeed the study of PDE (3) is crucial in this paper because its solution is used to define *virtual solutions* for both FBSDEs systems (1) and (2), as illustrated in Definition 12 and Definition 24. We solve the semi-linear PDE (3) by looking for mild solutions using a fixed-point argument. This is the same idea applied in [13, 18] where linear PDEs of transport-diffusion type with distributional coefficients analogous to b have been studied. The novelty here is the non linear term f , and for this we require Lipschitz continuity properties. Moreover there is a delicate issue about f that we want to mention at this point, namely the need to match the two set-ups in which the PDE and the FBSDE naturally live, which clearly reflects on the assumptions on the coefficients. The former (PDE) is solved as an infinite-dimensional equation, in particular the solution as a function of time takes values in a Sobolev space and so the Lipschitz continuity required for the non-linearity f must be set up in terms of Sobolev spaces (see Assumption 2). The latter (FBSDE) is set-up in \mathbb{R}^d and thus assumptions on the coefficients (including f) cannot be made in the Sobolev space, but are written in \mathbb{R}^d instead (see Assumption 1). Thus some care is needed to match the two settings and this is explained in Remark 3.

The final main result is about the FBSDE (2). This system is, in some sense, the generalization to multi dimensions of the BSDE studied in [31], but with deterministic terminal time. The system is decoupled and the forward equation is solved first. Here we study the forward equation with different techniques than in [31], in particular we invoke the results found in [13] about SDEs with distributional coefficients which can be applied to the forward component of (2). The forward solution $X^{t,x}$ is then used to find a virtual-weak solution $(X^{t,x}, Y^{t,x}, Z^{t,x})$ to the FBSDE (2), see Theorem 25. The virtual-weak solution is constructed using the mild solution u of the PDE (3) and is given in terms of a Feynman-Kac representation $(X^{t,x}, Y^{t,x}, Z^{t,x}) = (X^{t,x}, u(\cdot, X^{t,x}), \nabla u(\cdot, X^{t,x}))$.

For system (2) we do not find strong solutions but only weak solutions, because the solution of the forward equation is of weak type. We refer the reader to Section 5.1 for some extended and heuristic comments on the link between (1) and (2), and for open questions.

Organization of the paper. The paper is organised as follows: In Section 2 we introduce the notation and recall some useful results; In Section 3 we study the PDE (3) and find a unique mild solution with related smoothness properties. In Section 4 we introduce the notion of virtual-strong solution for backward SDE (1) and show that a unique virtual-strong solution exists. Moreover we establish the non-linear Feynman-Kac formula for (3) and (1). Finally in Section 5 we recall the

notion of virtual solution for the forward SDE in (2), we show existence of a virtual-weak solution to (2) by providing an explicit representation by means of a non-linear Feynman-Kac formula.

Throughout the paper the constants C and c can vary from line to line.

2. PRELIMINARIES

Here we recall some known facts, for more details see [13, Section 2.1] and references therein.

Let $(P(t), t \geq 0)$ be the heat semigroup on the space of \mathbb{R}^d -valued Schwartz functions $\mathcal{S}(\mathbb{R}^d)$ generated by $\frac{1}{2}\Delta$, that is the semigroup with kernel $p_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x|^2}{2t}\right)$, where $|\cdot|$ denotes the Euclidean norm in \mathbb{R}^d . The semigroup extends to the space of Schwartz distributions $\mathcal{S}'(\mathbb{R}^d)$ by duality, and in particular it maps any $L^p(\mathbb{R}^d)$ into itself for $1 < p < \infty$. This restriction to $L^p(\mathbb{R}^d)$, denoted by $(P_p(t), t \geq 0)$, is a bounded analytic semigroup (see [6, Theorems 1.4.1, 1.4.2]). Let $A_p := I - \frac{1}{2}\Delta$, then $-A_p$ also generates a bounded analytic semigroup which is given by $e^{-t}P_p(t)$ (i.e. with kernel $e^{-t}p_t(x)$). We can define fractional Sobolev spaces as images of fractional powers of A_p (which are well defined for any power $s \in \mathbb{R}$, see [29]) by $H_p^s(\mathbb{R}^d) := A_p^{-s/2}(L^p(\mathbb{R}^d))$. These are Banach spaces endowed with the norm $\|u\|_{H_p^s} := \|A_p^{s/2}u\|_{L^p}$. It turns out that these spaces correspond to the domain of fractional powers of A_p and of $-\frac{1}{2}\Delta$, that is $D(A_p^{s/2}) = D((-\frac{1}{2}\Delta)^{s/2}) = H_p^s(\mathbb{R}^d)$. Moreover $A_p^{-\alpha/2}$ is an isomorphism between $H_p^s(\mathbb{R}^d)$ and $H_p^{s+\alpha}(\mathbb{R}^d)$, for each $\alpha \in \mathbb{R}$. $H_p^s(\mathbb{R}^d; \mathbb{R}^n)$ are defined as above for each component. For shortness of notation we will sometimes denote them simply by H_p^s (note that the dimension n could be d, m or $m \times d$ depending on the context). When we write $u \in H_p^s$ we mean that each component u_i is in $H_p^s(\mathbb{R}^d)$. The norm will be denoted with the same notation for simplicity. One can also show that $\nabla : H_p^{1+\delta} \rightarrow H_p^\delta$ is a continuous map, so if $u \in H_p^{1+\delta}$ then $\|\nabla u\|_{H_p^\delta} \leq c\|u\|_{H_p^{1+\delta}}$ for some positive constant c .

The semigroup $(P_p(t), t \geq 0)$ is a contraction on the $H_p^s(\mathbb{R}^d)$ -spaces for all $t > 0$ and all $s \in \mathbb{R}$ and moreover it enjoys the following mapping property: for $\delta > \beta \geq 0, \delta + \beta < 1$ and $0 < t \leq T$ it holds $P_p(t) : H_p^{-\beta}(\mathbb{R}^d) \rightarrow H_p^{1+\delta}(\mathbb{R}^d)$, in particular we have

$$(4) \quad \|P_p(t)w\|_{H_p^{1+\delta}(\mathbb{R}^d)} \leq Ct^{-\frac{1+\delta+\beta}{2}} \|w\|_{H_p^{-\beta}(\mathbb{R}^d)}$$

for $w \in H_p^{-\beta}(\mathbb{R}^d), t > 0$, where $C = ce^T$ for some positive constant c . This follows from a similar property for the semigroup $(e^{-t}P_p(t), t \geq 0)$ which is stated in [13, Lemma 10], see also [18, Proposition 3.2] for the analogous on domains $D \subset \mathbb{R}^d$.

Here we recall the definition of the *pointwise product* between a function and a distribution (see [30]) as we will use it several times in this paper. Let $g \in \mathcal{S}'(\mathbb{R}^d)$. We choose a function $\psi \in \mathcal{S}(\mathbb{R}^d)$ such that $0 \leq \psi(x) \leq 1$, for every $x \in \mathbb{R}^d$ and

$$\psi(x) = \begin{cases} 1, & |x| < 1, \\ 0, & |x| \geq \frac{3}{2}. \end{cases}$$

For every $j \in \mathbb{N}$, we consider the approximation $S^j g$ of g as follows:

$$S^j g(x) := \mathcal{F}^{-1} \left(\psi \left(\frac{\xi}{2^j} \right) \mathcal{F}(g) \right) (x),$$

where $\mathcal{F}(g)$ and $\mathcal{F}^{-1}(g)$ are the Fourier transform and the inverse Fourier transform of g , respectively. The product gh of $g, h \in \mathcal{S}'(\mathbb{R}^d)$ is defined as

$$(5) \quad gh := \lim_{j \rightarrow \infty} S^j g S^j h,$$

if the limit exists in $\mathcal{S}'(\mathbb{R}^d)$. The convergence of the limit (5) in the case we are interested in is given by the following result (for a proof see [30, Theorem 4.4.3/1]).

Lemma 1. *Let $g \in H_q^{-\beta}(\mathbb{R}^d)$, $h \in H_p^\delta(\mathbb{R}^d)$ for $1 < p, q < \infty$, $q > \max(p, \frac{d}{\delta})$, $0 < \beta < \frac{1}{2}$ and $\beta < \delta$. Then the pointwise product gh is well defined, it belongs to the space $H_p^{-\beta}(\mathbb{R}^d)$ and we have the following bound*

$$\|gh\|_{H_p^{-\beta}(\mathbb{R}^d)} \leq c \|g\|_{H_q^{-\beta}(\mathbb{R}^d)} \cdot \|h\|_{H_p^\delta(\mathbb{R}^d)}.$$

For the following, see [32, Section 2.7.1]. The closures of \mathcal{S} with respect to the norms

$$\|h\|_{C^{0,0}} := \|h\|_{L^\infty}$$

and

$$\|h\|_{C^{1,0}} := \|h\|_{L^\infty} + \|\nabla h\|_{L^\infty}$$

respectively, are denoted by $C^{0,0}(\mathbb{R}^d; \mathbb{R}^m)$ and $C^{1,0}(\mathbb{R}^d; \mathbb{R}^m)$. For any $\alpha > 0$, we consider the Banach spaces

$$C^{0,\alpha} = \{h \in C^{0,0}(\mathbb{R}^d; \mathbb{R}^m) : \|h\|_{C^{0,\alpha}} < \infty\}$$

$$C^{1,\alpha} = \{h \in C^{1,0}(\mathbb{R}^d; \mathbb{R}^m) : \|h\|_{C^{1,\alpha}} < \infty\},$$

endowed with the norms

$$\|h\|_{C^{0,\alpha}} := \|h\|_{L^\infty} + \sup_{x \neq y \in \mathbb{R}^d} \frac{|h(x) - h(y)|}{|x - y|^\alpha}$$

$$\|h\|_{C^{1,\alpha}} := \|h\|_{L^\infty} + \|\nabla h\|_{L^\infty} + \sup_{x \neq y \in \mathbb{R}^d} \frac{|\nabla h(x) - \nabla h(y)|}{|x - y|^\alpha},$$

respectively.

Let B be a Banach space. We denote by $C^{0,\alpha}([0, T]; B)$ the space analogous to $C^{0,\alpha}$ but with values in B , and its norm by $\|\cdot\|_{C^{0,\alpha}([0, T]; B)}$.

We denote by $C([0, T]; B)$ the Banach space of B -valued continuous functions and its sup norm by $\|\cdot\|_{\infty, B}$. For $h \in C([0, T], B)$, we also use the family of equivalent norms $\{\|\cdot\|_{\infty, B}^{(\rho)}, \rho \geq 1\}$, defined by

$$\|h\|_{\infty, B}^{(\rho)} := \sup_{0 \leq t \leq T} e^{-\rho t} \|h(t)\|_B.$$

The usual esssup norm on $L^\infty(0, T; B)$ will also be denoted by $\|\cdot\|_{\infty, B}$ with a slight abuse of notation. The Euclidean norm in \mathbb{R} , \mathbb{R}^d , \mathbb{R}^m , and the Frobenius norm in $\mathbb{R}^{m \times d}$ will be denoted by $|\cdot|$.

The following lemma provides a generalization of the Morrey inequality to fractional Sobolev spaces. For the proof we refer to [32, Theorem 2.8.1, Remark 2].

Lemma 2 (Fractional Morrey inequality). *Let $0 < \delta < 1$ and $d/\delta < p < \infty$. If $h \in H_p^{1+\delta}(\mathbb{R}^d)$ then there exists a unique version of h (which we denote again by h) such that h is differentiable. Moreover $h \in C^{1,\alpha}(\mathbb{R}^d)$ with $\alpha = \delta - d/p$ and*

$$(6) \quad \|h\|_{C^{1,\alpha}} \leq c \|h\|_{H_p^{1+\delta}}, \quad \|\nabla h\|_{C^{0,\alpha}} \leq c \|\nabla h\|_{H_p^\delta},$$

where $c = c(\delta, p, d)$ is a universal constant.

Standing Assumption: Throughout the paper we will make the following standing assumption about the drift b and in particular about the parameters involved. We acknowledge that the set $K(\beta, q)$ is taken from [13].

Let $\beta \in (0, \frac{1}{2})$, $q \in (\frac{d}{1-\beta}, \frac{d}{\beta})$. Let the drift b be of the type

$$b \in L^\infty([0, T]; H_q^{-\beta}(\mathbb{R}^d; \mathbb{R}^d)).$$

Moreover for given β and q as above we define the set

$$(7) \quad K(\beta, q) := \left\{ \kappa = (\delta, p) : \beta < \delta < 1 - \beta, \frac{d}{\delta} < p < q \right\}.$$

We always choose $(\delta, p) \in K(\beta, q)$. Note that $K(\beta, q)$ is non-empty since $\beta < \frac{1}{2}$ and $\frac{d}{1-\beta} < q < \frac{d}{\beta}$.

Regarding the functions f and Φ , we make the following parallel sets of assumptions. This is because the PDE is set (and solved) using fractional Sobolev spaces, whereas the BSDE is typically set in \mathbb{R}^d . We discuss the link and implications of these two sets of Assumptions in Remark 3 below. Afterwards, we also give examples of possible f . Note that the notation for f is the same, even though the function is in principle different in the two sets of assumptions.

Assumption 1.

- $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ is such that $\Phi \in H_p^{1+\delta+2\gamma}$ for some $\gamma < \frac{1-\delta-\beta}{2}$;
- $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^m$ is continuous in (x, y, z) uniformly in t , and Lipschitz continuous in (y, z) uniformly in

t and x , i.e. $|f(t, x, y, z) - f(t, x, y', z')| \leq L(|y - y'| + |z - z'|)$
for any $y, y' \in \mathbb{R}^m$ and $z, z' \in \mathbb{R}^{m \times d}$;

- $\sup_{t,x} |f(t, x, 0, 0)| \leq C$ and $\sup_{t \in [0, T]} \int_{\mathbb{R}^d} |f(t, x, 0, 0)|^p dx \leq C$.

