Strong solutions of forward-backward stochastic differential equations with measurable coefficients

Peng Luo , Olivier Menoukeu-Pamen , Ludovic Tangpi

October 26, 2021

ABSTRACT

This paper investigates solvability of fully coupled systems of forward-backward stochastic differential equations (FBSDEs) with irregular coefficients. In particular, we assume that the coefficients of the FBSDEs are merely measurable and bounded in the forward process. We crucially use compactness results from the theory of Malliavin calculus to construct strong solutions. Despite the irregularity of the coefficients, the solutions turn out to be differentiable, at least in the Malliavin sense and, as functions of the initial variable, in the Sobolev sense.

KEYWORDS: Singular PDEs, Sobolev regularity, FBSDE, singular coefficients, strong solutions, Malliavin calculus.

MSC 2010: 60H10, 34F05, 60H07, 35D40

1. Introduction

The main result of this work concerns the existence of a (strong) solution of the forward-backward stochastic differential equation (FBSDE)

$$\begin{cases} X_t = x + \int_0^t b(u, X_u, Y_u, Z_u)\,\mathrm{d}u + \int_0^t \sigma\, dW_u \\ Y_t = h(X_T) + \int_t^T g(u, X_u, Y_u, Z_u)\mathrm{d}u - \int_t^T Z_u\, dW_u \end{cases} \tag{1.1}$$

with b, g and h measurable in (t, x), and uniformly continuous in (y, z), see Theorem 2.1. The proof of this result is partly inspired from results by Ma and Zhang [42] and Delarue and Guatteri [17] on *weak* solutions of FBSDE under similar conditions. Our contribution in this direction is to obtain *strong* solutions and allow irregularity of h. Because of the lack of regularity of the coefficients, usual fixed point and Picard iterations techniques cannot be applied here. Let us briefly describe our method:

We start as in [17, 42] by approximating the functions b, g and h by smooth function, e.g. by mollification. The FBSDE associated to these functions admit unique solutions (X^n, Y^n, Z^n) and a so-called decoupling field v_n which is the classical solution of an associated quasilinear PDE. The function v_n is called a decoupling field because it holds

$$Y^n_t = v_n(t, X^n_t) \quad \text{and} \quad Z^n_T = D_x v_n(t, X^n_t)\sigma, \tag{1.2}$$

which allows to decouple the system. The problem is now to derive strong limits for the above sequences and to show that these limits satisfy the desired equation. Using classical a priori estimations for such equations, (see e.g. [37] or the statements recalled in the Appendix) it can be shown that for every $\delta > 0$ and every $t \in [0, T - \delta]$ the sequence of functions v_n admits some compactness properties allowing to derive a limit v for v_n and a limit w for $D_x v_n$. When h is sufficiently regular, say Hölder continuous, δ can be taken equal to zero. In this setting, the idea of [17, 42] is to also gain sufficiently good control over the time-derivative and the Hessian using e.g. Calderon-Zygmund theory. The approach proposed here is to rather use ideas from Malliavin calculus, notably the compactness principle due to Da Prato et al. [13], to find a limit X of the sequence (X^n) in the strong sense. Together with the representation (1.2), this allows to find strong limits for Y and Z (at least for t small enough). It remains to verify that the limiting processes (X, Y, Z) actually solve the desired equation.

We further study regularity properties of solutions. In fact, despite the singularity of the coefficients, it turns out that the solutions enjoy satisfactory regularity, at least in the Malliavin and Sobolev sense. These are interesting results in that, the convention in the field is that solutions inherit the regularity properties of the coefficients [3, 41].

FBSDEs are an essential tool in the investigation of stochastic control problems and stochastic differential games. Due to Pontryagin's stochastic maximum principle, they can be used to characterize optimal controls and Nash equilibriums [11, 38, 55]. These equations also provide a probabilistic approach to deal with quasilinear parabolic partial differential equations via the nonlinear Feynman-Kac formula initiated by Pardoux and Peng [53] and further developed notably in [7, 16, 33, 54]. As a result, FBSDEs have received a lot of attention in the applied probability community and appear in various applications, we refer for instance to [12, 21, 23, 24, 49] and the references therein. When the coefficients of the equations, i.e. the functions b, g and h are sufficiently smooth, solvability of (1.1) is well-understood. Refer for instance to [15, 43, 56] for the case of equations with Lipschitz continuous coefficients and to [36, 41] for locally Lipschitz coefficients. When the coefficients are not regular enough, while an SDEs theory is well-developed (see e.g. [5, 34, 46, 47, 50]) BSDEs with irregular coefficients are less well-studied. A notable exception is the notion of *weak solution* of FBSDE (very analogous to weak solutions of SDEs) introduced by Buckdahn and Engelbert [10] and further investigated in [17, 42, 44]. These solutions are constructed on a probability space that is possibly different from the underlying probability space. On the other hand, more recently, Issoglio and Jing [29] studied two new classes of multidimensional FBSDEs with distributional coefficients. In many applications, for instance to the construction of feedback solutions of stochastic control problems, it is important to have *strong solutions*, and to analyze regularity properties thereof. More details on such applications to stochastic control theory are given in subsection 4.2.

The remainder of the paper is organized as follows: In the next section, we make precise the mathematical setting of the work and state the main results: Existence of strong solutions for FBSDEs with rough coefficients. The proof is given in Section 3.1. The regularity of the solutions of the FBSDE is analyzed in Section 3.2. We consider both regularity in the Malliavin (variational) sense and in the Sobolev sense.

2. Setting and main results

Let $T \in (0, \infty)$ and $d \in \mathbb{N}$ be fixed and consider a probability space $(\Omega, \mathcal{F}, P)$ equipped with the completed filtration $(\mathcal{F}_t)_{t\in[0,T]}$ of a d-dimensional Brownian motion W. Throughout the paper, the product $\Omega \times [0, T]$ is endowed with the predictable σ-algebra. Subsets of $\mathbb{R}^k$, $k \in \mathbb{N}$, are always endowed with the Borel σ-algebra induced by the Euclidean norm $|\cdot|$. Let us consider the following conditions:

(A1) The function $b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^{l\times d} \to \mathbb{R}^d$ is Borel measurable and it holds

$$|b(t, x, y, z)| \leq k_1(1 + |y|)$$

for some $k_1 \geq 0$ and every $(x, y, z) \in \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^{l\times d}$. Moreover, for each fixed (t, x) the restriction of $b(t, x, \cdot, \cdot)$ to $B_R(0) \times \mathbb{R}^{l\times d}$ is continuous, with $R := k_3 e^{Tk_2}$ and $B_R(0) := \{(y, z) : |y| \leq R\}$.

(A2) $\sigma \in \mathbb{R}^{d\times d}$ and $\xi\sigma\sigma^*\xi > \Lambda|\xi|^2$ for some $\Lambda > 0$ and for all $\xi \in \mathbb{R}^d$. Here * stands for the transpose of a matrix.

(A3) The function $g : [0, T] \times \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^{l\times d} \to \mathbb{R}^l$ is measurable, uniformly continuous in (y, z), uniformly in $(t, x) \in [0, T] \times \mathbb{R}^d$ and satisfies

$$|g(t, x, y, z)| \leq k_2(1 + |y| + |z|)$$

for some $k_2 \geq 0$, and for every $(t, x, y, z) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^{l\times d}$.

(A4) The function $h : \mathbb{R}^d \to \mathbb{R}^l$ is measurable and satisfies

$$|h(x)| \leq k_3$$

for some $k_3 \geq 0$ and for every $x \in \mathbb{R}^d$.

The following is our first main result: In its statement, the space $\mathcal{S}^2(\mathbb{R}^d) \times \mathcal{S}^2(\mathbb{R}^l) \times \mathcal{H}^2(\mathbb{R}^{l\times d})$ is defined as follows: For $p \in [1, \infty]$ and $k \in \mathbb{N}$, denote by $\mathcal{S}^p(\mathbb{R}^k)$ the space of all adapted continuous processes X with values in $\mathbb{R}^k$ such that $\|X\|^p_{\mathcal{S}^p(\mathbb{R}^k)} := E[(\sup_{t\in[0,T]} |X_t|)^p] < \infty$, and by $\mathcal{H}^p(\mathbb{R}^k)$ the space of all progressively measurable processes Z with values in $\mathbb{R}^k$ such that $\|Z\|_{\mathcal{H}^p(\mathbb{R}^k)} := E[(\int_0^T |Z_u|^2 \, \mathrm{d}u)^{p/2}] < \infty$.

Theorem 2.1. *Assume that the conditions (A1)-(A4) hold and that one of the following assumptions is satisfied:*

(B1) *b and g are bounded in z, i.e. $|b(t, x, y, z)| + |g(t, x, y, z)| \leq C(1 + |y|)$ for all t, x, y, z for some $C \geq 0$.*

(B2) *h is Lipschitz continuous: $|h(x) - h(x')| \leq k_3|x - x'|$ for every $x, x' \in \mathbb{R}^d$.*

Then the FBSDE (1.1) admits a solution $(X, Y, Z) \in \mathcal{S}^2(\mathbb{R}^d) \times \mathcal{S}^2(\mathbb{R}^l) \times \mathcal{H}^2(\mathbb{R}^{l\times d})$ such that

$$Y_t = v(t, X_t), \quad Z_t = w(t, X_t)\sigma \quad P \otimes dt\text{-a.s.}$$

for some measurable functions $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^l$ and $w : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{l\times d}$.

The proof of Theorem 2.1 will be given in Subsection 3.1 where we further prove regularity of the solutions both in the classical and variational sense. In particular, we will show that the solution (X, Y) is Malliavin differentiable, and if the generator g does not depend on x or does not depend on z, then (X, Y, Z) is Malliavin differentiable. Moreover, looking at (X, Y) as functions of the initial value of the forward process, with Probability one, this function belongs to a weighted Sobolev space. See Propositions 3.4 and 3.5 for details.

3. FBSDEs with measurable coefficients

3.1. Proof of Theorem 2.1

This section is entirely dedicated to the proof of Theorem 2.1. Throughout, the conditions (A1)-(A4) are in force. Let (b_n), (g_n) and (h_n) be sequences of smooth functions with compact support converging pointwise to b, g and h, respectively (e.g. obtained by standard mollification). We can assume without loss of generality that for each n, the functions h_n and g_n satisfy (A3)-(A4) in addition to being smooth and Lipschitz continuous (but with Lipschitz constant possibly depending on n). These sequences will be used throughout the proof. We begin with the following simple lemma which shows that the sequence b_n can be chosen so that the convergence holds uniformly on a given compact in (y, z) and g_n such that the convergence holds locally uniformly in (y, z). The lemma is well-known, we provide a proof since we did not find a directly citable reference. This will be needed at the end of the proof of the theorem.

Lemma 3.1. *The sequence of mollifiers (g_n) converge to g almost surely in (t, x) and locally uniformly in (y, z). That is, for every t, x and every compact set $K \subseteq \mathbb{R}^l \times \mathbb{R}^{l\times d}$ it holds that*

$$\lim_{n\to\infty} \sup_{(y,z)\in K} |g_n(t, x, y, z) - g(t, x, y, z)| = 0.$$

Similarly, (b_n) converges to b almost surely in (t, x) and locally uniformly in $(y, z) \in B_R(0) \times \mathbb{R}^{l\times d}$.

Proof. We prove the result for some generic function f since the proof for g and b is the same. Moreover, since the dimension does not matter, we let for simplicity $f : \mathbb{R}^d \times \mathbb{R}^l$ be a measurable function that is uniformly continuous in its second component. We consider the standard mollifiers ϕ^1_n, ϕ^2_n respectively defined on $\mathbb{R}^d$ and $\mathbb{R}^l$ as

$$\phi^1_n(\cdot) \quad = c_1 n^d \varphi(n| \cdot |), \quad \phi^2_n(\cdot) = c_2 n^l \varphi(n| \cdot |)$$

where for any $x \in \mathbb{R}^d, \varphi(x) = \exp(1/(|x|^2-1))1_{[0,1]}(|x|)$ and c_1, c_2 are two normalizing constants such that $\int_{\mathbb{R}^d}\phi_n^1\,dx = \int_{\mathbb{R}^l}\phi_n^2\,dx = 1$. Recall that the mollication f_n is defined as

$$f_n(x,y) := \int\limits_{\mathbb{R}^d\times\mathbb{R}^l} f(x-x',y-y')\phi_n^1(x')\phi_n^2(y')dx'dy'.$$

It is well-known that f_n converges to f almost surely. Let us show local uniform convergence in the second variable. Let $\mathcal{N}\subseteq\mathbb{R}^d$ bet a set of measure zero on the complement of which f_n converges to f, and let $K\subseteq\mathbb{R}^l$ be a compact set. Observe that for each n, the support of ϕ_n^1, ϕ_n^2 is in the closure of the ball of radius $1/n$. On the other hand, given $\varepsilon>0$, by uniform continuity of $f(x,\cdot)$, $x\in\mathcal{N}^c$, there is $\eta>0$ such that for $y,y'\in\mathbb{R}^l$ satisfying $|y-y'|\le\eta$, it holds that $|f(x,y)-g(x,y')|<\varepsilon$. Let $n\in\mathbb{N}$ be sufficiently large. We have

$$\begin{aligned}\sup_{y\in K}|f_n(x,y)-f(x,y)| &= \sup_{y\in K}\Big|\int\limits_{B_{1/n}(0)\times B_{1/n}(0)}\big(f(x',y')-f(x,y)\big)\phi_n^1(x-x')\phi_n^2(y-y')\,dx'\,dy'\Big|\\ &\le \sup_{y\in K}\Big|\int\limits_{B_{1/n}(0)\times B_{1/n}(0)}\big(f(x',y')-f(x',y)\big)\phi_n^1(x-x')\phi_n^2(y-y')\,dx'\,dy'\Big|\\ &\quad+\sup_{y\in K}\Big|\int\limits_{B_{1/n}(0)\times B_{1/n}(0)}\big(f(x',y)-f(x,y)\big)\phi_n^1(x-x')\phi_n^2(y-y')\,dx'\,dy'\Big|\\ &\le \varepsilon+\sup_{y\in K}\Big|\int\limits_{B_{1/n}(0)}\big(f(x',y)-f(x,y)\big)\phi_n^1(x-x')\,dx'\Big|.\end{aligned}$$

For each n, there is $y_n\in K$ such that

$$\sup_{y\in K}|f_n(x,y)-f(x,y)|\le\varepsilon+\Big|\int\limits_{B_{1/n}(0)}\big(f(x',y_n)-f(x,y_n)\big)\phi_n^1(x-x')\,dx'\Big|+1/n.$$

Since K is compact, up to a subsequence we can assume that y_n converges to some $\bar{y}\in K$. Thus, taking n large enough such that $|y_n-\bar{y}|\le\eta$ we have

$$\begin{aligned}\sup_{y\in K}|f_n(x,y)-f(x,y)| &\le \varepsilon+\Big|\int\limits_{B_{1/n}(0)}\big(f(x',y_n)-f(x',\bar{y})\big)+f(x',\bar{y})-f(x,y_n)\big)\phi_n^1(x-x')\,dx'\Big|+1/n\\ &\le 2\varepsilon+|f_n(x,\bar{y})-f(x,y_n)|+1/n.\end{aligned}$$

Taking the limit as n goes to infinity and ε goes to zero to conclude. □

Step 1: Construction of an approximating sequence of solutions. Let $n\in\mathbb{N}$ be fixed. According to [15, Theorem 2.6], for every $(s,x)\in[0,T]\times\mathbb{R}^d$ the FBSDE

$$\begin{cases}X_t = x+\int_s^t b_n(u,X_u,Y_u,Z_u)\,\mathrm{d}u+\int_s^t\sigma\,dW_u\\ Y_t = h_n(X_T)+\int_t^T g_n(u,X_u,Y_u,Z_u)\mathrm{d}u-\int_t^T Z_u\,dW_u \quad t\in[s,T]\end{cases}\tag{3.1}$$

admits a unique solution $(X^{s,x,n},Y^{s,x,n},Z^{s,x,n})\in\mathcal{S}^2(\mathbb{R}^d)\times\mathcal{S}^\infty(\mathbb{R}^l)\times\mathcal{H}^2(\mathbb{R}^{l\times d})$. Denote by $\mathcal{L}^n$ the differential operator

$$\mathcal{L}^n v := b_n(t,x,v,D_xv\sigma)D_xv+\frac{1}{2}\mathrm{trace}(\sigma\sigma^*D_{xx}v),$$

where D_x and D_{xx} denote the first and second derivatives acting on the space variable.