Assumption 2.

- $\Phi \in H_p^{1+\delta+2\gamma}(\mathbb{R}^d; \mathbb{R}^m)$ for some $\gamma < \frac{1-\delta-\beta}{2}$;
- $f : [0, T] \times H_p^{1+\delta}(\mathbb{R}^d; \mathbb{R}^m) \times H_p^\delta(\mathbb{R}^d; \mathbb{R}^{m \times d}) \rightarrow H_p^0(\mathbb{R}^d; \mathbb{R}^m)$ is Lipschitz continuous in the second and third variable uniformly in t , that is, there exists a positive constant L such that for any $u_1, u_2 \in H_p^{1+\delta}$ and $v_1, v_2 \in H_p^\delta$ then

$$\|f(t, u_1, v_1) - f(t, u_2, v_2)\|_{H_p^0} \leq L \left(\|u_1 - u_2\|_{H_p^{1+\delta}} + \|v_1 - v_2\|_{H_p^\delta} \right);$$

- $\sup_{t,x} |f(t, x, 0, 0)| \leq C$ and $\sup_{t \in [0, T]} \|f(t, 0, 0)\|_{H_p^0} \leq C$, where 0 here denotes the constant zero function.

Notation: In Assumption 2 the functional f is a function of time t and of two other functions, often denoted by u and v (or u and ∇u). In this paper we write $f(t, u, v)$, or $f(t, \cdot, u, v)$, or also $f(t, \cdot, u(\cdot), v(\cdot))$, and this is an element of the space $H_p^{1+\delta}$ by Assumption 2.

Remark 3.

- By applying the Fractional Morrey inequality we see that $\Phi \in C^{1,\alpha}$ with $\alpha = \delta + 2\gamma - \frac{d}{p} > 0$. This implies in particular that Φ is bounded and continuous. Note that the latter would be the standard assumption on the terminal condition Φ when solving the BSDE, but our setting to solve the PDE requires that Φ is an element of fractional Sobolev spaces and we will use the fact that Assumption 1 implies Assumption 2, as illustrated below.
- Assumption 1 implies Assumption 2. Indeed take f according to Assumption 1. Then we can define the functional \bar{f} as follows $\bar{f}(t, u, v)(\cdot) := f(t, \cdot, u(\cdot), v(\cdot))$ for $u \in H_p^{1+\delta}$ and $v \in H_p^\delta$. The first and third bullet points of Assumption 2 are obvious. The second bullet point can be proven as follows. First we show that for $(t, u, v) \in [0, T] \times H_p^{1+\delta} \times H_p^\delta$ then $\bar{f}(t, u, v) \in H_p^0$. Indeed we have

$$\begin{aligned} & \int_{\mathbb{R}^d} |\bar{f}(t, u, v)(x)|^p dx = \int_{\mathbb{R}^d} |f(t, x, u(x), v(x))|^p dx \\ & \leq c \int_{\mathbb{R}^d} |f(t, x, u(x), v(x)) - f(t, x, 0, 0)|^p dx \\ & \quad + c \int_{\mathbb{R}^d} |f(t, x, 0, 0)|^p dx \\ & \leq cL^p (\|u\|_{H_p^{1+\delta}}^p + \|v\|_{H_p^\delta}^p) + \sup_{0 \leq t \leq T} \|f(t, 0, 0)\|_{H_p^0}^p < \infty, \end{aligned}$$

where the constant c depends on p .

Now with similar calculations one can prove that given any $u, u' \in H_p^{1+\delta}$ and $v, v' \in H_p^\delta$ it holds

$$\|\bar{f}(t, u, v) - \bar{f}(t, u', v')\|_{H_p^0} \leq cL \left(\|u - u'\|_{H_p^0} + \|v - v'\|_{H_p^0} \right),$$

where the constant c depends on p , and L is the Lipschitz constant for f .

Example.

- An easy case is the class of functions f linear in (y, z) , for example $f(t, x, y, z) = c(t) \cdot (y + z) + d(x)$, where $t \mapsto c(t)$ is continuous on $[0, T]$ and $x \mapsto d(x)$ is bounded in \mathbb{R}^d and $L^p(\mathbb{R}^d)$ -integrable, for example $d(x) = e^{-|x|^2}$. In this case we would have $\bar{f}(t, u, v) = c(t) \cdot (u + v) + d$.
- A non-linear example is given by $f(t, x, y, z) = c(t) \cdot \sin(y + z) + d(x)$, where c and d are as above. Then we would get $\bar{f}(t, u, v) = c(t) \cdot \sin(u + v) + d$, which is Lipschitz continuous in (u, v) and bounded at 0 uniformly in (t, x) .

3. THE SEMI-LINEAR PDE

In this section we analyse the PDE (3) and obtain several bounds for its solution and for the mollified version. We refer the reader to [16, 18] for results on different (S)PDEs obtained using similar techniques, and [19] for the general case of linear equations in metric measure spaces.

3.1. Existence and uniqueness of a mild solution. We recall the PDE below for ease of reading:

$$(8) \quad \begin{cases} u_t(t, x) + L^b u(t, x) + f(t, x, u(t, x), \nabla u(t, x)) = 0, \\ u(T, x) = \Phi(x), \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d. \end{cases}$$

Here the operator $L^b u = \frac{1}{2} \Delta u + \nabla u b$ is defined component by component by $(L^b u)_i(t, x) = \frac{1}{2} \Delta u_i(t, x) + \nabla u_i(t, x) b(t, x)$ for all $i = 1, \dots, d$. The peculiarity of this PDE is that it involves a distributional coefficient b and in particular its product with ∇u . The meaning we give to this product makes use of the pointwise product recalled in Section 2. We follow the study of a similar equation from the first author in [18]. Here the novelty is that the PDE is non-linear, with the extra term f appearing. We are going to look for mild solutions, hence the following definition is in order.

Definition 4. A mild solution of (8) is an element u of $C([0, T], H_p^{1+\delta})$ which is a solution of the following integral equation

$$(9) \quad \begin{aligned} u(t) = & P_p(T-t)\Phi + \int_t^T P_p(r-t) (\nabla u(r) b(r)) dr \\ & + \int_t^T P_p(r-t) f(r, u(r), \nabla u(r)) dr, \end{aligned}$$

where $(P_p(t), t \geq 0)$ is the semigroup generated by $\frac{1}{2}\Delta$ and recalled in Section 2.

To solve the PDE (8) we will use a fixed point argument in equation (9) and for that we need f to be an element of a fractional Sobolev space as function of x and further to be Lipschitz continuous in such space: this is what is stated in Assumption 2.

Theorem 5. Suppose that Assumption 2 holds. Then there exists a unique mild solution $u \in C([0, T], H_p^{1+\delta})$ of (8).

Proof. The idea of the proof is similar to the proof of [18, Theorem 3.5] and [13, Theorem 14]: We look for a fixed point in $C([0, T], H_p^{1+\delta})$, in particular we show that the mapping defined by the right-hand side of (9) is a contraction by using the family of equivalent norms $\|\cdot\|_{\infty, H_p^{1+\delta}}^{(\rho)}$.

To this aim, we rewrite the mild solution in a forward form for $\bar{u}(t) = u(T-t)$. We get

$$(10) \quad \begin{aligned} \bar{u}(t) = & P_p(t)\Phi \\ & + \int_0^t P_p(t-r) (\nabla \bar{u}(r) b(T-r) + f(T-r, \bar{u}(r), \nabla \bar{u}(r))) dr \\ = & P_p(t)\Phi + \int_0^t P_p(t-r) (\nabla \bar{u}(r) \bar{b}(r) + \bar{f}(r, \bar{u}(r), \nabla \bar{u}(r))) dr, \end{aligned}$$

where $\bar{b}(r) = b(T-r)$ and $\bar{f}(r, \bar{u}(r)) = f(T-r, \bar{u}(r), \nabla \bar{u}(r))$. Since $\bar{b}, \bar{f}, \bar{u}$ and b, f, u share the same regularities in r , with a slight abuse of notations, in the following we still write b, f and u instead of \bar{b}, \bar{f} and \bar{u} .

If we denote by $I_t(u)$ the right-hand side of (10), then we need to control the norm $\|I(u_1) - I(u_2)\|_{\infty, H_p^{1+\delta}}^{(\rho)}$ for any $u_1, u_2 \in C([0, T], H_p^{1+\delta})$, which is the sum of three terms: One with the initial condition, one term with b and one term with f . The initial condition $P_p(t)\Phi$ belongs to $H_p^{1+\delta}$ since $\Phi \in H_p^{1+\delta+2\gamma} \subset H_p^{1+\delta}$ and the semigroup is a contraction on $H_p^{1+\delta}$. The term including the distributional coefficient b can be treated exactly like in [18, Theorem 3.4] because the pointwise product

is linear. One gets the bound

$$\begin{aligned} & \left\| \int_0^\cdot P_p(\cdot - r) ((\nabla u_1(r) - \nabla u_2(r))b(r)) dr \right\|_{\infty, H_p^{1+\delta}}^{(\rho)} \\ & \leq C \rho^{\frac{\delta+\beta-1}{2}} \|b\|_{\infty, H_p^{-\beta}} \|u_1 - u_2\|_{\infty, H_p^{1+\delta}}^{(\rho)}, \end{aligned}$$

which is finite and the constant $C \rho^{\frac{\delta+\beta-1}{2}}$ tends to zero as $\rho \rightarrow \infty$ since $\delta + \beta - 1 < 0$ by assumption on the parameters.

The third term involves f and is estimated using the Lipschitz regularity of f and the mapping property (4) of $P_p(t)$ with $\beta = 0$. We get

$$\begin{aligned} & \left\| \int_0^\cdot P_p(\cdot - r) f(r, u_1(r), \nabla u_1(r)) dr - \int_0^\cdot P_p(\cdot - r) f(r, u_2(r), \nabla u_2(r)) dr \right\|_{\infty, H_p^{1+\delta}}^{(\rho)} \\ & \leq \sup_{0 \leq t \leq T} e^{-\rho t} \int_0^t \|P(t-r)(f(r, u_1(r), \nabla u_1(r)) - f(r, u_2(r), \nabla u_2(r)))\|_{H_p^{1+\delta}} dr \\ & \leq \sup_{0 \leq t \leq T} e^{-\rho t} \int_0^t C r^{-\frac{1+\delta}{2}} \|f(r, u_1(r), \nabla u_1(r)) - f(r, u_2(r), \nabla u_2(r))\|_{H_p^0} dr \\ & \leq C \sup_{0 \leq t \leq T} \int_0^t e^{-\rho(t-r)} e^{-\rho r} r^{-\frac{1+\delta}{2}} L \left(\|u_1(r) - u_2(r)\|_{H_p^{1+\delta}} + \|\nabla u_1(r) - \nabla u_2(r)\|_{H_p^\delta} \right) dr \\ & \leq 2C \|u_1 - u_2\|_{\infty, H_p^{1+\delta}}^{(\rho)} \sup_{0 \leq t \leq T} \int_0^t e^{-\rho(t-r)} r^{-\frac{1+\delta}{2}} dr \\ & \leq C \rho^{\frac{\delta-1}{2}} \|u_1 - u_2\|_{\infty, H_p^{1+\delta}}^{(\rho)}, \end{aligned}$$

where in the second to last inequality we used the definition of ρ -equivalent norm and the continuity of $\nabla : H_p^{1+\delta} \rightarrow H_p^\delta$. Note that again the exponent of ρ is negative since $\delta < 1$ by assumption. Thus for ρ large enough we have

$$\|I(u_1) - I(u_2)\|_{\infty, H_p^{1+\delta}}^{(\rho)} \leq C \|u_1 - u_2\|_{\infty, H_p^{1+\delta}}^{(\rho)},$$

where $C < 1$ does not depend on u_1 and u_2 . Hence by Banach's contraction principle there exists a unique solution $u \in C([0, T], H_p^{1+\delta})$. \square

Remark 6. Thanks to the choice of the parameters δ and p in $K(\beta, q)$ (which is always possible since $p > d/\delta$, see [13] for more details) and to Lemma 2, we have the embedding of $H_p^{1+\delta}$ in $C^{1,\alpha}$, where $\alpha = \delta - d/p$. So for each $t \in [0, T]$, the solution $u(t)$ as a function of x is in fact bounded, differentiable and the first derivative is Hölder continuous, $u(t) \in C^{1,\alpha}$.

We will use [13, Proposition 11] several times in this paper. We recall it here for the reader's convenience.

Proposition 7. *Let $h \in L^\infty([0, T]; H_p^{-\beta})$ and $g : [0, T] \rightarrow H_p^{-\beta}$ for $\beta \in \mathbb{R}$ be defined as*

$$g(t) = \int_0^t P_p(t-r)h(r)dr.$$

Then $g \in C^{0,\gamma}([0, T]; H_p^{2-2\varepsilon-\beta})$ for every $\varepsilon > 0$ and $\gamma \in (0, \varepsilon)$. Moreover, we have

$$(11) \quad \|g(t) - g(s)\|_{H_p^{2-2\varepsilon-\beta}} \leq C(t-s)^\gamma ((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma}) \|h\|_{\infty, H_p^{-\beta}}.$$

The proof of bound (11) can be found in the proof of [13, Proposition 11].

Additionally we can show the following lemma.