By [37, Theorem VII.7.1] (or see also [43, Proposition 3.3]) the PDE

$$\begin{cases} \partial_t v_n(t,x) + \mathcal{L}^n v_n(t,x) + g_n(t,x,v_n(t,x),D_x v_n(t,x)\sigma) = 0 \\ v_n(T,x) = h_n(x) \end{cases} \tag{3.2}$$

admits a unique (classical) solution $v_n \in C^{1,2}([0,T]\times\mathbb{R}^d)$ that is bounded and with bounded gradient. Moreover, the solutions of (3.2) and (3.1) are linked through the identities (see [43])

$$Y_t^{s,x,n} = v_n(t,X_t^{s,x,n}) \quad \text{and} \quad Z_t^{s,x,n} = D_x v_n(t,X_t^{s,x,n})\sigma, \quad t\in[s,T]. \tag{3.3}$$

The rest of the proof will consist in proving (strong) convergence of the above defined sequence of stochastic processes $(X^{s,x,n},Y^{s,x,n},Z^{s,x,n})$ and to verify that the limiting process satisfies the FBSDE with measurable drift. Our method will make use of a priori (gradient) estimates for Sobolev solutions of parabolic quasilinear PDEs which can be found e.g. in [16] or [37] and that we recall in the Appendix. These estimates allow us to have:

Lemma 3.2. *There is a constant $R>0$ depending only on k_2,k_3,T (but not on n) and for every $\delta>0$, $\alpha\in(0,1)$ there are constants C_δ and $C_{\alpha,\delta}$ depending on k_1,k_2,k_3,σ,d,l and T, and which do not depend on n such that*

$$|v_n(t,x)| \le R \quad \textit{for all} \quad (t,x)\in[0,T]\times\mathbb{R}^d,$$

the derivatives satisfy

$$|D_x v_n(t,x)| \le C_\delta \quad \textit{for every} \quad (t,x)\in[0,T-\delta]\times\mathbb{R} \tag{3.4}$$

and if h is α-Hölder continuous, then

$$|v_n(t,x)-v_n(t',x')| \le C(|t-t'|^{\alpha'/2}+|x-x'|^{\alpha'}) \tag{3.5}$$

for every $(t,x),(t',x')\in[0,T]\times\mathbb{R}^d$ and some $\alpha'\in(0,\alpha]$. Moreover, if h is Lipschitz continuous, then (3.4) holds with $\delta=0$.

Proof. The boundedness of v_n is well-known. We provide it to explicitly derive the constant R. We have

$$v_n(t,x) = Y_t^{t,x,n} = h_n(X_T^{t,x,n}) + \int_t^T\int_0^1 \partial_z g_n(u,X_u^{t,x,n},Y_u^{t,x,n},\lambda Z_u^{t,x,n})\,d\lambda Z_u^{t,x,n}\,du - \int_t^T Z_u^{t,x,n}\,dW_u$$
$$+\int_t^T g_n(u,X_u^{t,x,n},Y_u^{t,x,n},0)\,du.$$

Therefore, by the Girsanov's theorem, conditions (A3)-(A4) and Gronwall's inequality we have

$$|v_n(t,x)| \le k_3 e^{Tk_2} = R \quad \text{for all} \quad (t,x)\in[0,T]\times\mathbb{R}^d.$$

The bounds (3.4) and (3.5) follow by Theorem A.1. Furthermore, since v_n is a classical solution of (3.2), i.e. $v_n \in C^{1,2}([0,T]\times\mathbb{R}^d)$, it is in particular a Sobolev solution, and $v_n\in W^{1,2}_{d+1,\mathrm{loc}}([0,T]\times\mathbb{R}^d,\mathbb{R}^l)$ (see definition in Appendix). Moreover, if h is Lipschitz continuous then by definition of (h_n), it holds $|h_n(x)-h_n(x')|\le k_3|x-x'|$ for every $x,x'\in\mathbb{R}^d$ and all $n\in\mathbb{N}$. Therefore, the last claims follow by Theorem A.2. □

Step 2: Candidate solution for the forward equation. In this step, we show that the sequence $(X^{s,x,n})$ converges in the strong topology of $\mathcal{S}^2(\mathbb{R}^d)$. We first show existence of a weak limit. To ease the presentation, we omit the superscript (s,x) and put

$$X^n := X^{s,x,n}, \quad Y^n := Y^{s,x,n} \quad \text{and} \quad Z^n := Z^{s,x,n}.$$

Step 2a: Weak limit. It follows from Step 1 that the process X^n satisfies the forward SDE

$$X_t^n = x + \int_s^t b_n(u, X_u^n, v_n(u, X_u^n), D_x v_n(u, X_u^n)\sigma)\, du + \int_s^t \sigma dW_u. \tag{3.6}$$

Lemma 3.3. *Consider the function* $\widetilde{b}_n : (t,x) \mapsto b_n(t, x, v_n(t,x), D_x v_n(u, X_u^n)\sigma)$. *Under either of the conditions (B1) or (B2), the function* $\widetilde{b}_n$ *is continuously differentiable and uniformly bounded, i.e. there is a constant* $C \geq 0$ *which does not depend on* n *such that*

$$|\widetilde{b}_n(t,x)| \leq C \quad \textit{for all} \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

Proof. That $\widetilde{b}_n$ is continuously differentiable follows from the fact that b_n is smooth and v_n is twice continuously differentiable. By (A1) and Lemma 3.2, if condition (B1) holds, then for every $(t,x) \in [0,T] \times \mathbb{R}^d$ we have

$$\begin{aligned} |\widetilde{b}_n(t,x)| &\leq k_1(1 + |v_n(t,x)|) \\ &\leq k_1(1 + C). \end{aligned}$$

When condition (B2) holds, it follows by Lemma 3.2 that $D_x v_n$ is bounded. Thus the result follows from the linear growth of b, i.e. (A1). □

Due to Lemma 3.3, it follows from standard SDE estimates that the sequence (X^n) satisfies

$$\sup_n E\Big[\sup_{t\in[s,T]} |X_t^n|^2\Big] < \infty.$$

Therefore (X^n) admits a subsequence which converges to some $\widetilde{X}$ in the weak topology of $\mathcal{S}^2(\mathbb{R}^d)$. This subsequence will be denoted again (X^n).

Step 2b: Strong limit. Since $\widetilde{b}_n$ is Lipschitz continuous, the solution X^n of the SDE (3.6) is Malliavin differentiable and since $\widetilde{b}_n$ is a smooth function with compact support, it follows by [47, Lemma 3.5] that

$$E\Big[\Big\|D_{t'}^i X_r^n - D_t^i X_r^n\Big\|^2\Big] \leq \mathcal{C}_{d,T}(\|\widetilde{b}_n\|_\infty)|t-t'|^\alpha \leq \mathcal{C}_{d,T}(\|\widetilde{b}\|_\infty)|t-t'|^\alpha \tag{3.7}$$

and

$$\sup_{0\leq t\leq T} E\Big[\Big\|D_t X_r^n\Big\|^2\Big] \leq \mathcal{C}_{d,T}(\|\widetilde{b}_n\|_\infty) \leq \mathcal{C}_{d,T}(\|\widetilde{b}\|_\infty) \tag{3.8}$$

for a strictly positive constant $\mathcal{C}_{d,T}(\|\widetilde{b}_n\|_\infty)$ such that $\mathcal{C}_{d,T}$ is a continuous increasing function, and with $\alpha = \alpha(r) > 0$. Whereby, $D_t^i X$ denotes the Malliavin derivative of the random variable X at time t in the direction of the Brownian motion W^i and $\|x\|$ denotes the Euclidean norm of x irrespective of the dimension. Since the sequence $\widetilde{b}_n$ is bounded (see Lemma 3.6), it follows that the bounds on the right hand sides of (3.7) and (3.8) do not depend on n.

Therefore, it follows from the relative compactness criteria from Malliavin calculus of [13] that the sequence (X_r^n) admits a subsequence $(X_r^{n_k})_k$ converging to some X_r in L^2.