Lemma 8. *The mild solution u of (8) is Hölder continuous in time of any order $\gamma < \frac{1-\delta-\beta}{2}$, that is, $u \in C^{0,\gamma}([0, T]; H_p^{1+\delta})$.*

Proof. This is done using the results of Proposition 7 with $\varepsilon = \frac{1-\delta-\beta}{2}$ and noting that $P_p(\cdot)\Phi$ is γ -Hölder continuous if $\Phi \in H_p^{1+\delta+2\gamma}$, with $2\gamma < 1 - \delta - \beta$. \square

3.2. Uniform bounds on mollified mild solution. In the next sections we will make use of an approximating sequence b^n in place of b . We therefore need to describe its effect on the solution of the PDE (8) where the coefficient b is replaced by a coefficient b^n , that is

$$(12) \quad \begin{cases} u_t^n(t, x) + L^{b^n}u^n(t, x) + f(t, x, u^n(t, x), \nabla u^n(t, x)) = 0, \\ u^n(T, x) = \Phi(x), \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d, \end{cases}$$

where $L^{b^n}u^n(t, x) := \frac{1}{2}\Delta u^n(t, x) + \nabla u^n(t, x)b^n(t, x)$ is the analogue of L^b .

If b^n is smooth, for example $b^n \in C([0, T]; C_b^1(\mathbb{R}^d; \mathbb{R}^d))$ (bounded with bounded first derivatives), then u^n is a classical solution and it coincides with the mild solution found in Theorem 5. We will use this fact for example in the proof of Theorem 18. In what follows we state and prove some continuity results which hold also for b^n non-smooth.

Lemma 9. *Let Assumption 2 hold, and let $b^n \rightarrow b$ in $L^\infty([0, T]; H_q^{-\beta})$. Then*

- (i) $u^n \rightarrow u$ in $C([0, T]; H_p^{1+\delta})$ and there exists a constant C independent of n such that

$$\|u^n - u\|_{\infty, H_p^{1+\delta}} \leq C\|b^n - b\|_{\infty, H_q^{-\beta}}.$$

- (ii) $u^n \rightarrow u$ and $\nabla u^n \rightarrow \nabla u$ uniformly on $[0, T] \times \mathbb{R}^d$.

Proof. (i) By similar calculations as in Theorem 5 and by adding and subtracting $b^n(r)\nabla u(r)$ we have

$$\begin{aligned}
& \|u - u^n\|_{\infty, H_p^{1+\delta}}^{(\rho)} = \sup_{t \in [0, T]} e^{-\rho t} \|u(t) - u^n(t)\|_{H_p^{1+\delta}} \\
& \leq \sup_{t \in [0, T]} e^{-\rho t} \left(\int_0^t \|P_p(t-r)(\nabla u^n(r)b^n(r) - u(r)b(r))\|_{H_p^{1+\delta}} dr \right. \\
& \quad \left. + \int_0^t \|P_p(t-r)(f(r, u^n(r), \nabla u^n(r)) - f(r, u(r), \nabla u(r)))\|_{H_p^{1+\delta}} dr \right) \\
& \leq \sup_{t \in [0, T]} \left(C \int_0^t e^{-\rho(t-r)} (t-r)^{-\frac{1+\delta+\beta}{2}} e^{-\rho r} \|b^n(r)\|_{H_q^{-\beta}} \|u^n(r) - u(r)\|_{H_p^{1+\delta}} dr \right. \\
& \quad \left. + C \|b^n - b\|_{\infty, H_q^{-\beta}} \int_0^t e^{-\rho(t-r)} (t-r)^{-\frac{1+\delta+\beta}{2}} e^{-\rho r} \|u(r)\|_{H_p^{1+\delta}} dr \right) \\
& \quad + \sup_{t \in [0, T]} e^{-\rho t} \int_0^t r^{-\frac{1+\delta}{2}} \|f(r, u^n(r), \nabla u^n(r)) - f(r, u(r), \nabla u(r))\|_{H_p^0} dr \\
& \leq C \|b\|_{\infty, H_q^{-\beta}} \|u^n - u\|_{\infty, H_p^{1+\delta}}^{(\rho)} \rho^{\frac{\delta+\beta-1}{2}} + C \|b^n - b\|_{\infty, H_p^{1+\delta}} \|u\|_{H_p^{1+\delta}}^{(\rho)} \rho^{\frac{\delta+\beta-1}{2}} \\
& \quad + C \|u^n - u\|_{\infty, H_p^{1+\delta}}^{(\rho)} \rho^{\frac{\delta-1}{2}}.
\end{aligned}$$

Therefore there exists a ρ big enough so that

$$1 - C \left(\rho^{\frac{\delta+\beta-1}{2}} + \rho^{\frac{\delta-1}{2}} \right) > 0.$$

Hence for such ρ ,

$$\begin{aligned}
\|u - u^n\|_{\infty, H_p^{1+\delta}}^{(\rho)} & \leq \frac{C \|u\|_{H_p^{1+\delta}}^{(\rho)} \rho^{\frac{\delta+\beta-1}{2}}}{1 - C \left(\rho^{\frac{\delta+\beta-1}{2}} + \rho^{\frac{\delta-1}{2}} \right)} \|b^n - b\|_{\infty, H_p^{1+\delta}} \\
& = C \|b^n - b\|_{\infty, H_p^{1+\delta}}.
\end{aligned}$$

Part (ii) follows from part (i) and by the Fractional Morrey inequality (Lemma 2). \square

Lemma 10. *Let Assumption 2 hold and let b^n be such that $b^n \rightarrow b$ in $L^\infty(0, T; H_q^{-\beta})$. The mild solution u^n of (12) is Hölder continuous in time of any order $\gamma < \frac{1-\delta-\beta}{2}$, that is, $u^n \in C^{0,\gamma}([0, T]; H_p^{1+\delta})$. Moreover, we have the uniform bound:*

$$(13) \quad \|u^n\|_{C^{0,\gamma}([0, T]; H_p^{1+\delta})} \leq C$$

for some C independent of n .

Proof. We recall that

$$\begin{aligned}
(14) \quad \|u^n\|_{C^{0,\gamma}([0, T]; H_p^{1+\delta})} & = \sup_{0 \leq t \leq T} \|u^n(t)\|_{H_p^{1+\delta}} \\
& \quad + \sup_{0 \leq s < t \leq T} \frac{\|u^n(t) - u^n(s)\|_{H_p^{1+\delta}}}{|t - s|^\gamma}.
\end{aligned}$$

By Lemma 9, the first term on the right-hand side of (14) is bounded by

$$\|u^n\|_{\infty, H_p^{1+\delta}} \leq C \|u\|_{\infty, H_p^{1+\delta}},$$

where the constant C is independent of n . To bound the second term, let us consider the difference $u^n(t) - u^n(s)$ as the sum of three terms

$$(P_p(t)\Phi - P_p(s)\Phi) + (g_1^n(t) - g_1^n(s)) + (g_2^n(t) - g_2^n(s)),$$

where

$$g_1^n(t) = \int_0^t P_p(t-r) \nabla u^n(r) b^n(r) dr$$

and

$$g_2^n(t) = \int_0^t P_p(t-r) f(r, u^n(r), \nabla u^n(r)) dr.$$

Observe that since $\Phi \in H_p^{1+\delta+2\gamma}$, then $A^{\frac{1+\delta}{2}} \Phi \in H^{2\gamma}$ hence it belongs to $D(A^\gamma)$ and so does $P_p(s)A^{\frac{1+\delta}{2}} \Phi$. We have

$$\begin{aligned} & \|P_p(t)\Phi - P_p(s)\Phi\|_{H_p^{1+\delta}} \\ & \leq C \|(P_p(t-s) - I)P_p(s)A^{\frac{1+\delta}{2}} \Phi\|_{H_p^0} \\ & \leq C(t-s)^\gamma \|P_p(s)A^{\frac{1+\delta}{2}+\gamma} \Phi\|_{H_p^0} \\ & \leq C(t-s)^\gamma, \end{aligned}$$

where we have used the fact that for any $\phi \in D(A^\gamma)$ then $\|P_t\phi - \phi\|_{H_p^0} \leq C_\gamma t^\gamma \|A^\gamma \phi\|_{H_p^0}$. Observe also that for $\varepsilon > 0$ such that $1 + \delta \leq 2 - 2\varepsilon - \beta$, *i.e.*, $\varepsilon \leq \frac{1-\delta-\beta}{2}$, we have, for $i = 1, 2$,

$$(15) \quad \|g_i^n(t) - g_i^n(s)\|_{H_p^{1+\delta}} \leq \|g_i^n(t) - g_i^n(s)\|_{H_p^{2-2\varepsilon-\beta}}.$$

Moreover, for fixed $r \in [0, T]$, we have

$$\begin{aligned} \|\nabla u^n(r) b^n(r)\|_{H_p^{-\beta}} & \leq C \|b^n(r)\|_{H_q^{-\beta}} \|\nabla u^n(r)\|_{H_p^\delta} \\ & \leq C \|b^n\|_{\infty, H_q^{-\beta}} \|u^n\|_{\infty, H_p^{1+\delta}}. \end{aligned}$$

Hence by Proposition 7 applied to g_1^n and using (15) we get

$$\begin{aligned} \|g_1^n(t) - g_1^n(s)\|_{H_p^{1+\delta}} & \leq C(t-s)^\gamma \left((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma} \right) \|\nabla u^n b^n\|_{\infty, H_p^{-\beta}} \\ & \leq C(t-s)^\gamma \left((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma} \right), \end{aligned}$$

where C is independent of n because $u^n \rightarrow u$ in $C([0, T], H_p^{1+\delta})$ by Lemma 9 and $b^n \rightarrow b$ in $L^\infty(0, T; H_q^{-\beta})$ by hypothesis.

The difference involving g_2 is similar, but instead we use the Lipschitz property of f to get

$$\begin{aligned}
& \|f(r, u^n(r), \nabla u^n(r))\|_{H_p^{-\beta}} \\
& \leq \|f(r, u^n(r), \nabla u^n(r))\|_{H_p^0} \\
& \leq C \|f(r, u^n(r), \nabla u^n(r)) - f(r, 0, 0)\|_{H_p^0} + C \|f(r, 0, 0)\|_{H_p^0} \\
& \leq C \left(1 + \|u^n(r)\|_{H_p^{1+\delta}} + \|\nabla u^n(r)\|_{H_p^\delta} \right) \\
& \leq C \left(1 + \|u^n\|_{\infty, H_p^{1+\delta}} \right),
\end{aligned}$$

having also used the fact that $\sup_r \|f(r, 0, 0)\|_{H_p^0} < c$ by Assumption 2. Hence by Proposition 7 we get

$$\begin{aligned}
& \|g_2^n(t) - g_2^n(s)\|_{H_p^{1+\delta}} \\
& \leq C(t-s)^\varepsilon \|f(\cdot, u^n, \nabla u^n)\|_{\infty, H_p^{-\beta}} + C(t-s)^\gamma s^{\varepsilon-\gamma} \|f(\cdot, u^n, \nabla u^n)\|_{\infty, H_p^{-\beta}} \\
& \leq C \left(1 + \|u^n\|_{\infty, H_p^{1+\delta}} \right) \left((t-s)^\varepsilon + (t-s)^\gamma s^{\varepsilon-\gamma} \right) \\
& \leq C(t-s)^\gamma \left((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma} \right).
\end{aligned}$$

where C is independent of n . Putting the three terms together we get

$$\begin{aligned}
& \|u^n(t) - u^n(s)\|_{H_p^{1+\delta}} \\
& \leq \|P_p(t)\Phi - P_p(s)\Phi\|_{H_p^{1+\delta}} + \|g_1^n(t) - g_1^n(s)\|_{H_p^{1+\delta}} + \|g_2^n(t) - g_2^n(s)\|_{H_p^{1+\delta}} \\
& \leq C(t-s)^\gamma + 2C(t-s)^\gamma \left((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma} \right),
\end{aligned}$$

and so the second term on the right-hand side of (14) is bounded by

$$C + 2C \left((t-s)^{\varepsilon-\gamma} + s^{\varepsilon-\gamma} \right) \leq C(T),$$

for ε such that $\gamma < \varepsilon \leq \frac{1-\delta-\beta}{2}$, which is always possible since $2\gamma < 1 - \delta - \beta$ by assumption. \square

Both for u and u^n we have desirable continuity properties and bounds which are uniform in n .

Lemma 11. *Let Assumption 2 hold and let u and u^n be the solutions of (8) and (12) respectively. For $\nu = u$ and $\nu = u^n$, the following properties hold:*

For each $t \in [0, T]$ we have $\nu(t) \in C^{1,\alpha}$ and there exists a positive constant C independent of n such that

$$(16) \quad \sup_{0 \leq t \leq T} \left(\sup_{x \in \mathbb{R}^d} |\nu(t, x)| \right) \leq C,$$

and

$$(17) \quad \sup_{0 \leq t \leq T} \left(\sup_{x \in \mathbb{R}^d} |\nabla \nu(t, x)| \right) \leq C.$$

Moreover, there exists a positive constant C independent of n such that for any $t, s \in [0, T]$ and $x, y \in \mathbb{R}^d$ we have

$$(18) \quad |\nu(t, x) - \nu(s, y)| \leq C (|t - s|^\gamma + |x - y|),$$

and

$$(19) \quad |\nabla \nu(t, x) - \nabla \nu(s, y)| \leq C (|t - s|^\gamma + |x - y|^\alpha),$$

for any $\gamma < 1 - \beta - \delta$ and for $\alpha = \delta - \frac{d}{p}$.

Proof. Since $u \in C([0, T]; H_p^{1+\delta})$ and $(\delta, p) \in K(\beta, q)$, we can apply the fractional Morrey inequality (Lemma 2) and for all $t \in [0, T]$ we get $u(t) \in C^{1,\alpha}$ with $\alpha = \delta - \frac{d}{p}$. By using the definition of the norms in $C^{1,\alpha}$ and in $C([0, T]; H_p^{1+\delta}(\mathbb{R}))$ we get (16) for $\nu = u$.

For $\nu = u^n$, since from Lemma 9 part (i) it holds $u^n \rightarrow u$ in $C([0, T]; H_p^{1+\delta})$, then there exists a constant C such that

$$(20) \quad \|u^n\|_{\infty, H_p^{1+\delta}} \leq C \|u\|_{\infty, H_p^{1+\delta}}, \quad \forall n \geq 0.$$

Then we have

$$\sup_{0 \leq t \leq T} \left(\sup_{x \in \mathbb{R}^d} |u^n(t, x)| \right) \leq \|u^n\|_{\infty, H_p^{1+\delta}} \leq C \|u\|_{\infty, H_p^{1+\delta}}.$$

For (17), we first observe that by the definition of the norm in $C^{1,\alpha}$ and the continuous embedding $H_p^{1+\delta} \subset C^{1,\alpha}$ we have

$$\begin{aligned} \sup_{x \in \mathbb{R}^d} |\nabla u(t, x)| &\leq \|u(t)\|_{C^{1,\alpha}} \leq \|u(t)\|_{H_p^{1+\delta}} \\ &\leq \sup_{t \in [0, T]} \|u(t)\|_{H_p^{1+\delta}} = \|u\|_{C([0, T]; H_p^{1+\delta})} =: C, \end{aligned}$$

where the last bound is due to the fact that $u \in C([0, T]; H_p^{1+\delta})$. This proves (17) for $\nabla \nu = \nabla u$. Bound (17) for $\nabla \nu = \nabla u^n$ is obtained analogously by using (20).

To prove (18), let $(t, x), (s, y) \in [0, T] \times \mathbb{R}^d$. We have

$$\begin{aligned} &|u(t, x) - u(s, y)| \\ &\leq |u(t, x) - u(s, x)| + |u(s, x) - u(s, y)| \\ &\leq \sup_{x \in \mathbb{R}^d} |u(t, x) - u(s, x)| + |u(s, x) - u(s, y)| \\ &\leq \|u(t, \cdot) - u(s, \cdot)\|_{C^{1,\alpha}} + \|u(s, \cdot)\|_{C^{1,\alpha}} |x - y| \\ &\leq \|u(t, \cdot) - u(s, \cdot)\|_{H_p^{1+\delta}} + \|u(s, \cdot)\|_{H_p^{1+\delta}} |x - y| \\ &\leq \|u\|_{C^{0,\gamma}([0, T]; H_p^{1+\delta})} |t - s|^\gamma + \|u\|_{C^{0,\gamma}([0, T]; H_p^{1+\delta})} |x - y| \\ &\leq \|u\|_{C^{0,\gamma}([0, T]; H_p^{1+\delta})} (|t - s|^\gamma + |x - y|), \end{aligned}$$

having used the embedding property (fractional Morrey inequality) with $\alpha = \delta - d/p$, the Lipschitz property of $u(t, \cdot)$ (due to the fact that it is differentiable) and the Hölder property of $u(\cdot)$ with values

in $H_p^{1+\delta}$. Setting $C = \|u\|_{C^{0,\gamma}([0,T];H_p^{1+\delta})}$ concludes the proof of (18) for $\nu = u$.

The bound (18) for $\nu = u^n$ is obtained from the previous one: we proceed as the proof for $\nu = u$ and get

$$(21) \quad |u^n(t, x) - u^n(s, y)| \leq \|u^n\|_{C^{0,\gamma}([0,T];H_p^{1+\delta})} (|t - s|^\gamma + |x - y|).$$

Plugging (13) from Lemma 10 into (21), we get the desired result.

To show (19) for $\nabla\nu = \nabla u$ we proceed with very similar computations for $|\nabla u(t, x) - \nabla u(s, y)|$ as in the proof of (18), but now we use the fact that $\nabla u(s, \cdot)$ is only Hölder continuous of order α rather than Lipschitz continuous, that is $|\nabla u(s, x) - \nabla u(s, y)| \leq \|u(s, \cdot)\|_{C^{1,\alpha}} |x - y|^\alpha$, so we finally have

$$\begin{aligned} & |\nabla u(t, x) - \nabla u(s, y)| \\ & \leq |\nabla u(t, x) - \nabla u(s, x)| + |\nabla u(s, x) - \nabla u(s, y)| \\ & \leq \|u(t, \cdot) - u(s, \cdot)\|_{C^{1,\alpha}} + \|u(s, \cdot)\|_{C^{1,\alpha}} |x - y|^\alpha \\ & \leq \|u\|_{C^{0,\gamma}([0,T];H_p^{1+\delta})} (|t - s|^\gamma + |x - y|^\alpha), \end{aligned}$$

which is the claim with C as in the previous bound.

The proof of (19) for $\nabla\nu = \nabla u^n$ is similar and uses (13) in the last part. \square

4. SOLUTION OF BSDE (1)

4.1. Definition of solution, existence and uniqueness. In this section we consider FBSDE (1), which we write again below for convenience

$$(22) \quad \begin{cases} X_s^{t,x} = x + \int_t^s dW_r, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) - \int_s^T Z_r^{t,x} dW_r + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr \\ \quad + \int_s^T Z_r^{t,x} b(r, X_r^{t,x}) dr, \\ \forall s \in [t, T], \end{cases}$$

where $(W_s)_s$ is a given Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ and the filtration \mathbb{F} is the Brownian filtration. Here $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^m$ and $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$. We note that $X^{t,x} := (X_s^{t,x})_{s \in [t, T]}$ is in fact a Brownian motion starting from x at time t . The major difficulty related to (22) is the term $\int_s^T Z_r^{t,x} b(r, X_r^{t,x}) dr$ because $b \in L^\infty([0, T]; H_q^{-\beta})$. Given $X^{t,x}$, we introduce the notion of *virtual-strong solution* for the backward SDE in (22). To do so, we first consider the following auxiliary PDE

$$(23) \quad \begin{cases} w_t + \frac{1}{2} \Delta w = \nabla u b, \\ w(T, x) = 0, \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d, \end{cases}$$

where u is the mild solution of (8). The term $\nabla u b$ is defined by means of the pointwise product, and thanks to the semigroup properties (see

Section 2 for more details) there exists a unique mild solution $w \in C([0, T]; H_p^{1+\delta})$ to (23) which is given by

$$(24) \quad \begin{aligned} w(t) &= P_p(T-t)w(T) + \int_t^T P_p(r-t)\nabla u(r)b(r)dr \\ &= \int_t^T P_p(r-t)\nabla u(r)b(r)dr. \end{aligned}$$

Note that by the Fractional Morrey inequality (Lemma 2) we have that w can be evaluated pointwisely since $w \in C([0, T]; C^{1,\alpha})$ for $\alpha = \delta - \frac{d}{p}$. We use this function w to give a meaning to the backward SDE in (22) as follows. In the sequel we will drop the superscript t, x for simplicity of notation.

Definition 12. A virtual-strong solution to the backward SDE in (22) is a couple (Y, Z) such that

- Y is continuous and \mathbb{F} -adapted and Z is \mathbb{F} -progressively measurable;
- $\mathbb{E} [\sup_{r \in [t, T]} |Y_r|^2] < \infty$ and $\mathbb{E} \left[\int_t^T |Z_r|^2 dr \right] < \infty$;
- for all $s \in [t, T]$, the couple satisfies the following backward SDE

$$(25) \quad \begin{aligned} Y_s &= \Phi(X_T) - \int_s^T Z_r dW_r + \int_s^T f(r, X_r, Y_r, Z_r) dr \\ &\quad - w(s, X_s) - \int_s^T \nabla w(r, X_r) dW_r \end{aligned}$$

\mathbb{P} -almost surely, where w is the solution of (23) given by (24).

An intuitive explanation on why we define virtual-strong solutions like this is the fact that if b were smooth, also w would be smooth and we could apply Itô's formula to $w(\cdot, X)$, where $X_s = x + W_s - W_t$, to get

$$\begin{aligned} dw(s, X_s) &= w_t(s, X_s)ds + \nabla w(s, X_s)dX_s + \frac{1}{2}\Delta w(s, X_s)ds \\ &= \nabla u(s, X_s)b(s, X_s)ds + \nabla w(s, X_s)dW_s. \end{aligned}$$

Therefore, we could write

$$\begin{aligned} w(T, X_T) - w(s, X_s) &- \int_s^T \nabla w(r, X_r) dW_r \\ &= -w(s, X_s) - \int_s^T \nabla w(r, X_r) dW_r \\ &= \int_s^T \nabla u(r, X_r)b(r, X_r)dr \\ &= \int_s^T Z_r b(r, X_r)dr, \end{aligned}$$

where the last equality holds because in the smooth case the solution (Y, Z) could be written as $(u(\cdot, X), \nabla u(\cdot, X))$. This is why the term $-w(s, X_s) - \int_s^T \nabla w(r, X_r) dW_r$ appears in (25) in place of $\int_s^T Z_r b(r, X_r) dr$.

We recall that a *strong solution* of (25) is a couple (Y, Z) such that

- Y is continuous and \mathbb{F} -adapted, Z is \mathbb{F} -progressively measurable;
- $\mathbb{E} [\sup_{r \in [t, T]} |Y_r|^2] < \infty$ and $\mathbb{E} [\int_t^T |Z_r|^2 dr] < \infty$;
- (25) holds \mathbb{P} -almost surely.

Note that the terms involving w in (25) do not pose any extra condition because we can prove that w is continuous and bounded (see Lemma 16 below).

The notion of virtual-strong solution for BSDE is in alignment with classical strong solutions when the drift b is a function with classical regularity properties. In this case a virtual-strong solution is also a strong solution, as illustrated in the proposition below.

Proposition 13. *Let $b \in C([0, T]; C_b^1(\mathbb{R}^d, \mathbb{R}^d))$ (bounded with bounded first derivatives). Then the virtual-strong solution (Y, Z) of the backward SDE in (22) is also a strong solution.*

Proof. First observe that the first two conditions for Y and Z in Definition 12 are the same as for strong solutions.

Let u be the classical solution of (8) and w be the classical solution of (23). Then u and w are both at least of class $C^{1,2}$ and by Itô's formula applied to w we have that the term $-w(s, X_s) - \int_s^T \nabla w(r, X_r) dW_r$ is equal to $\int_s^T Z_r b(r, X_r) dr$, hence the BSDE in (22) holds \mathbb{P} -a.s.. \square

We remark that, although every term in the backward SDE (25) is well defined, this SDE is not written in a classical form. Hence to find a virtual-strong solution we transform (25) using the solution of the PDE (23), in particular we apply the transformation $y \mapsto y + w(s, x)$ where w is the solution of the PDE (23). This transformation could be regarded as the analogous of the Zvonkin transformation for SDEs to get rid of a (singular) drift. More precisely, we set $\widehat{Y}_s := Y_s + w(s, X_s)$ and $\widehat{Z}_s := Z_s + \nabla w(s, X_s)$ for all $s \in [t, T]$ and $\widehat{f}(r, x, y, z) := f(r, x, y - w(r, x), z - \nabla w(r, x))$, and we get the following *auxiliary backward SDE*

$$(26) \quad \widehat{Y}_s = \Phi(X_T) - \int_s^T \widehat{Z}_r dW_r + \int_s^T \widehat{f}(r, X_r, \widehat{Y}_r, \widehat{Z}_r) dr,$$

for all $s \in [t, T]$.

It turns out that indeed the BSDEs (25) and (26) are equivalent as shown in the following proposition.

Proposition 14. *Let X be a Brownian motion starting from x at time t and \mathbb{F} be the Brownian filtration generated by W . Then*

- (i) If (Y, Z) is a virtual-strong solution of the backward SDE in (22), then

$$(\widehat{Y}, \widehat{Z}) := (Y + w(\cdot, X), Z + \nabla w(\cdot, X))$$

is a strong solution of (26).

- (ii) If $(\widehat{Y}, \widehat{Z})$ is a strong solution of (26), then

$$(Y, Z) := (\widehat{Y} - w(\cdot, X), \widehat{Z} - \nabla w(\cdot, X))$$

is a virtual-strong solution of the backward SDE in (22).

Proof. The proof is very easy and straight-forward, so we omit it. \square

We will now prove existence and uniqueness of the virtual-strong solution for the FBSDE (22). For this we need Assumption 1.

Theorem 15. *Under Assumption 1 there exists a unique virtual-strong solution (Y, Z) to the backward SDE in (22).*

Proof. By definition, a virtual-strong solution of the backward SDE in (22) is a couple that solves BSDE (25), if u exists. Note that by Remark 3 we know that Assumption 1 implies Assumption 2, hence u does exist by Theorem 5. Moreover BSDE (25) is equivalent to BSDE (26) by Proposition 14.

Using the Lipschitz assumption on f from Assumption 1 and the definition of \widehat{f} , we have for any $y, y' \in \mathbb{R}^m$ and $z, z' \in \mathbb{R}^{m \times d}$ that

$$\begin{aligned} & |\widehat{f}(t, x, y, z) - \widehat{f}(t, x, y', z')| \\ &= |f(t, x, y - w(t, x), z - \nabla w(t, x)) - f(t, x, y' - w(t, x), z' - \nabla w(t, x))| \\ &\leq C(|y - y'| + |z - z'|). \end{aligned}$$

Moreover by definition of \widehat{f} we have

$$\begin{aligned} & \mathbb{E} \left[\int_0^T |\widehat{f}(r, x + W_r, 0, 0)|^2 dr \right] \\ &= \mathbb{E} \left[\int_0^T |f(r, x + W_r, -w(r, x + W_r), -\nabla w(r, x + W_r))|^2 dr \right] \\ &\leq C \left(1 + E \left[\int_0^T |f(r, x + W_r, 0, 0)|^2 dr \right] \right), \end{aligned}$$

where we have used the fact that w and ∇w are uniformly bounded by Lemma 16. The latter integral is bounded using the assumption of $f(t, x, 0, 0)$, indeed

$$\mathbb{E} \left[\int_0^T |f(r, x + W_r, 0, 0)|^2 dr \right] \leq \mathbb{E} \int_0^T \sup_{t,x} |f(t, x, 0, 0)|^2 dr \leq c.$$

Hence equation (26) has a unique strong solution by classical results (see for example [9, Theorem 2.1]). \square

4.2. The auxiliary PDE and the auxiliary BSDE. We now establish several useful properties for the auxiliary PDE (23) and for the auxiliary BSDE (26), which will be used in the next Section to prove the non-linear Feynman-Kac formula.