It remains to show that the choice of the subsequence $(X_r^{n_k})_k$ does not depend on r. That is, for every $t \in [s,T]$, $(X_t^{n_k})_k$ converges to X_t in L^2. In fact, we will show that the whole sequence converges. This is done as in the proof of [46, Proposition 2.6]. Assume by contradiction that for some $t \in [s,T]$, there is a subsequence $(n_k)_{k\geq 0}$ such that

$$\|X_t^{n_k} - X_t\|_{L^2} \geq \varepsilon. \tag{3.9}$$

Since (3.7) is proved for arbitrary n, it follows again by the compactness criteria of [13] that $(X^{n_k})_k$ admits a further subsequence $(X_t^{n_{k_1}})_{k_1}$ which converges in L^2 to X_t. But since we showed in Step 2a that the whole sequence of processes (X^n) converges weakly to the process $\widetilde{X}$, it follows that $(X_t^{n_{k_1}})_{k_1}$ converges weakly to $\widetilde{X}_t$ and therefore, by uniqueness of the limit, $\widetilde{X}_t = X_t$. Since by (3.9) it holds

$$\|X_t^{n_{k_1}} - X_t\|_{L^2} \geq \varepsilon,$$

we have a contradiction. Thus,

$$X_t^n \to X_t \quad \text{in} \quad L^2 \quad \text{for every } t \in [s,T].$$

Step 3: Candidate solution for the value process Y and the control process Z. In this part we show that the sequence (Y^n, Z^n) converges strongly in $\mathcal{H}^2(\mathbb{R}^l) \times \mathcal{H}^2(\mathbb{R}^{l\times d})$ to some (Y, Z).

First recall that (Y^n) is a bounded sequence in the Hilbert space $\mathcal{H}^2(\mathbb{R}^l)$. Thus, it admits a subsequence again denoted (Y^n) which converges weakly in $\mathcal{H}^2(\mathbb{R}^l)$ to some Y. We will show that the convergence is actually strong, provided that we restrict ourselves to a small enough time interval. Let $\delta \in (0,T)$ be fixed. By Lemma 3.2, the sequence of functions (v_n) is bounded and equicontinuous on $[0, T-\delta] \times \mathbb{R}^d$. Thus, by the Arzela-Ascoli theorem, there is a subsequence again denoted (v_n) which converges locally uniformly to a continuous function v^δ. Since by Lemma 3.2 the functions v_n are Hölder continuous with a coefficient that does not depend on n and with common Hölder exponents α' (in x) and $\alpha'/2$ (in t), we have

$$\begin{aligned} E\big[|v_n(t, X_t^n) - v^\delta(t, X_t)|^2\big] &\leq E\big[|v_n(t, X_t^n) - v_n(t, X_t)|^2\big] + E\big[|v_n(t, X_t) - v^\delta(t, X_t)|^2\big] \\ &\leq CE\big[|X_t - X_t^n|^{2\alpha'}\big] + E\big[|v_n(t, X_t) - v^\delta(t, X_t)|^2\big] \to 0. \end{aligned} \tag{3.10}$$

Therefore, $Y_t^n = v_n(t, X_t^n)$ converges to $v^\delta(t, X_t)$ in L^2 for each $t \in [0, T-\delta]$. It then follows by uniqueness of the limit that

$$Y_t = v^\delta(t, X_t) \quad \text{for all} \quad t \in [0, T-\delta]. \tag{3.11}$$

It then follows by Lebesgue dominated convergence (in view of Lemma 3.2) that (Y^n) converges to Y in $\mathcal{H}^2(\mathbb{R}^l)$ restricted to $[0, T-\delta]$, i.e.

$$\lim_{n\to\infty} E\Big[\int_0^{T-\delta} |Y_t^n - Y_t|^2 \, dt\Big] = 0. \tag{3.12}$$

The equation (3.11) further shows that v^δ does not depend on δ. Thus, we will henceforth write

$$Y_t = v(t, X_t) \quad \text{for all} \quad t \in [0, T-\delta] \quad \text{and} \quad \text{for all} \quad \delta > 0.$$

We now turn to the construction of the candidate control process Z. We want to justify that under both conditions (B1) and (B2) the sequence b_n can be taken uniformly bounded. In fact, if the function b satisfies (B1), and since (Y^n) is uniformly bounded (this fcomes from the representation $v_n(t, X_t^n) = Y_t^n$ and Lemma 3.2) it follows by uniqueness of solution that (X^n, Y^n, Z^n) also solves the FBSDE (3.1) with b_n replaced by its restriction on $[0,T]\times\mathbb{R}^d\times B_R(0)\times\mathbb{R}^{l\times d}$. Similarly, if condition (B2) holds, then (Y^n) and (Z^n) are bounded, and by uniqueness, (X^n, Y^n, Z^n) also solves the FBSDE (3.1) with b_n replaced by its restriction on $[0,T] \times \mathbb{R}^d \times B_R(0) \times B_R(0)$. In particular, we can assume without loss of generality that b_n is uniformly bounded, i.e. $|b_n(t,x,y,z)| \leq C$ for all n, t, x, y, z and for some constant $C > 0$. Therefore, it follows by Theorem A.1 that for every $\delta > 0$ and $\kappa \in (0,1)$ there is a constant $C_{\delta,\kappa}$ independent on the derivatives of the coefficient (which in particular does not depend on n) such that for every $t, t' \in [0, T-\delta]$ and $x, x' \in \mathbb{R}^d$ it holds that

$$|D_x v_n(t,x) - D_x v_n(t', x')| \leq C_{\delta,\kappa}(|x - x'|^\kappa + |t - t'|^{\kappa/2}).$$

Now, let (δ^k) be a strictly decreasing sequence converging to 0. By Arzela-Ascoli theorem, there is a subsequence $w_{n,k} := D_x v_n|_{[0,T-\delta^k]\times\mathbb{R}^d}$ which converges locally uniformly to some function w_k on $[0,T-\delta^k]\times\mathbb{R}^d$. Since $Z^n_t = D_x v_n(t,X^n_t)\sigma$ for all $t\in[0,T]$ (recall (3.3)) we then have $Z^{n,k}_t = w_{n,k}(t,X^n_t)\sigma$ for every $t\in[0,T-\delta^k]$ and every $k\in\mathbb{N}$, for some subsequence of Z^n. And arguing as in Equation (3.10), we have

$$Z^{n_k}_t = D_x w_{n,k}(t,X^{n_k}_t)\sigma \to w_k(t,X_t)\sigma =: Z^k \quad \text{in} \quad L^2 \quad \text{for every} \quad t\in[0,T-\delta^k].$$

Assumption (A2) and uniqueness of the limit (w_k of the sequence $(w_{n,k})_n$) show that $w_k = w_{k+1}$ on $[0,T-\delta^k]$ for every k. Thus, the function

$$w(t,x) := w_1(t,x)1_{[0,T-\delta^1]}(t) + \sum_{k=1}^{\infty} w_k(t,x)1_{[T-\delta^k,T-\delta^{k+1}]}(t)$$

is a well-defined Borel measurable function. In particular, the limit Z^k does not depend on k. In fact, putting

$$Z_t := w(t,X_t)\sigma, \tag{3.13}$$

we have by Lebesgue dominated convergence that $Z^{n,k}\to Z$ in $\mathcal{H}^2(\mathbb{R}^{l\times d})$ restricted to the interval $[0,T-\delta^k]$. In particular, it follows by Itô isometry that

$$\int_0^{T-\delta^k} Z^{n,k}_t\,dW_t \to \int_0^{T-\delta^k} Z_t\,dW_t \quad \text{in} \quad L^2 \quad \text{for every } k. \tag{3.14}$$

Step 4: Verification. The goal of this step is to show that the triple of processes (X,Y,Z) constructed above satisfies the coupled system (1.1). This part of the proof will be further split into 2 steps: We first show that (X,Y,Z) satisfies the forward equation. This step uses the representations $Y_t = v(t,X_t)$ and $Z_t = w(t,X_t)\sigma$ in a crucial way. In fact, these representation allow to obtain a solution $\bar{X}$ of a decoupled SDE with measurable drift that we can then show to coincide with the candidate solution X constructed above. In the last part we show that (X,Y,Z) satisfies the backward equation.

Step 4a: The forward equation. Using either of the conditions (B1) or (B2), we can show as above that the function $x\mapsto b(t,x,v(t,x),w(t,x)\sigma)$ is bounded. Therefore, [47] gives existence of a unique solution $\bar{X}$ to the SDE

$$\bar{X}_t = x + \int_s^t b(u,\bar{X}_u, v(u,\bar{X}_u), w(u,\bar{X}_u)\sigma)\,du + \int_s^t \sigma\,dW_u.$$

Hence, in view of (3.11) and (3.13), it remains to show that $\bar{X}_t = X_t$ P-a.s. for every $t\in[s,T]$ to conclude that the forward SDE is satisfied, that is, that

$$X_t = x + \int_s^t b(u,X_u,Y_u,Z_u)\,du + \int_s^t \sigma\,dW_u. \tag{3.15}$$

To that end, continuity of the paths of X and $\bar{X}$ and uniqueness of the limit, it suffices to show that for each fixed $t\in[s,T]$ the sequence (X^n_t) converges to $\bar{X}_t$ in the weak topology of $L^2(P)$. For any progressive and square integrable process q, we will use the notation

$$\mathcal{E}(q\cdot W)_{s,t} := \exp\Big(\int_s^t q_u\,dW_u - \frac{1}{2}\int_s^t |q_u|^2\,du\Big),$$

for the stochastic exponential of the martingale $\int q\,dW$. Since the set

$$\left\{\mathcal{E}(\dot{\varphi}\cdot W)_{0,T}:\ \varphi\in C_b^1([0,T],\mathbb{R}^d)\right\}$$

is dense in $L^2(P)$, in order to get weak convergence it is enough to show that $(X_t^n\mathcal{E}(\dot{\varphi}_u\cdot W)_{0,T})$ converges to $\bar{X}_t\mathcal{E}(\dot{\varphi}_u\cdot W)_{0,T}$ in expectation, for every $\varphi\in C_b^1([0,T],\mathbb{R}^d)$. Hereby $C_b^1([0,T],\mathbb{R}^d)$ denotes the space of bounded continuously differentiable functions on $[0,T]$ with values in $\mathbb{R}^d$, and $\dot{\varphi}$ is the derivative of φ. Put $\widetilde{X}_t^n(\omega):=X_t^n(\omega+\varphi)$ and $\widetilde{X}_t(\omega):=\bar{X}_t(\omega+\varphi)$. It follows by the Cameron-Martin theorem, see e.g. [60] that $\widetilde{X}^n$ satisfies the SDE

$$d\widetilde{X}_t^n=\Big(b_n(t,\widetilde{X}_t^n,v_n(t,\widetilde{X}_t^n),D_xv_n(t,\widetilde{X}_t^n)\sigma)+\sigma\dot{\varphi}_t\Big)\,dt+\sigma dW_t.$$

In fact, for every $H\in L^2(P;\mathcal{F}_t)$, using Cameron-Martin-Girsanov (see for example [60]) it holds

$$\begin{aligned}
E[\widetilde{X}_t^nH]&=E\Big[X_t^nH(\omega-\varphi)\mathcal{E}(\dot{\varphi}_u\cdot W)_{s,T}\Big]\\
&=E\Big[\Big(x+\int_s^t b_n(u,X_u^n,v_n(u,X_u^n),D_xv_n(u,X_u^n)\sigma)\,du+\sigma(W_t-W_s)\Big)H(\omega-\varphi)\mathcal{E}(\dot{\varphi}_u\cdot W)_{s,T}\Big]\\
&=E\Big[\Big(x+\int_s^t b_n(u,X_u^n,v_n(u,X_u^n),D_xv_n(u,X_u^n)\sigma)(\omega+\varphi)\,du+\sigma(W_t-W_s)(\omega+\varphi)\Big)H\Big]\\
&=E\Big[\Big(x+\int_s^t b_n(u,\widetilde{X}_u^n,v_n(u,\widetilde{X}_u^n),D_xv_n(u,\widetilde{X}_u^n)\sigma)+\sigma\dot{\varphi}_u\,du+\sigma(W_t-W_s)(\omega)\Big)H\Big],
\end{aligned}$$

where the latter equality follows by applying once more the Cameron-Martin-Girsanov theorem and the fact that $W_t(\omega+\varphi)=W_t(\omega)+\varphi_t=W_t(\omega)+\int_0^t\dot{\varphi}_u\,du$ since W is the canonical process. This proves the claim. That $\widetilde{X}$ satisfies

$$d\widetilde{X}_t=\Big(b(t,\widetilde{X}_t,v(t,\widetilde{X}_t),w(t,\widetilde{X}_t)\sigma)+\sigma\dot{\varphi}_t\Big)\,dt+\sigma dW_t$$

is proved similarly. Now put

$$u_n(t,x):=\sigma^*(\sigma\sigma^*)^{-1}b_n(t,x,v_n(t,x),D_xv_n(t,x)\sigma)\quad\text{and}\quad u:=\sigma^*(\sigma\sigma^*)^{-1}b(t,x,v(t,x),w(t,x)\sigma).$$

Recall that the law of $\widetilde{X}_t^n$ under the probability measure Q^n with density $\mathcal{E}(u_n(r,\widetilde{X}_r^n)+\dot{\varphi}_r\cdot W)_{0,T}$ coincides with the law of $x+\sigma W_t$ under P. Similarly, the law of $\widetilde{X}_t$ under the probability measure Q with density $\mathcal{E}(u(r,\widetilde{X}_r)+\dot{\varphi}_r\cdot W)_{0,T}$ coincides with the law of $x+\sigma W_t$ under P. To show $\bar{X}_t=X_t$ P-a.s. for every $t\in[s,T]$ as needed, we will show that (X_t^n) converges weakly to $\bar{X}_t$ (in the weak topology of L^2) and conclude by uniqueness of the limit. Thus, it follows by

Girsanov's theorem and the inequality $|e^a - e^b| \leq |e^a + e^b||a-b|$

$$
\begin{aligned}
&E\left[X_t^n \mathcal{E}(\dot{\varphi}_u \cdot W)_{0,T}\right] - E\left[\bar{X}_t \mathcal{E}(\dot{\varphi}_u \cdot W)_{0,T}\right]\\
&= E\Big[(x+\sigma W_t)\Big(\mathcal{E}\left(\{u_n(r,x+\sigma W_r)+\dot{\varphi}_r\}\cdot W\right)_{0,T} - \mathcal{E}(\{u(r,x+\sigma W_r)+\varphi_r\}\cdot W)_{0,T}\Big)\Big]\\
&\leq CE\left[|x+\sigma\cdot W_t|^2\right]^{\frac{1}{2}}\\
&\quad\times E\Big[\Big(\mathcal{E}\Big(\Big\{u_n\Big(r,x+\sigma W_r\Big)+\dot{\varphi}_r\Big\}\cdot W\Big)_{0,T} + \mathcal{E}\Big(\Big\{u\Big(r,x+\sigma W_r\Big)_{0,T}+\dot{\varphi}_r\Big\}\cdot W\Big)_{0,T}\Big)^4\Big]^{\frac{1}{4}}\\
&\quad\times\Big\{E\Big[\Big(\int_0^T\Big(u_n\Big(r,x+\sigma W_r\Big)-u\Big(r,x+\sigma W_r\Big)\Big)dW_r\Big)^4\Big]\\
&\quad+E\Big[\Big(\int_0^T\Big\{\|u_n(r,x+\sigma W_r)+\dot{\varphi}_r\|^2-\|u(r,x+\sigma W_r)+\dot{\varphi}_r\|^2\Big\}\mathrm{d}r\Big)^4\Big]\Big\}^{\frac{1}{4}}\\
&= I_1\times I_{2,n}\times(I_{3,n}+I_{4,n})^{1/4}.
\end{aligned} \tag{3.16}
$$

That I_1 is finite is clear, by properties of Brownian motion. Since b_n is bounded, so is u_n. Thus, by boundedness of $\dot{\varphi}$, it holds that $\sup_n I_{2,n}$ is finite.