We start by proving a result analogous to Lemma 11.

Lemma 16. *Let Assumption 2 hold and $b \in L^\infty(0, T; H_q^{-\beta})$. Then the solution w is an element of $C^{0,\gamma}([0, T]; H_p^{1+\delta})$ for all $2\gamma < 1 - \delta - \beta$ and it enjoys the following bounds*

$$(27) \quad \sup_{0 \leq t \leq T} \left(\sup_{x \in \mathbb{R}^d} |w(t, x)| \right) \leq C,$$

$$(28) \quad \sup_{0 \leq t \leq T} \left(\sup_{x \in \mathbb{R}^d} |\nabla w(t, x)| \right) \leq C.$$

Furthermore, for all $t, s \in [0, T]$ and $x, y \in \mathbb{R}^d$ we have

$$(29) \quad |w(t, x) - w(s, y)| \leq C (|t - s|^\gamma + |x - y|),$$

$$(30) \quad |\nabla w(t, x) - \nabla w(s, y)| \leq C (|t - s|^\gamma + |x - y|^\alpha),$$

where $\alpha = \delta - \frac{d}{p}$.

Proof. To show that $w \in C([0, T]; H_p^{1+\delta})$ we first observe that $\nabla u b \in L^\infty([0, T]; H_p^{-\beta})$ since

$$\|\nabla u(s)b(s)\|_{H_p^{-\beta}} \leq C \|\nabla u(s)\|_{H_p^\delta} \|b(s)\|_{H_q^{-\beta}},$$

and taking the supremum over $s \in [0, T]$ the right-hand side is bounded by a constant which is independent of s . Hence

$$\|\nabla u b\|_{\infty, H_p^{-\beta}} \leq \sup_{0 \leq s \leq T} C \|u(s)\|_{H_p^{1+\delta}} \|b\|_{\infty, H_q^{-\beta}} \leq C(b, u).$$

By Proposition 7 applied to equation (24) we have that $w \in C^{0,\gamma}([0, T]; H_p^{2-2\varepsilon-\beta})$ for every $\varepsilon > 0$ and $\gamma \in (0, \varepsilon)$, and setting with $\varepsilon = \frac{1-\delta-\beta}{2}$ it implies $w \in C([0, T]; H_p^{1+\delta})$.

The bounds (27) and (28) follow by fractional Morrey inequality (Lemma 2)

$$w \in C^{0,\gamma}([0, T]; H_p^{1+\delta}) \subset C^{0,\gamma}([0, T]; C^{1,\alpha}),$$

where $\alpha = \delta - \frac{d}{p}$. Hence the $\sup_{t,x}$ of the functions w and ∇w are finite.

The bound (29) is clear by using the norm definition in $C^{0,\gamma}$, whereas (30) can be obtained by using the fact that $w \in C^{0,\gamma}([0, T]; C^{0,1+\alpha})$ implies $\nabla w \in C^{0,\gamma}([0, T]; C^{0,\alpha})$ and applying the definition of the norm in the latter space. \square

If we now consider a smooth coefficient b^n in place of b then the PDE (23) becomes

$$(31) \quad \begin{cases} w_t^n + \frac{1}{2}\Delta w^n = \nabla u^n b^n, \\ w^n(T, x) = 0, \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d. \end{cases}$$

For this approximating PDE we have nice convergence properties as follows.

Lemma 17. *Let Assumption 2 hold and let $b^n \rightarrow b$ in $L^\infty([0, T]; H_q^{-\beta})$. Then $w^n \rightarrow w$ in $C^{0,\gamma}([0, T]; C^{1,\alpha})$ and $\nabla w^n \rightarrow \nabla w$ in $C^{0,\gamma}([0, T]; C^{0,\alpha})$. In particular, $w^n(t, x) \rightarrow w(t, x)$ and $\nabla w^n(t, x) \rightarrow \nabla w(t, x)$ uniformly on $[0, T] \times \mathbb{R}^d$.*

Proof. By Lemma 16 we have that w and w^n are both elements of $C^{0,\gamma}([0, T]; H_p^{1+\delta})$. The norm of $w - w^n$ in $C^{0,\gamma}([0, T]; H_p^{1+\delta})$ has two terms, as recalled in Section 2. The first one can be bounded by observing that

$$\begin{aligned} w(T-t) - w^n(T-t) &= \int_0^t P_p(r) (\nabla u(r+T-t)b(r+T-t) \\ &\quad - \nabla u^n(r+T-t)b^n(r+T-t)) dr \end{aligned}$$

and by abuse of notation we consider the semigroup simply acting on $\nabla u(r)b(r) - \nabla u^n(r)b^n(r)$ because the regularity properties are the same. So

$$\begin{aligned} &\|w(T-t) - w^n(T-t)\|_{H_p^{1+\delta}} \\ &\leq \left\| \int_0^t P_p(r) (\nabla u(r)b(r) - \nabla u(r)b^n(r)) dr \right\|_{H_p^{1+\delta}} \\ &\quad + \left\| \int_0^t P_p(r) (\nabla u(r)b^n(r) - \nabla u^n(r)b^n(r)) dr \right\|_{H_p^{1+\delta}} \\ &\leq \int_0^t r^{-\frac{1-\delta-\beta}{2}} \left(\|u(r)\|_{H_p^{1+\delta}} \|b(r) - b^n(r)\|_{H_q^{-\beta}} \right. \\ &\quad \left. + \|b^n(r)\|_{H_q^{-\beta}} \|\nabla u(r) - \nabla u^n(r)\|_{H_p^{1+\delta}} \right) dr \\ &\leq CT^{\frac{1+\delta+\beta}{2}} \|b - b^n\|_{\infty, H_q^{-\beta}}, \end{aligned}$$

where the constant C is independent of n (for n large enough) because $u^n \rightarrow u$ as shown in Lemma 9, part (i) and $b_n \rightarrow b$ by hypotheses. Thus

$$\begin{aligned} \sup_{t \in [0, T]} \|w(t) - w^n(t)\|_{H_p^{1+\delta}} &= \sup_{t \in [0, T]} \|w(T-t) - w^n(T-t)\|_{H_p^{1+\delta}} \\ &\leq C \|b - b^n\|_{\infty, H_q^{-\beta}}. \end{aligned}$$

The Hölder term in the norm of $w - w^n$ can be bounded by using Proposition 7 with $\varepsilon = \frac{1-\delta-\beta}{2}$, since the integrand $h(r) := b(r)\nabla u(r) - b^n(r)\nabla u^n(r)$ belongs to $H_p^{-\beta}$. Then we have

$$\frac{\|w^n(t) - w(t) - (w^n(s) - w(s))\|_{H^{1+\delta}}}{|t - s|^\gamma} \leq C \|h\|_{\infty, H_p^{-\beta}},$$

where C is independent of n and the norm of h is bounded by $C\|b - b^n\|_{\infty, H_q^{-\beta}}$ as done above. Hence we have shown that

$$w^n \rightarrow w \text{ in } C^{0,\gamma}([0, T]; H_p^{1+\delta})$$

which implies

$$\nabla w^n \rightarrow \nabla w \text{ in } C^{0,\gamma}([0, T]; H_p^\delta)$$

by the continuity of the mapping $\nabla : H_p^{1+\delta} \rightarrow H_p^\delta$.

By the Sobolev embedding (Lemma 2) we have $C^{0,\gamma}([0, T]; H_p^{1+\delta}) \subset C^{0,\gamma}([0, T]; C^{1,\alpha})$ and so it follows that

$$\sup_{0 \leq t \leq T} \sup_{x \in \mathbb{R}^d} |w^n(t, x) - w(t, x)| \leq C \|b - b^n\|_{\infty, H_q^{-\beta}}$$

and

$$\sup_{0 \leq t \leq T} \sup_{x \in \mathbb{R}^d} |\nabla w^n(t, x) - \nabla w(t, x)| \leq C \|b - b^n\|_{\infty, H_q^{-\beta}},$$

which is the uniform convergence claimed. \square

4.3. Feynman-Kac representation formula. In this last section we will establish a non-linear Feynman-Kac representation formula for the FBSDE (22) using the solution of the PDE (8) and of the auxiliary PDE (23). In particular, we will construct the virtual-strong solution of (22) –that is a strong solution of (25)– by means of the mild solution of the PDE (8), and we will also show that the unique mild solution can be obtained as the first component Y at initial time t of the virtual-strong solution (Y, Z) , and in this case the gradient of the solution corresponds to Z .

Theorem 18. *Let Assumption 1 hold. Let u be the unique mild solution of (8) and X be the solution of the forward equation in (22), namely $X_s = x + W_s - W_t$, $s \in [t, T]$. Then the couple $(u(\cdot, X), \nabla u(\cdot, X))$ is a virtual-strong solution of the backward SDE in (22).*

Proof. First we note that by Remark 3 we can consider the composition of f with $u, \nabla u$ and this satisfies Assumption 2. Hence by Theorem 5 we know that a solution u to PDE (8) exists and it is unique. Furthermore this solution is in $C([0, T]; C^{1,\alpha})$ for some small $\alpha > 0$ by Lemma 11 and it is uniformly bounded in (t, x) . These properties, together with the fact that X is a Brownian motion starting in x at time t , imply that the first two bullet points of Definition 12 are easily satisfied for the couple $(u(\cdot, X), \nabla u(\cdot, X))$. The only non-trivial point to verify in this definition is to show that $(u(\cdot, X), \nabla u(\cdot, X))$ satisfies (25), where w is given by (23).

To show this we take a smooth approximating sequence, e.g. $b^n \in L^\infty(0, T; C_b^1(\mathbb{R}^d; \mathbb{R}^d))$, such that $b^n \rightarrow b$ converges in $L^\infty(0, T; L_q^{-\beta})$. The PDE (8) then becomes (12) and PDE (23) becomes (31). These approximations are smooth so we can apply Itô's formula to both $u^n(\cdot, X)$ and $w^n(\cdot, X)$, and get

$$\begin{aligned} du^n(s, X_s) &= -\nabla u^n(s, X_s)b^n(s, X_s)ds \\ &\quad - f(s, X_s, u^n(s, X_s), \nabla u^n(s, X_s))ds + \nabla u^n(s, X_s)dW_s, \end{aligned}$$

and

$$dw^n(s, X_s) = \nabla w^n(s, X_s)b^n(s, X_s)ds + \nabla w^n(s, X_s)dW_s.$$

Adding the second equation to the first we get rid of the term with $\nabla u^n b^n$ and we end up with

$$\begin{aligned} du^n(s, X_s) &= -dw^n(s, X_s) - f(s, X_s, u^n(s, X_s), \nabla u^n(s, X_s))ds \\ &\quad + \nabla w^n(s, X_s)dW_s + \nabla u^n(s, X_s)dW_s. \end{aligned}$$

Integrating from s to T gives

$$\begin{aligned} (32) \quad u^n(s, X_s) &= \Phi(X_T) - w^n(s, X_s) \\ &\quad + \int_s^T f(r, X_r, u^n(r, X_r), \nabla u^n(r, X_r))dr \\ &\quad - \int_s^T \nabla w^n(r, X_r)dW_r - \int_s^T \nabla u^n(r, X_r)dW_r. \end{aligned}$$

Our aim to show that the limit of (32) is given by

$$\begin{aligned} (33) \quad u(s, X_s) &= \Phi(X_T) - w(s, X_s) \\ &\quad + \int_s^T f(r, X_r, u(r, X_r), \nabla u(r, X_r))dr \\ &\quad - \int_s^T \nabla w(r, X_r)dW_r - \int_s^T \nabla u(r, X_r)dW_r. \end{aligned}$$

We will consider the limit in \mathbb{S}^2 : For a stochastic process $(\xi_s)_{t \leq s \leq T}$ the norm in \mathbb{S}^2 is given by $\mathbb{E}[\sup_{t \leq s \leq T} |\xi_s|^2]$. We take the difference

of (32) and (33), then by triangular inequality is enough to show \mathbb{S}^2 -convergence to zero for each of the following five terms:

$$\begin{aligned} & u^n(\cdot, X) - u(\cdot, X) \\ & w^n(\cdot, X) - w(\cdot, X) \\ & \int_{\cdot}^T f(r, X_r, u^n(r, X_r), \nabla u^n(r, X_r)) dr \\ & \quad - \int_{\cdot}^T f(r, X_r, u(r, X_r), \nabla u(r, X_r)) dr \\ & \int_{\cdot}^T \nabla u^n(r, X_r) dW_r - \int_{\cdot}^T \nabla u(r, X_r) dW_r \\ & \int_{\cdot}^T \nabla w^n(r, X_r) dW_r - \int_{\cdot}^T \nabla w(r, X_r) dW_r. \end{aligned}$$

The first two are a consequence of uniform convergence of u^n to u and w^n to w (which is proven in Lemma 9 and 17). The third term converges to zero thanks to the Lipschitz continuity of f (by Assumption 1) and uniform convergence of u^n and ∇u^n (again by Lemma 9). The last two terms can be bounded using BDG inequality and Lemma 17) as follows (we show it only for w , the same applies to u thanks to Lemma 9).

$$\begin{aligned} & \mathbb{E} \left[\sup_{t \leq s \leq T} \left| \int_s^T (\nabla w^n(r, X_r) - \nabla w(r, X_r)) dW_r \right|^2 \right] \\ & \leq c \mathbb{E} \left[\int_s^T (\nabla w^n(r, X_r) - \nabla w(r, X_r))^2 dr \right] \\ & \leq c \mathbb{E} \left[\int_s^T \left(\sup_{r,x} |\nabla w^n(r, X_r) - \nabla w(r, X_r)| \right)^2 dr \right] \rightarrow 0. \end{aligned}$$

This concludes the proof. \square

From Theorem 18 and using Proposition 14, it is also easily seen that $(u(\cdot, X) + w(\cdot, X), \nabla u(\cdot, X) + \nabla w(\cdot, X))$ is a strong solution of (26), where u is the solution of PDE (8) and w is the solution of (23).