Now if we show that the sequence (u_n) converges to u pointwise, it would follow by Lebesgue's dominated convergence theorem, to get that $I_{3,n}$ and $I_{4,n}$ converge to 0 as n goes to infinity, hence concluding the proof. In fact, there is $R>0$ such that $|v_n|\leq R$ and there is R' such that[1] for every $t\in[0,T)$, it holds that $|D_x v_n(t,x)|\leq R'$ for all n. Thus, by definition of u_n and u, for almost every $(t,x)\in[0,T)\times\mathbb{R}^d$ we have

$$
\begin{aligned}
|u_n(t,x)-u(t,x)| &\leq C|b_n(t,x,v_n(t,x),D_xv_n(t,x)\sigma)-b(t,x,v(t,x),w(t,x)\sigma)|\\
&\leq C|b_n(\cdot,v_n,D_xv_n\sigma)-b(\cdot,v_n,D_xv_n\sigma)|(t,x)+C|b(\cdot,v_n,D_xv_n\sigma)-b(\cdot,v,w\sigma)|(t,x)\\
&\leq C\sup_{y\in B_R(0),z\in B_{R'}(0)}|b_n(t,x,y,z)-b(t,x,y,z)|\\
&\quad+C|b(t,x,v_n(t,x),v_n(t,x)\sigma)-b(t,x,v(t,x),w(t,x)\sigma)|.
\end{aligned}
$$

The first term converges to zero since b_n converges to b locally uniformly in (y,z) (Lemma 3.1); and the second term converges to zero because v_n and $D_xv_n\sigma$ converge to v and $w\sigma$ respectively, and the function $b(t,x,\cdot,\cdot)$ is continuous on the ball $B_R(0)\times B_{R'}(0)$. Therefore, (X_t^n) converges to $\bar{X}$ in the weak topology of L^2, therefore $\bar{X}=X$ satisfies the forward equation (3.15).

Step 4b: The backward equation. In this final step of the proof we show that the process (X,Y,Z) satisfies the backward equation. The arguments is very similar to those of the Step 4a and also rely on the existence of the decoupling fields v and w and Girsanov's transform.

By Steps 2 and 3 we know that (X_t^n) converges to X_t in L^2 and $(Y^{n,k},Z^{n,k})$ converges to (Y,Z) in $\mathcal{H}^2(\mathbb{R}^l)\times\mathcal{H}^2(\mathbb{R}^{l\times d})$ (restricted to the interval $[0,T-\delta^k]$), where $(Y^{n,k},Z^{n,k})$ is the sequence corresponding to (Y^n,Z^n) restricted to $[0,T-\delta^k]$. Let k be fixed and let $X^{n,k}$ be a subsequence corresponding to $(Y^{n,k},Z^{n,k})$. For every n,k we have

$$
Y_t^{n,k}=Y_{T-\delta^k}^{n,k}+\int_t^{T-\delta^k}g_n(u,X_u^{n,k},Y_u^{n,k},Z_u^{n,k})\,du-\int_t^{T-\delta^k}Z_u^{n,k}\,dW_u. \tag{3.17}
$$

[1] Under the condition (B1) and when $t=T$, the sequence (D_xv_n) might not be bounded and (u_n) does not necessarily converge to u but convergence for almost every t is enough.

Now, we would like to take first the limit in n and then limit in k on both sides. By Step 3, the sequences of random variables $Y_t^{n,k}$, $Y_{T-\delta^k}^{n,k}$ and $\int_t^{T-\delta^k} Z_u^{n,k}\,dW_u$ respectively converge to Y_t, $Y_{T-\delta^k}$ and $\int_t^{T-\delta^k} Z_u\,dW_u$ in L^2. Thus, it suffices to show that $\int_t^{T-\delta^k} g_n(X_u^{n,k}, Y_u^{n,k}, Z_u^{n,k})\,du$ converges to $\int_t^{T-\delta^k} g(u, X_u, Y_u, Z_u)\,du$ in L^2. To this end, define

$$\widetilde{g}_{n,k}(t,x) := g_n(t,x,v_n(t,x), D_x v_n(t,x)\sigma)_{|_{[0,T-\delta^k]}} \quad \text{and} \quad \widetilde{g}(t,x) := g(t,x,v(t,x),w(t,x)\sigma).$$

Observe that $\widetilde{g}_{n,k}$ is uniformly bounded (this can be shown using similar arguments as in Lemma 3.3) and converges to g almost surely on $[0, T-\delta^k] \times \mathbb{R}^d$. In fact,

$$\begin{aligned}
|\widetilde{g}_{n,k}(t,x) - \widetilde{g}(t,x)| &= |g_n(t,x,v_n(t,x), D_x v_n(t,x)\sigma) - g(t,x,v_n(t,x), D_x v_n(t,x)\sigma)| \\
&\quad + |g(t,x,v_n(t,x), D_x v_n(t,x)\sigma) - g(t,x,v(t,x),w(t,x)\sigma)| \\
&\le \sup_{y,z} |g_n(t,x,y,z) - g(t,x,y,z)| \\
&\quad + |g(t,x,v_n(t,x), D_x v_n(t,x)\sigma) - g(t,x,v(t,x),w(t,x)\sigma)| \to 0,
\end{aligned}$$

where we used Lemma 3.1 and continuity of g in (y,z) . Recall the representations $Y_u^{n,k} = v_n(u, X_u^{n,k})$, $Z_u^{n,k} = D_x v^n(u, X_u^{n,k})\sigma$ and $Y_u = v(u, X_u)$, $Z_u = w(u, X_u)\sigma$. For any $m \in \mathbb{N}$, we have

$$\begin{aligned}
&E\Big[\int_t^{T-\delta^k} |g_{n,k}(u, X_u^{n,k}, Y_u^{n,k}, Z_u^{n,k}) - g(u, X_u, Y_u, Z_u)|^2\,du\Big] = E\Big[\int_t^{T-\delta^k} |\widetilde{g}_n(u, X_u^{n,k}) - \widetilde{g}(u, X_u)|^2\,du\Big] \\
&\le E\Big[\int_t^{T-\delta^k} |\widetilde{g}_n(u, X_u^{n,k}) - \widetilde{g}(u, X_u^{n,k})|^2 + |\widetilde{g}(u, X_u^{n,k}) - \widetilde{g}_m(u, X_u^{n,k})|^2 + |\widetilde{g}_m(u, X_u^{n,k}) - \widetilde{g}(u, X_u)|^2\,du\Big] \\
&\le E\Big[\mathcal{E}\big(\widetilde{b}_n(u, x+\sigma W_u)\cdot W\big)_{0,T}\Big\{\int_t^{T-\delta^k} |\widetilde{g}_{n,k}(u, x+\sigma W_u) - \widetilde{g}(u, x+\sigma W_u)|^2 \\
&\quad + |\widetilde{g}(u, x+\sigma W_u) - \widetilde{g}_m(u, x+\sigma W_u)|^2\,du\Big\}\Big] + E\Big[\int_t^{T-\delta^k} |\widetilde{g}_m(u, X_u^{n,k}) - \widetilde{g}(u, X_u)|^2\,du\Big],
\end{aligned}$$

where the last inequality follows by Girsanov's theorem and where we used the notation

$$\widetilde{b}_n(t,x) := \sigma^*(\sigma\sigma^*)^{-1} b_n(t,x,v_n(t,x), D_x v_n(t,x)\sigma). \tag{3.18}$$

Therefore, using Hölder's inequality the above estimation continues as

$$\begin{aligned}
&E\Big[\int_t^{T-\delta^k} |g_{n,k}(u, X_u^{n,k}, Y_u^{n,k}, Z_u^{n,k}) - g(u, X_u, Y_u, Z_u)|^2\,du\Big] \\
&\le C E\Big[\mathcal{E}\big(\widetilde{b}_n(u, x+\sigma W_u)\cdot W\big)_{0,T}^2\Big]^{1/2} E\Big[\int_t^{T-\delta^k} |\widetilde{g}_n(u, x+\sigma W_u) - \widetilde{g}(u, x+\sigma W_u)|^4 \\
&\quad + |\widetilde{g}(u, x+\sigma W_u) - \widetilde{g}_m(u, x+\sigma W_u)|^4\,du\Big]^{1/2} + E\Big[\int_t^{T-\delta^k} |\widetilde{g}_m(u, X_u^{n,k}) - \widetilde{g}(u, X_u)|^2\,du\Big].
\end{aligned}$$