Next we have the opposite result, namely that the BSDE provides a representation for the mild solution of the PDE. For this result we resume the use of the superscript t, x for better clarity.

Theorem 19. *Let Assumption 2 hold, and let $(Y^{t,x}, Z^{t,x})$ be a virtual-strong solution of the backward SDE in (22). Assume further that there exists deterministic functions $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ such that*

$$Y_s^{t,x} = \alpha(s, X_s^{t,x}) \quad \text{and} \quad Z_s^{t,x} = \beta(s, X_s^{t,x})$$

for all $s \in [0, T]$. Moreover assume that $\alpha \in C^\varepsilon([0, T]; H_p^{1+\delta})$ (form some $\varepsilon > 0$) and $\beta \in C([0, T]; H_p^\delta)$. Then the unique mild solution of

(8) can be written as $u(t, x) = Y_t^{t,x}$. Moreover we have that $\nabla u(t, x) = Z_t^{t,x}$.

Proof. Since $(Y^{t,x}, Z^{t,x}) = (\alpha(\cdot, X^{t,x}), \beta(\cdot, X^{t,x}))$ is a virtual-strong solution of the backward SDE in (22), we have for $s = t$

$$(34) \quad \begin{aligned} \alpha(t, x) &= \Phi(X_T^{t,x}) - \int_t^T \beta(r, X_r^{t,x}) dW_r \\ &\quad + \int_t^T f(r, X_r^{t,x}, \alpha(r, X_r^{t,x}), \beta(r, X_r^{t,x})) dr \\ &\quad - w(t, x) - \int_t^T \nabla w(r, X_r^{t,x}) dW_r. \end{aligned}$$

Note that the stochastic integrals in (34) have zero-mean because both integrands are square integrable. We denote by $\mathbb{P}_{t,x}$ the probability measure of $X^{t,x}$ (which we recall is a Brownian motion starting in x at t) and by $\mathbb{E}_{t,x}$ the expectation under this measure, namely $\mathbb{E}[X_s^{t,x}] = \mathbb{E}_{t,x}[X_s]$, where X_s is the canonical process. Moreover, this process X generates the heat semigroup under this measures, namely for all bounded and measurable a we have

$$\mathbb{E}[a(s, X_s^{t,x})] = \mathbb{E}_{t,x}[a(s, X_s)] = (P(s-t)a(s, \cdot))(x).$$

The heat semigroup P coincides with the semigroup P_p when it acts on elements in L^p . Then taking the expectation \mathbb{E} on both sides of (34) we get

$$(35) \quad \begin{aligned} \alpha(t, \cdot) &= \mathbb{E}[\Phi(X_T^{t,\cdot})] - w(t, \cdot) \\ &\quad + \mathbb{E}\left[\int_t^T f(r, X_r^{t,\cdot}, \alpha(r, X_r^{t,\cdot}), \beta(r, X_r^{t,\cdot})) dr\right] \\ &= P_p(T-t)\Phi - w(t) + \int_t^T P_p(r-t)f(r, \cdot, \alpha(r), \beta(r)) dr \\ &= P_p(T-t)\Phi + \int_t^T P_p(r-t)(\nabla u(r) b(r)) dr \\ &\quad + \int_t^T P_p(r-t)f(r, \cdot, \alpha(r), \beta(r)) dr, \end{aligned}$$

having used in the last equality that w is the mild solution of (23). Next we calculate the covariation of Y and W . We use the covariation defined in [15], recalled below for convenience:

$$[Y, W]_s := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^s (Y_{r+\varepsilon} - Y_r)(W_{r+\varepsilon} - W_r) dr,$$

if the limit exists *u.c.p.* in s . Notice that $\alpha \in C^\varepsilon([0, T]; H_p^{1+\delta})$ implies by fractional Morrey inequality (Lemma 2) that α is continuous in time and $C^{1,\gamma}$ in space with $\gamma = \delta - \frac{d}{p}$. Moreover one can show that

$\alpha \in C^{0,1}([0, T] \times \mathbb{R}^d)$ by similar computations as [13, Lemma 21], thus we can apply [15, Corollary 3.13] and get

$$[Y, W]_s = [\alpha(\cdot, X^{t,x}), W]_s = \int_0^s \nabla \alpha(r, X_r^{t,x}) dr.$$

On the other hand, the covariation calculated using the BSDE (34) gives

$$\begin{aligned} & [Y, W]_s \\ &= \left[\Phi(X_T^{t,x}) - \int_{\cdot}^T Z_r^{t,x} dW_r + \int_{\cdot}^T f(r, X_r^{t,x}, \alpha(r, X_r^{t,x}), \beta(r, X_r^{t,x})) dr, W \right]_s \\ & \quad + \left[-w(\cdot, X^{t,x}) - \int_{\cdot}^T \nabla w(r, X_r^{t,x}) dW_r, W \right]_s \\ &= - \left[\int_{\cdot}^T Z_r^{t,x} dW_r, W \right]_s - [w(\cdot, X^{t,x}), W]_s \\ & \quad - \left[\int_{\cdot}^T \nabla w(r, X_r^{t,x}) dW_r, W \right]_s \\ &= \int_0^s Z_r^{t,x} dr - \int_0^s \nabla w(r, X_r^{t,x}) dr + \int_0^s \nabla w(r, X_r^{t,x}) dr \\ &= \int_0^s \beta(r, X_r^{t,x}) dr. \end{aligned}$$

Therefore $\beta(s, X_s^{t,x}) = \nabla \alpha(s, X_s^{t,x})$ for all s . Equation (35) becomes

$$\begin{aligned} \alpha(t) &= P_p(T-t)\Phi + \int_t^T P_p(r-t) (\nabla u(r) b(r)) dr \\ & \quad + \int_t^T P_p(r-t) f(r, \alpha(r), \nabla \alpha(r)) dr. \end{aligned}$$

We remark that this is exactly the mild formulation of

$$(36) \quad \begin{cases} \alpha_t(t, x) + \frac{1}{2} \Delta \alpha(t, x) + \nabla u(t, x) b(t, x) + f(t, \alpha(t, x), \nabla \alpha(t, x)) = 0, \\ \alpha(T, x) = \Phi(x), \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d, \end{cases}$$

where u is the mild solution of (8). With a very similar proof of Theorem 5 one can show that there exists a unique mild solution $\alpha \in C([0, T]; H_p^{1+\delta})$ to (36). But by Theorem 5 we also know that u is a solution of (36) hence we have $\alpha = u$. The claims $Y_t^{t,x} = u(t, x)$ and $Z_t^{t,x} = \nabla u(t, x)$ are thus proved. \square

5. SOLUTION OF FBSDE (2)

5.1. **Some heuristic comments.** In this last section we study the forward-backward system (2) recalled again below for ease of reading:

$$(37) \quad \begin{cases} X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s dW_r, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) - \int_s^T Z_r^{t,x}dW_r + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr, \\ \forall s \in [t, T]. \end{cases}$$

We will go into more technical details in Section 5.2 and below, but first we want to make some heuristic comments on the link between the system above and the other FBSDE, given by (1).

If we were in the classical (and smooth enough) case where b is a suitable function, we would be able to change measure in (37) and apply Girsanov's theorem: We could find a new measure $\tilde{\mathbb{P}}$ defined by $d\tilde{\mathbb{P}} := M_T d\mathbb{P}$ under which $\tilde{W}_s := W_s + \int_0^s b(r, X_r^{t,x})dr$ is a Brownian motion. Here $M_s := \exp(-\int_0^s b(r, X_r) dW_r - \frac{1}{2} \int_0^s b^2(r, X_r) dr)$ is a martingale. Under the new measure $\tilde{\mathbb{P}}$, the system (37) would read

$$(38) \quad \begin{cases} \tilde{X}_s^{t,x} = x + \tilde{W}_s - \tilde{W}_t, \\ \tilde{Y}_s^{t,x} = \Phi(\tilde{X}_T^{t,x}) - \int_s^T \tilde{Z}_r^{t,x} d\tilde{W}_r + \int_s^T f(r, \tilde{X}_r^{t,x}, \tilde{Y}_r^{t,x}, \tilde{Z}_r^{t,x})dr \\ \quad + \int_s^T \tilde{Z}_r^{t,x} b(r, \tilde{X}_r^{t,x})dr, \\ \forall s \in [t, T], \end{cases}$$

which is exactly equation (1) mentioned above. In both cases the associated PDE would be the same, namely (3), recalled below

$$(39) \quad \begin{cases} u_t(t, x) + L^b u(t, x) + f(t, x, u(t, x), \nabla u(t, x)) = 0, \\ u(T, x) = \Phi(x), \\ \forall (t, x) \in [0, T] \times \mathbb{R}^d. \end{cases}$$

This can be easily checked by applying Itô's formula to $u(s, X_s^{t,x})$ (respectively $u(s, \tilde{X}_s^{t,x})$), and identifying Y and Z (respectively \tilde{Y} and \tilde{Z}) with u and ∇u calculated in X (respectively \tilde{X}).

The fact that the same PDE leads to two different FBSDEs can be interpreted analytically by looking at the PDE from two different viewpoints. On one hand we can look at the PDE and the semigroup generated by the Laplacian ($\frac{1}{2}\Delta$), which is also the generator of the forward component. In this case the process generated is a Brownian motion (which is X), so one gets to (38). Alternatively, we can look at the semigroup generated by the Laplacian *and* the term involving b (that is $L^b = \frac{1}{2}\Delta + (\nabla \cdot) b$), which is again the generator of the forward component, but in this case this process is a Brownian motion with drift, more specifically it is the solution of $\tilde{X}_s = x + \int_t^s b(r, \tilde{X}_r)dr + \int_t^s d\tilde{W}_r$. This second viewpoint leads to (37).

Clearly when the drift b is a distribution, this argument is no longer rigorous: We are not able to justify the change of measure (which would involve two measures which are not equivalent). From the analytical

point of view, it is unclear to us how to characterize the “semigroup” generated by L^b . We do not have answers to those questions yet.

What we achieve here instead, is an independent study of the system (37). We will define what a solution is, show its existence (but not uniqueness) and prove rigorously the link between the system (37) and the PDE (39).

5.2. The forward component X . It is easy to see that the forward-backward system (37) can be decoupled and the forward component solved first. We define a solution of (37) using both classical literature about weak solutions of FBSDEs (see for example [5, 7, 25]) and the notion of virtual solution for an SDE with distributional drift from [13]. Here the authors introduced and studied (in the special case where $t = 0$) equations in \mathbb{R}^d of the form

$$(40) \quad X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x}) dr + \int_t^s dW_r, \quad s \in [t, T]$$

with drift b being a distribution as specified in the standing assumption, with the extra L^q -condition that $b \in L^\infty([0, T]; H_q^{-\beta} \cap H_{\tilde{q}}^{-\beta})$, where q is as usual and $\tilde{q} := \frac{d}{1-\beta}$. In this Section we recall some of their results for the reader’s convenience. Notice that Lemma 23 is a new result.

To define a virtual solution we need to consider the following auxiliary PDE

$$(41) \quad \begin{cases} \xi_s(s, y) + L^b \xi(s, y) - (\lambda + 1)\xi(s, y) = -b(s, y), \\ \xi(T, y) = 0, \\ \forall (s, y) \in [0, T] \times \mathbb{R}^d. \end{cases}$$

This PDE is similar to (8) and can be treated with similar techniques. In [13, Theorem 14] the authors show that the PDE (41) admits a unique mild solution in $C([0, T], H_p^{1+\delta})$. This solution enjoys several smoothness properties and in particular it has a continuous version that can be evaluated pointwise and that will be used in the definition of virtual solution and in the construction of the auxiliary SDE below.

By *standard set-up* we mean a quintuple $(\Omega, \mathcal{F}, P, \mathbb{F}, (W_t)_t)$ where (Ω, \mathcal{F}, P) is a complete probability space, \mathbb{F} is a filtration satisfying the usual hypotheses and $W = (W_t)_t$ is an \mathbb{F} -Brownian motion. According to [13] we give the following definition.

Definition 20. [13, Definition 25] *A standard set-up $(\Omega, \mathcal{F}, P, \mathbb{F}, (W_t)_t)$ and a continuous stochastic process $X := (X_s^{t,x})_s$ on it are said to be a virtual solution of (40) if X is \mathbb{F} -adapted and the integral equation*

$$(42) \quad \begin{aligned} X_s^{t,x} &= x + \xi(t, x) - \xi(s, X_s^{t,x}) + (\lambda + 1) \int_t^s \xi(r, X_r^{t,x}) dr \\ &+ \int_t^s (\nabla \xi(r, X_r^{t,x}) + \mathbf{I}_d) dW_r, \end{aligned}$$

holds for all $s \in [t, T]$, \mathbb{P} -a.s.