Since $\widetilde{b}_n$ is bounded, the quantity $E\Big[\mathcal{E}\big(\widetilde{b}_{n,k}(u, x+\sigma W_u)\cdot W\big)_{0,T}^2\Big]$ is bounded. Thus, letting m fixed and taking the limit as n goes to infinity we obtain by Lebesgue dominated convergence that

$$\begin{aligned}
&\lim_{n\to\infty} E\Big[\int_t^{T-\delta^k} |g_{n,k}(u, X_u^{n,k}, Y_u^{n,k}, Z_u^{n,k}) - g(u, X_u, Y_u, Z_u)|^2\, du\Big] \\
&\le CE\Big[\int_t^{T-\delta^k} |\widetilde{g}(u, x+\sigma W_u) - \widetilde{g}_m(u, x+\sigma W_u)|^4\, du\Big] + E\Big[\int_t^{T-\delta^k} |\widetilde{g}_m(u, X_u) - \widetilde{g}(u, X_u)|^2\, du\Big].
\end{aligned}$$

Letting m go to infinity it follows again by dominated convergence that the right hand side above goes to zero. (We also used that the normal distribution is absolutely continuous.) Thus

$$\int_t^{T-\delta^k} g_n(u, X_u^{n,k}, Y_u^{n,k}, Z_u^{n,k}) \to \int_t^{T-\delta^k} g(u, X_u, Y_u, Z_u)\, du \quad \text{in} \quad L^2.$$

Hence, (X, Y, Z) satisfies

$$Y_t = Y_{T-\delta^k} + \int_t^{T-\delta^k} g(u, X_u, Y_u, Z_u)\, du - \int_t^{T-\delta^k} Z_u\, dW_u \quad P\text{-a.s. for every } k.$$

Next, we take the limit as k goes to infinity. Since $\delta^k \downarrow 0$, we only need to justify that $(Y_{T-\delta^k})$ converges to Y_T P-a.s. Indeed, since (Y_T^n) converges to Y_T in the weak topology of L^2, there exists a subsequence $(\widetilde{Y}_T^n)$ in the asymptotic convex hull of (Y_T^n) such that $(\widetilde{Y}_T^n)$ converges to Y_T in L^2. Moreover, $\widetilde{Y}_T^n$ satisfies

$$\widetilde{Y}_T^n = \widetilde{Y}_t^n - \int_t^T G_u^n du + \int_t^T \widetilde{Z}_u^n dW_u$$

where $(\widetilde{Y}_t^n, G_u^n, \widetilde{Z}_u^n)$ is the convex combination of $(Y_t^n, g_n(u, X_u^n, Y_u^n, Z_u^n), Z_u^n)$ corresponding to $\widetilde{Y}_T^n$. If the condition (B1) is satisfied, then $|g_n(u, X_u^n, Y_u^n, Z_u^n)|$ is dominated by $|Y_u^n|$ which is bounded, and if the condition (B2) is satisfied, then $Z_u^n = D_x v^n(u, X_u^n)\sigma$ is bounded (by Lemma 3.2), thus it follows by (A3) that $|g_n(u, X_u^n, Y_u^n, Z_u^n)|$ is bounded. Hence, G_u^n is bounded under both conditions. Therefore it follows by triangular inequality that for every $k, n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\big|Y_{T-\delta^k} - E[Y_T \,|\, \mathcal{F}_{T-\delta^k}]\big| \\
&\quad \le C\big(|Y_{T-\delta^k} - \widetilde{Y}_{T-\delta^k}^n| + |E[\widetilde{Y}_{T-\delta^k}^n - \widetilde{Y}_T^n \,|\, \mathcal{F}_{T-\delta^k}]| + E[|\widetilde{Y}_T^n - Y_T| \,|\, \mathcal{F}_{T-\delta^k}]\big) \\
&\quad \le C\Big(|Y_{T-\delta^k} - \widetilde{Y}_{T-\delta^k}^n| + E\Big[\int_{T-\delta^k}^T |G_u^n|\, du \,|\, \mathcal{F}_{T-\delta^k}\Big] + E[|\widetilde{Y}_T^n - Y_T| \,|\, \mathcal{F}_{T-\delta^k}]\Big) \\
&\quad \le C(|Y_{T-\delta^k} - \widetilde{Y}_{T-\delta^k}^n| + \delta^k + E[|\widetilde{Y}_T^n - Y_T| \,|\, \mathcal{F}_{T-\delta^k}])
\end{aligned}$$

for some constant $C > 0$. Since $(\widetilde{Y}_T^n)$ converges to Y_T in L^2, $(Y_{T-\delta^k}^n)$ converges to $Y_{T-\delta^k}$ in L^2 and $\widetilde{Y}_{T-\delta^k}^n$ is the convex combination of $Y_{T-\delta^k}^n$, taking the limit first in n and then in k as they go to infinity shows that $|Y_{T-\delta^k} - E[Y_T \,|\, \mathcal{F}_{T-\delta^k}]| \to 0$ P-a.s. On the other hand, in our filtration every martingale has a continuous version. Thus,

$E[Y_T \,|\, \mathcal{F}_{T-\delta^k}] \to Y_T$ P-a.s. as k goes to infinity. We can therefore conclude that $Y_{T-\delta^k} \to Y_T$ P-a.s. when k goes to infinity, which yields

$$Y_t = Y_T + \int_t^T g(u, X_u, Y_u, Z_u)\, du - \int_t^T Z_u \, dW_u.$$

It finally remains to show that $Y_T = h(X_T)$. Since (Y_T^n) converges to Y_T in the weak topology of L^2 (see the beginning of Step 3) if we show that (Y_T^n) converges to $h(X_T)$ in L^2 then we can conclude that $Y_T = h(X_T)$. If (B2) holds, this is clear. In case (B1) holds, this is done using again a Girsanov change of measure and boundedness of $\widetilde{b}_n$ (recall definition given in (3.18)). In fact, for every $m \in \mathbb{N}$ it holds that

$$\begin{aligned}
&E[|h_n(X_T^n) - h(X_T)|^2] \\
&\leq C\Big(E[|h_n(X_T^n) - h(X_T^n)|^2] + E[|h(X_T^n) - h_m(X_T^n)|^2] + E[|h_m(X_T^n) - h(X_T)|^2]\Big) \\
&\leq C\Big(E\Big[\mathcal{E}\big(\widetilde{b}_n(t, x + \sigma W_t) \cdot W\big)_{0,T}\Big\{|h(x + \sigma W_T) - h_m(x + \sigma W_T)|^2 + |h_m(x + \sigma W_T) - h(x + \sigma W_T)|^2\Big\}\Big] \\
&\quad + E[|h_m(X_T^n) - h(X_T)|^2]\Big) \\
&\leq C\Big(E\Big[\mathcal{E}\big(\widetilde{b}_n(t, x + \sigma W_t) \cdot W\big)_{0,T}^2\Big]^{1/2} \\
&\quad \times E\Big[|h(x + \sigma W_T) - h_m(x + \sigma W_T)|^4 + |h_m(x + \sigma W_T) - h(x + \sigma W_T)|^4\Big]^{1/2} \\
&\quad + E[|h_m(X_T^n) - h(X_T)|^2]\Big).
\end{aligned}$$

Since $\widetilde{b}_n$ is bounded, the first term on the right hand side above is bounded. Thus, fix m then take the limit $n \to \infty$ and then the limit $m \to \infty$ to get by dominated convergence

$$E[|h_n(X_T^n) - h(X_T)|^2] \to 0.$$

This concludes the proof. $\square$

3.2. Regularity of solutions

In this section we investigate regularity properties of the solution (X, Y, Z) of the FBSDE (1.1). We will consider two types of regularity properties. We start by proving Malliavin differentiability of the solution. This follows as a direct consequence of the method of proof of the existence result. Then, we continue to consider smoothness of the solution as function of the initial position of the forward process. We will show that for each $s \in [0, T]$ and $t \geq s$, the mapping $x \mapsto (X_t^{s,x}, Y_t^{s,x})$ belongs to a weighted Sobolev space for almost every path. The last result will be central for applications to PDEs.

3.2.1. Malliavin differentiability

Let $0 \leq s \leq t \leq T$ and $x \in \mathbb{R}^d$. Let (X, Y, Z) be the solution of FBSDE (1.1) given by Theorem 2.1. The next result gives the Malliavin differentiability of (X, Y, Z). We additionally consider the following conditions:

(A5) The function $g(t, x, y, z) = g(t, x, y)$ does not depend on z and is Lipschitz continuous in (x, y).

(A6) The function $g(t, x, y, z) = g(t, y, z)$ does not depend on x and is continuously differentiable and Lipschitz continuous in (y, z).