To construct a virtual solution to (40) we transform (42) using the auxiliary PDE (41) and we get an auxiliary SDE (see equation (44) below) which we solve in the weak sense. Let us define $\varphi(s, y) := y + \xi(s, y)$ and let

$$(43) \quad \psi(s, \cdot) := \varphi^{-1}(s, \cdot)$$

be the inverse of $y \mapsto \varphi(s, y)$ for any fixed s , which is shown to exist and to be jointly continuous, see [13, Lemma 22]. Let V be the weak solution of the following auxiliary SDE

$$(44) \quad \begin{aligned} V_s^{t,x} &= v + (\lambda + 1) \int_t^s \xi(r, \psi(r, V_r^{t,x})) dr \\ &+ \int_t^s (\nabla \xi(r, \psi(r, V_r^{t,x})) + \mathbf{I}_d) dW_r, \end{aligned}$$

for $s \in [t, T]$, where \mathbf{I}_d is the $d \times d$ identity matrix and ξ is the solution of (41). Equation (44) is exactly [13, equation (34)], where the authors show that a unique weak solution exists. Then in [13, Theorem 28] the authors show existence and uniqueness of a virtual solution according to Definition 20 by making use of the weak solution of the SDE (44) with initial condition $v = \varphi(t, x) = x + \xi(t, x)$. This result is recalled in what follows.

Proposition 21. [13, Theorem 28] *Let Assumption 1 hold and let $b \in L^\infty([0, T], H_q^{-\beta} \cap H_{\tilde{q}}^{-\beta})$ where $\tilde{q} := \frac{d}{1-\beta}$. Then for every $x \in \mathbb{R}^d$ and $0 \leq t < T$, there exists a unique virtual solution of (40) which has the form $X_s^{t,x} = \psi(s, V_s^{t,x})$, where V is the unique weak solution of (44) and ψ is given by (43).*

Finally let us remark that, although the transformation ψ appearing in (44) involves a parameter λ not included in the original SDE for X , the virtual solution does not actually depend on λ . This is a consequence of [13, Proposition 29].

The next results are important in the proof of Theorem 25 below, when we approximate the coefficient b with a smooth sequence b^n . Let us denote by ψ_n, φ_n, ξ^n and V^n the same objects as above associated to equations (44) and (41) but with b replaced by a smooth sequence b^n . In this case it was shown in [13, Lemma 23 and Lemma 24, (iii)] that the following property holds:

Lemma 22. [13, Lemma 23 and Lemma 24, (iii)]

If $b^n \rightarrow b$ in $L^\infty([0, T], H_q^{-\beta} \cap H_{\bar{q}}^{-\beta})$, then $\xi^n \rightarrow \xi$ in $C([0, T], H_p^{1+\delta})$. Moreover, $\xi^n \rightarrow \xi$ and $\nabla \xi^n \rightarrow \nabla \xi$ uniformly in $[0, T] \times \mathbb{R}^d$.

For ψ_n and V^n , we have the following result.

Lemma 23. Let $b^n \rightarrow b$ in $L^\infty([0, T], H_q^{-\beta} \cap H_{\bar{q}}^{-\beta})$. Then

- (i) the functions ψ_n and ψ are jointly γ -Hölder continuous (for any $\gamma < 1 - \delta - \beta$) in the first variable and Lipschitz continuous in the second variable, uniformly in n , in particular there exists a constant $C > 0$ independent of n such that

$$(45) \quad |\psi_n(t, x) - \psi_n(s, y)| \leq C(|t - s|^\gamma + |x - y|).$$

- (ii) the moments of V^n can be controlled uniformly in n , in particular there exists a constant $C = C(p) > 0$ independent of n such that, for every $a > 2$,

$$(46) \quad \mathbb{E} [|V_t^n - V_s^n|^a] \leq C(|t - s|^a + |t - s|^{a/2}).$$

Proof. (i) Let $t, s > 0$ and $x, y \in \mathbb{R}^d$. Then

$$|\psi_n(t, x) - \psi_n(s, y)| \leq |\psi_n(t, x) - \psi_n(t, y)| + |\psi_n(t, y) - \psi_n(s, y)|.$$

The first term on the right hand side is bounded by $2|x - y|$ since $\sup_{(t,x)} |\nabla \psi_n(t, x)| < 2$ by [13, Lemma 24 (ii)]. The second term can be bounded with a similar proof as [13, Lemma 22, Step 3] and one gets

$$|\psi_n(t, y) - \psi_n(s, y)| \leq \frac{1}{2} |\psi_n(t, y) - \psi_n(s, y)| + |\xi^n(t, y) - \xi^n(s, y)|.$$

Using the fractional Morrey inequality (Lemma 2) we have

$$|\xi^n(t, y) - \xi^n(s, y)| \leq C \|\xi^n(t, \cdot) - \xi^n(s, \cdot)\|_{H_p^{1+\delta}} \leq C \|\xi^n\|_{C^{0,\gamma}} |t - s|^\gamma,$$

where $\|\xi^n\|_{C^{0,\gamma}} \leq C$ with C independent of n (proof similar to Lemma 11, (i)).

(ii) This bound is proven by similar arguments as in Step 3 in the proof of [13, Proposition 29], with the only difference that the exponent 4 is replaced by a for any $a > 2$. \square

5.3. Definition of solution for FBSDE and existence. Let us consider the virtual solution to the forward equation in (37), which is a *standard set-up* $(\Omega, \mathcal{F}, P, \mathbb{F}, (W_t)_t)$ and a process $(X_s^{t,x})_s$ that solves (42). We introduce the following definition.

Definition 24. A virtual-weak solution to the FBSDE (37) is a *standard set-up* $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, (W_t)_t)$ and a triplet of processes $(X^{t,x}, Y^{t,x}, Z^{t,x})$ such that

- $X^{t,x}, Y^{t,x}$ and $Z^{t,x}$ are \mathbb{F} -adapted, $X^{t,x}$ and $Y^{t,x}$ are continuous;
- $\mathbb{P} \left(|\Phi(X_T^{t,x})| + \int_0^T (|f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})| + |Z_s^{t,x}|^2) ds < \infty \right) = 1$;

- $(X^{t,x}, Y^{t,x}, Z^{t,x})$ verifies, \mathbb{P} -a.s.,

$$(47) \quad \begin{cases} X_s^{t,x} = x + \xi(t, x) - \xi(s, X_s^{t,x}) + (\lambda + 1) \int_t^s \xi(r, X_r^{t,x}) dr \\ \quad + \int_t^s (\nabla \xi(r, X_r^{t,x}) + \text{Id}) dW_r, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) - \int_s^T Z_r^{t,x} dW_r + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr, \\ \forall s \in [t, T]. \end{cases}$$

We conclude the paper with an existence result for the virtual-weak solution (X, Y, Z) of (37), which is constructed using the solution u to the PDE (8) and a Feynman-Kac representation for (Y, Z) .

Theorem 25. *Let Assumption 1 hold and let $b \in L^\infty([0, T], H_q^{-\beta} \cap H_q^{-\beta})$. Then a virtual-weak solution to the FBSDE system (37) is given by $(X^{t,x}, u(\cdot, X^{t,x}), \nabla u(\cdot, X^{t,x}))$, where $X^{t,x}$ is the unique virtual solution of (40) and u is the solution of PDE (8).*

Proof. In this proof we will drop the superscript t, x for shortness.

By Remark 3 and Theorem 5 there exists a unique mild solution to (8), which we denote by u . To prove that $(X, u(\cdot, X), \nabla u(\cdot, X))$ is a virtual-weak solution of (37) with X being the virtual solution of the forward component, it is enough to show that $(u(\cdot, X), \nabla u(\cdot, X))$ solves the backward component in (37) \mathbb{P} -a.s. Indeed, the first bullet point of Definition 24 is easy to check thanks to regularity of X and u , and the integrability conditions on f stated in the second bullet point of the Definition are fulfilled because f is Lipschitz continuous in (y, z) , bounded at $(t, x, 0, 0)$ uniformly in (t, x) and u and ∇u are uniformly bounded by Lemma 11; and $Z = \nabla u(\cdot, X)$ is square integrable because ∇u is uniformly bounded again by Lemma 11.

Let us denote by (X^n, Y^n, Z^n) the classical strong solution of the FBSDE

$$(48) \quad \begin{cases} X_s^n = x + \int_t^s b^n(r, X_r^n) dr + \int_t^s dW_r, \\ Y_s^n = \Phi(X_T^n) - \int_s^T Z_r^n dW_r + \int_s^T f(r, X_r^n, Y_r^n, Z_r^n) dr \end{cases}$$

in $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, (W_t)_t)$, where $b^n \in C([0, T]; C_b^1(\mathbb{R}^d; \mathbb{R}^d))$ such that $b^n \rightarrow b$ in $L^\infty([0, T]; H_q^{-\beta} \cap H_q^{-\beta})$. This strong solution X^n converges in law to X thanks to [13, Proposition 29]. Moreover we define

$$M_s^n := \int_t^s Z_r^n dW_r \quad \text{and} \quad F_s^n := \int_t^s f(r, X_r^n, Y_r^n, Z_r^n) dr$$

for any $t \leq s \leq T$. Note that from classical theory of BSDEs (see for example [9]) we have that $Y_s^n = u^n(s, X_s^n)$ and $Z_s^n = \nabla u^n(s, X_s^n)$.

We will show that there exists a subsequence of $(X^n, Y^n, Z^n, M^n, F^n, W)$ that converges in law to a limit vector and then we will identify this limit vector with the components of the solution of (37) and thus show that $(X, u(\cdot, X), \nabla u(\cdot, X))$ is a solution.

First, we prove the tightness of the sequence

$$\nu^n = (X^n, Y^n, Z^n, M^n, F^n, W)$$

in the space of continuous paths $C([0, T]; \mathbb{R}^{d'})$, where $d' = 2d + 3m + m \times d$. To do so, we use the following tightness criterion (see for example, [20, Corollary 16.9]): A sequence of stochastic processes $(\nu^n)_n$ with values in \mathbb{R}^d is tight in $C([0, T]; \mathbb{R}^d)$ if $(\nu_0^n)_n$ is tight and there exists $a, b, C > 0$ (independent of n) such that

$$\mathbb{E}[|\nu_r^n - \nu_s^n|^a] \leq C|r - s|^{1+b}.$$

First note that the initial condition ν_0^n is deterministic and it converges pointwise to ν_0 , hence it is tight. As for the other bound, we look for an estimate of the quantity

$$\begin{aligned} \mathbb{E}|\nu_r^n - \nu_s^n|^a &\leq C\mathbb{E}(|X_r^n - X_s^n|^a + |Y_r^n - Y_s^n|^a + |Z_r^n - Z_s^n|^a \\ &\quad + |M_r^n - M_s^n|^a + |F_r^n - F_s^n|^a + |W_r^n - W_s^n|^a), \end{aligned}$$

for $a > 2$, where the constant C depends only on a .

The first term is defined as $X_r^n = \psi_n(r, V_r^n)$. By Lemma 23 part (i) we get

$$\begin{aligned} |X_r^n - X_s^n|^a &= |\psi_n(r, V_r^n) - \psi_n(s, V_s^n)|^a \\ &\leq C(|V_r^n - V_s^n| + |r - s|^\gamma)^a \\ &\leq C(|V_r^n - V_s^n|^a + |r - s|^{a\gamma}), \end{aligned}$$

and by using Lemma 23 part (ii) we get

$$\begin{aligned} \mathbb{E}|X_r^n - X_s^n|^a &\leq C(\mathbb{E}|V_r^n - V_s^n|^a + |r - s|^{a\gamma}) \\ &\leq C(|r - s|^a + |r - s|^{a/2} + |r - s|^{a\gamma}) \\ &\leq C(|r - s|^{a/2} + |r - s|^{a\gamma}). \end{aligned}$$

Next we look at $\mathbb{E}|Y_r^n - Y_s^n|^a$, and using equation (18) from Lemma 11 we have

$$\begin{aligned} \mathbb{E}|Y_r^n - Y_s^n|^a &= \mathbb{E}|u^n(r, X_r) - u^n(s, X_s)|^a \\ &\leq C\mathbb{E}(|X_r - X_s| + |r - s|^\gamma)^a \\ &\leq C(\mathbb{E}|X_r - X_s|^a + |r - s|^{a\gamma}) \\ &\leq C(|r - s|^{a/2} + |r - s|^{a\gamma}). \end{aligned}$$

The third term $\mathbb{E}|Z_r^n - Z_s^n|^a$ is done similarly using equation (19) from Lemma 11 to get $\mathbb{E}|Z_r^n - Z_s^n|^a \leq C(|r - s|^{a\alpha/2} + |r - s|^{a\gamma})$.

Concerning the term involving M^n , using equation (17) from Lemma 11, we get

$$\begin{aligned} \mathbb{E}|M_r^n - M_s^n|^a &\leq C\mathbb{E} \left(\left| \int_t^s \nabla u^n(v, X_v^n) dW_v \right|^2 \right)^{a/2} \\ &\leq C\mathbb{E} \left(\int_s^r |\nabla u^n(v, X_v)|^2 dv \right)^{a/2} \\ &\leq C|r - s|^{a/2}. \end{aligned}$$

The last non-trivial term is

$$\mathbb{E}|F_r^n - F_s^n|^a = \mathbb{E} \left(\int_s^r |f(v, X_v, u^n(v, X_v), \nabla u^n(v, X_v))| dv \right)^a.$$

The function f inside the integral can be bounded using Assumption 1 as follows

$$\begin{aligned} & \sup_{(v,x)} |f(v, x, u^n(v, x), \nabla u^n(v, x))| \\ & \leq \sup_{(v,x)} |f(v, x, u^n(v, x), \nabla u^n(v, x)) - f(v, x, 0, 0)| + \sup_{v,x} |f(v, x, 0, 0)| \\ & \leq \sup_{(v,x)} C(1 + |u^n(v, x)| + |\nabla u^n(v, x)|) \\ & \leq C, \end{aligned}$$

where we have used equation (16) from Lemma 11. Thus

$$\mathbb{E}|F_r^n - F_s^n|^a \leq \mathbb{E} \left(\int_s^r C dv \right)^a \leq C|r - s|^a.$$

Putting everything together we have

$$\mathbb{E}|\nu_r^n - \nu_s^n|^a \leq C(|r - s|^{a/2} + |r - s|^{a\gamma} + |r - s|^a),$$

so choosing a big enough such that $\min\{a/2, a\gamma\} > 1$, then by the tightness criteria we have that ν^n is tight.

Next we want to identify the limit of $(X^n, Y^n, Z^n, M^n, F^n, W)$. Let us denote by ν one limit of ν^n (or of a subsequence) in $C([0, T]; \mathbb{R}^d)$, which exists by tightness shown as above. Note that the limit might not be unique. By Skorohod theorem there exists another probability space $(\tilde{\Omega}, \tilde{F}, \tilde{\mathbb{P}})$ and other random variables $\tilde{\nu}^n$ and $\tilde{\nu}$ on this space with values in $C([0, T]; \mathbb{R}^d)$ such that $\tilde{\nu}^n \rightarrow \tilde{\nu}$, $\tilde{\mathbb{P}}$ -a.s. and they have the same laws as the original random variables, in particular $\tilde{\mathbb{P}} \circ (\tilde{\nu}^n)^{-1} = \mathbb{P} \circ (\nu^n)^{-1}$ and $\tilde{\mathbb{P}} \circ (\tilde{\nu})^{-1} = \mathbb{P} \circ (\nu)^{-1}$.

Recall that for fixed n (some of) the components of the vector ν^n satisfy

$$Y_s^n = Y_t^n + M_s^n - F_s^n, \quad \mathbb{P}\text{-a.s.},$$

hence

$$\tilde{Y}_s^n = \tilde{Y}_t^n + \tilde{M}_s^n - \tilde{F}_s^n, \quad \tilde{\mathbb{P}}\text{-a.s.}.$$

Now taking the limit (along a subsequence) as $n \rightarrow \infty$ and by the $\tilde{\mathbb{P}}$ -almost sure convergence of $\tilde{\nu}^n$ to $\tilde{\nu}$ we get

$$\tilde{Y}_s = \tilde{Y}_t + \tilde{M}_s - \tilde{F}_s, \quad \tilde{\mathbb{P}}\text{-a.s.},$$

and since $\tilde{\mathbb{P}} \circ (\tilde{\nu})^{-1} = \mathbb{P} \circ (\nu)^{-1}$ we also have that the components of the limit vector ν satisfy

$$Y_s = Y_t + M_s - F_s, \quad \mathbb{P}\text{-a.s.}.$$

The last step in the proof consists in showing that the limiting components are of the desired form, for example that $M_s = \int_t^s Z_r dW_r$ etc.

We start by showing the convergence in law of $u^n(s, X_s^n) \rightarrow u(s, X_s)$. We do so by using the following result from [4, Section 3, Theorem 3.1]: Let (S, μ) be a metric space and let us consider S -valued random variables such that $\xi_n \rightarrow \xi$ in law and $\mu(\xi_n, \zeta_n) \rightarrow 0$ in probability. Then $\zeta_n \rightarrow \xi$ in law.

In the present case, on one hand we have that for any bounded and continuous functional $G : C([0, T]; \mathbb{R}^m) \rightarrow \mathbb{R}$, then $G \circ u$ is also bounded and continuous because u is uniformly continuous by equation (18) from Lemma 11. Hence by weak convergence of $X^n \rightarrow X$ we obtain weak convergence of $G(u(\cdot, X^n)) \rightarrow G(u(\cdot, X))$, that is $u(\cdot, X^n) \rightarrow u(\cdot, X)$ in law. On the other hand $u^n(\cdot, X^n) - u(\cdot, X^n) \rightarrow 0$ in $C([0, T]; \mathbb{R}^m)$, \mathbb{P} -a.s., because $u^n \rightarrow u$ uniformly by Lemma 9 part (ii), hence $|u^n(\cdot, X^n) - u(\cdot, X^n)| \rightarrow 0$ in probability. These two facts imply the convergence in law of $u^n(s, X_s^n) \rightarrow u(s, X_s)$ by [4, Section 3, Theorem 3.1]. A similar argument can be applied to $\nabla u^n(s, X_s^n) \rightarrow \nabla u(s, X_s)$ by using equation (19) instead of (18).

Similarly as above, one can see that the convergence in law means that the components Y and Z in the limit vector ν satisfy $Y_s = u(s, X_s)$ and $Z_s = \nabla u(s, X_s)$ \mathbb{P} -a.s. in $C([0, T]; \mathbb{R}^m)$ and $C([0, T]; \mathbb{R}^{m \times d})$, since $Y_s^n = u^n(s, X_s^n)$ and $Z_s^n = \nabla u^n(s, X_s^n)$.

For the component F , we use the continuity assumption of f in (x, y, z) and the continuity of u and ∇u in x to show that the map

$$X^n \mapsto \int_t^\cdot f(r, X_r^n, u(r, X_r^n), \nabla u(r, X_r^n)) dr$$

composed with any bounded and continuous functional $G : C([0, T]; \mathbb{R}^m) \rightarrow \mathbb{R}$ is still bounded and continuous, hence we have that

$$\begin{aligned} & \mathbb{E} \left[G \left(\int_t^\cdot f(r, X_r^n, u(r, X_r^n), \nabla u(r, X_r^n)) dr \right) \right] \\ & \rightarrow \mathbb{E} \left[G \left(\int_t^\cdot f(r, X_r, u(r, X_r), \nabla u(r, X_r)) dr \right) \right] \end{aligned}$$

from the weak convergence of $X^n \rightarrow X$. Moreover the convergence in probability of $|u^n(\cdot, X^n) - u(\cdot, X^n)| \rightarrow 0$ in $C([0, T]; \mathbb{R}^m)$ and the Lipschitz character of f imply that

$$\begin{aligned} & \int_t^\cdot |f(r, X_r^n, u^n(r, X_r^n), \nabla u^n(r, X_r^n)) - f(r, X_r^n, u(r, X_r^n), \nabla u(r, X_r^n))| dr \\ & \leq L \int_t^\cdot (|u^n(r, X_r^n) - u(r, X_r^n)| + |\nabla u^n(r, X_r^n) - \nabla u(r, X_r^n)|) dr \\ & \leq C \int_t^\cdot |u^n(r, X_r^n) - u(r, X_r^n)| dr \rightarrow 0 \end{aligned}$$

in probability in $C([0, T]; \mathbb{R}^m)$. Hence applying again [4, Section 3, Theorem 3.1] we obtain that $\int_t^s f(r, X_r^n, u^n(r, X_r^n), \nabla u^n(r, X_r^n))dr$ converges to $\int_t^s f(r, X_r, u(r, X_r), \nabla u(r, X_r))dr$ in law. Thus, for the component F of the limit vector ν we have that

$$F_s = \int_t^s f(r, X_r, u(r, X_r), \nabla u(r, X_r))dr = \int_t^s f(r, X_r, Y_r, Z_r)dr,$$

\mathbb{P} -a.s..

It remains to show that $M_s = \int_t^s Z_r dW_r$, \mathbb{P} -a.s. This follows from [23, Theorem 7.10] (see also [10, Section 2.2]) and from the fact that $Z^n \rightarrow Z$ weakly.

Putting everything together and using the fact that

$$Y_t = Y_T - \int_t^T Z_r dW_r + \int_t^T f(r, X_r, Y_r, Z_r)dr$$

we have

$$\begin{aligned} Y_s &= Y_t + \int_t^s Z_r dW_r - \int_t^s f(r, X_r, Y_r, Z_r)dr \\ &= Y_T - \int_s^T Z_r dW_r + \int_s^T f(r, X_r, Y_r, Z_r)dr, \quad \mathbb{P}\text{-a.s.}, \end{aligned}$$

where $Y_s = u(s, X_s)$ and $Z_s = \nabla u(s, X_s)$, as wanted. \square

Acknowledgement: The authors would like to thank Tiziano De Angelis for various suggestions and fruitful discussions on this topic, and the anonymous referees for reading the manuscript and providing useful hints that led to several improvements.

REFERENCES

- [1] F. Antonelli. Backward-forward stochastic differential equations. *Ann. Appl. Probab.*, 3(3):777–793, 1993.
- [2] F. Antonelli and J. Ma. Weak solutions of forward-backward SDE's. *Stoch. Anal. Appl.*, 21(3):493–514, 2003.
- [3] V. Bally and A. Matoussi. Weak solutions for spdes and backward doubly stochastic differential equations. *Journal of Theoretical Probability*, 14(1):125–164, 2001.
- [4] P. Billingsley. *Convergence of Probability Measures*. Wiley series in probability and statistics. Wiley, 2nd edition, 1999.
- [5] R. Buckdahn, H.-J. Engelbert, and A. Răşcanu. On weak solutions of backward stochastic differential equations. *Teor. Veroyatn. Primen.*, 49(1):70–108, 2004.
- [6] E. B. Davies. *Heat kernels and spectral theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.
- [7] F. Delarue and G. Guatteri. Weak existence and uniqueness for forward-backward SDEs. *Stoch. Proc. Appl.*, 116:1712–1742, 2006.
- [8] J. Diehl and J. Zhang. Backward stochastic differential equations with Young drift. *Probab. Uncertain. Quant. Risk*, 2:Paper No. 5, 17, 2017.
- [9] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

- [10] H-J. Engelbert and G. Peskir. Stochastic differential equations for sticky Brownian motion. *Stochastics: An International Journal of Probability and Stochastic Reports*, 86(6):993–1021, 2014.
- [11] M. Erraoui, Y. Ouknine, and A. Sbi. Backward stochastic differential equations with distribution as terminal condition. *Random Operators and Stochastic Equations*, 5(4):349–356, 1997.
- [12] M. Erraoui, Y. Ouknine, and A. Sbi. Reflected solutions of backward stochastic differential equations with distribution as terminal condition. *Random Operators and Stochastic Equations*, 6(1):1–16, 1998.
- [13] F. Flandoli, E. Issoglio, and F. Russo. Multidimensional SDEs with distributional coefficients. *Trans. Amer. Math. Soc.*, 369:1665–1688, 2017.
- [14] K. Gawedzki. Stochastic processes in turbulent transport. ArXiv.0806.1949, 2008.
- [15] F. Gozzi and F. Russo. Weak Dirichlet processes with a stochastic control perspective. *Stochastic Processes and their Applications*, 116 (11):1563–1583, 2006.
- [16] M. Hinz, E. Issoglio, and M. Zähle. Elementary pathwise methods for nonlinear parabolic and transport type spde with fractal noise. *Modern Stochastics and Applications - conference proceedings*, 2013.
- [17] Y. Hu and G. Tessitore. BSDE on an infinite horizon and elliptic PDEs in infinite dimension. *Nonlinear Differential Equations and Applications NoDEA*, 14(5):825–846, 2007.
- [18] E. Issoglio. Transport equations with fractal noise - existence, uniqueness and regularity of the solution. *J. Analysis and its App.*, 32(1):37–53, 2013.
- [19] E. Issoglio and M. Zähle. Regularity of the solutions to SPDEs in metric measure spaces. *Stochastic Partial Differential Equations: Analysis and Computations*, 3 (2):272–289, 2015.
- [20] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, 2nd edition, 2001.
- [21] I. Kharroubi, N. Langrenè, and H. Pham. Discrete time approximation of fully nonlinear hjb equations via bsdes with nonpositive jumps. *Ann. Appl. Probab.*, 25(4):2301–2338, 08 2015.
- [22] R. H. Kraichnan. Small scale structure of a scalar field convected by turbulence. *Physics of Fluids*, 11(5):945–953, 1968.
- [23] T. G. Kurtz and P. E. Protter. Weak convergence of stochastic integrals and differential equations. *Lecture Notes in Math.*, 1627:1–41, 1996.
- [24] A. Lejay. A probabilistic representation of the solution of some quasi-linear PDE with a divergence form operator. application to existence of weak solutions of FBSDE. *Stoch. Proc. Appl.*, 110(1):145–176, 2004.
- [25] J. Ma, J. Zhang, and Z. Zheng. Weak solutions for forward-backward SDEs – a martingale problem approach. *Ann. Probab.*, 36(6):2092–2125, 2008.
- [26] C. Pagani. Functional renormalization group approach to the Kraichnan model. *Phys. Rev. E*, 92:033016, 2015.
- [27] É. Pardoux and S. G. Peng. Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.
- [28] É. Pardoux and S. G. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. *Rożwskii B L, Sowers R B, eds. Stochastic Partial Differential Equations and Their Applications. Lect Notes Control Inf Sci.*, 176:200–217, 1992.
- [29] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

- [30] T. Runst and W. Sickel. *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*, volume 3 of *de Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter & Co., Berlin, 1996.
- [31] F. Russo and L. Wurzer. Elliptic PDEs with distributional drift and backward SDEs driven by a càdlàg martingale with random terminal time. *Stoch. Dyn.*, 17(4):1750030 (36 pages), 2017.
- [32] H. Triebel. *Interpolation theory, function spaces, differential operators*. North-Holland Amsterdam, Heidelberg, first edition, 1978.
- [33] B. S. Tsirel'son. An example of a stochastic differential equation having no strong solution. *Theor. Prob. Appl.*, 20:427–430, 1975.