Proposition 3.4. *Assume that the conditions (A1)-(A4) are satisfied.*

(i) If (B1) is satisfied, then X_t is Malliavin differentiable for all $t \in [0,T]$ and for every $\delta > 0$, Y_t is Malliavin differentiable for all $t \in [0, T-\delta]$.

(ii) If (B2) is satisfied, then (X_t, Y_t) is Malliavin differentiable for all $t \in [0,T]$.

(iii) If (B2) and either of the conditions (A5) or (A6) hold, then (X_t, Y_t, Z_t) is Malliavin differentiable for all $t \in [0,T]$.

Proof. Consider the sequence (X^n) constructed in the proof of Theorem 2.1. Recall that under both (B1) and (B2) we have

$$X_t^n \to X_t \quad \text{in} \quad L^2 \quad \text{for every } t \in [0,T]$$

and (see Equation (3.7) with $t' = 0$ therein) we have

$$E\left[|D_t X_s^n|^2\right] \leq \sum_{i=1}^{d} E\left[\left\|D_t^i X_s^n\right\|^2\right] \leq d\mathcal{C}_{d,T}(\|\widetilde{b}_n\|_\infty)t$$

where $\widetilde{b}_n$ is a uniformly bounded sequence. Therefore, by [51, Lemma 1.2.3] we conclude that X_t is Malliavin differentiable for all $t \in [0,T]$. In particular, $\sup_t E[|D_t X_s|^2] < \infty$. To deduce the differentiability of Y, recall that for every $\delta > 0$ and every $t \in [0, T-\delta]$ the function $x \mapsto v(t,x)$ is Lipschitz continuous. Thus, it follows by chain rule (see [51, Proposition 1.2.4]) that Y_t is Malliavin differentiable for all $t \in [0, T-\delta]$.

When condition (B2) is satisfied, the function $x \mapsto v(t,x)$ is Lipschitz continuous for every $t \in [0,T]$. In fact, it follows from Lemma 3.2 that v_n is Lipschitz continuous with Lipschitz constant C_δ (see Equation (3.4)) which does not depend on n. Thus, since v_n converges pointwise to v the claim follows. Again by chain rule, Y_t is Malliavin differentiable for all $t \in [0,T]$. Thus, (X_t, Y_t) is Malliavin differentiable.

If furthermore condition (A5) holds, then in view of the identity

$$\int_t^T Z_s dW_s = h(X_T) - Y_t + \int_t^T g(s, X_s, Y_s) ds,$$

it follows from the chain rule and [53, Lemma 2.3] that Z_t is Malliavin differentiable for all $t \in [0,T]$. If we rather assume (A6), then since X_t is Malliavin differentiable, the Malliavin differentiability of (Y_t, Z_t) follows from the chain rule and [30, Proposition 5.3] since $\int_0^T E[|D_s h(X_T)|^2]\, ds < \infty$. $\square$

3.2.2. Weighted Sobolev differentiable flow

We now investigate differentiability properties of the solution with respect to the initial variable of the forward process. Let $0 \leq s \leq t \leq T$ and $x \in \mathbb{R}^d$. We denote by $(X^{s,x}, Y^{s,x}, Z^{s,x})$ the solution of the FBSDE

$$\begin{cases} X_t = x + \int_s^t b(u, X_u, Y_u, Z_u)\, \mathrm{d}u + \int_0^t \sigma dW_u \\ Y_t = h(X_T) + \int_t^T g(u, X_u, Y_u, Z_u) \mathrm{d}u - \int_t^T Z_u\, dW_u \quad t \in [s,T] \end{cases} \tag{3.19}$$

given by Theorem 2.1. The next result gives regularity of the function $x \mapsto (X^{s,x}, Y^{s,x})$. We now introduce the Sobolev space where the derivatives will be defined. Let ρ be a weight function, that is, a measurable function $\rho : \mathbb{R}^d \to [0,\infty)$ satisfying

$$\int_{\mathbb{R}^d} (1 + |x|^p) \rho(x)\, dx < \infty$$

for some $p > 1$. Let $L^p(\mathbb{R}^d, \rho)$ be the weighted Lebesgue space of (classes) of measurable functions $f : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\|f\|^p_{L^p(\mathbb{R}^d,\rho)} := \int_{\mathbb{R}^d} |f(x)|^p \rho(x)\, dx < \infty.$$

For functions $f : \mathbb{R}^d \to \mathbb{R}^l$ satisfying this integrability property we analogously define the space $L^p(\mathbb{R}^l, \rho)$. Further denote by $\mathcal{W}^1_p(\mathbb{R}^d, \rho)$ the weighted Sobolev space of functions $f \in L^p(\mathbb{R}^d, \rho)$ admitting weak derivatives of first order $\partial_{x_i} f$ and such that

$$\|f\|_{\mathcal{W}^1_p(\mathbb{R}^d,\rho)} := \|f\|_{L^p(\mathbb{R}^d,\rho)} + \sum_{i=1}^d \|\partial_{x_i} f\|_{L^p(\mathbb{R}^d,\rho)} < \infty.$$

Proposition 3.5. *Assume that the conditions (A1)-(A4) are satisfied.*

(i) If condition (B1) holds, then we have

$$X^{s,x}_t \in L^2\big(\Omega; \mathcal{W}^1_p(\mathbb{R}^d, \rho)\big) \quad \text{for every } t \in [0, T] \tag{3.20}$$

and if $l = 1$, then for every bounded open set $U \subseteq \mathbb{R}^d$ we have

$$Y^{s,x}_t \in L^2\big(\Omega; \mathcal{W}^1_1(U)\big) \quad \text{for every } t \in [0, T - \delta] \quad \text{and every } \delta > 0. \tag{3.21}$$

(ii) If condition (B2) holds and $l = 1$, then (3.20) and (3.21) hold with $\delta = 0$.

Proof. Recall from Theorem 2.1 that the solution (X, Y, Z) of the FBSDE (1.1) satisfies $Y^{s,x}_t = v(t, X^{s,x}_s)$ and $Z^{s,x}_t = w(t, X^{s,x}_t)\sigma$ for some bounded measurable function $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^l$ and a measurable function $w : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{l\times d}$. Thus, $X^{s,x}$ satisfies

$$X^{s,x}_t = x + \int_s^t b(u, X^{s,x}_u, v(u, X^{s,x}_u), w(u, X^{s,x}_u)\sigma)\, du + \sigma(W_t - W_s).$$

Under both conditions (B1) and (B2) the function $x \mapsto b(t, x, v(t, x), w(t, x)\sigma)$ is bounded and measurable. Thus, it follows from [50, Theorem 3] that $X^{s,x}_t \in L^2\big(\Omega; \mathcal{W}^1_p(\mathbb{R}^d, \rho)\big)$.

To deduce differentiability of Y, recall that for every $\delta > 0$ and every $t \in [0, T - \delta]$ the function $x \mapsto v(t, x)$ is Lipschitz continuous. Let ρ be the weight function given by $\rho(x) := 1_U(x)$. There is a measurable $N \subseteq \Omega$ such that $X^{s,\cdot}_t(\omega) \in \mathcal{W}^1_p(U)$ for all $\omega \in N^c$ and $P(N) = 0$. Thus, since $l = 1$ it follows by the chain rule formula of [39, Theorem 1.1] that for every $\omega \in N^c$ the function $Y^{s,x}_t(\omega) = v(t, X^{s,x}_t(\omega))$ belongs to the Sobolev space $\mathcal{W}^1_1(U)$.

When condition (B2) is satisfied, the function $x \mapsto v(t, x)$ is Lipschitz continuous for every $t \in [0, T]$. The claim (ii) then follows from the same arguments as above. □

4. Links to partial differential equations and stochastic control with rough coefficients

4.1. Link to partial differential equation

Since the FBSDE under consideration is Markovian, it can be argued that the solvability to (1.1) is a direct consequence of existence and uniqueness of solution of the partial differential equation (PDE)

$$\begin{cases} \partial_t v(t, x) + \mathcal{L}v(t, x) + g(t, x, v(t, x), D_x v(t, x)\sigma) = 0 \\ v(T, x) = h(x) \end{cases} \tag{4.1}$$

with

$$\mathcal{L}v := b(t, x, v, D_x v\sigma)D_x v + \frac{1}{2}\text{trace}(\sigma\sigma^* D_{xx}v)$$

in the classical, or Sobolev sense. While such equations with b non-smooth are well-studied, see e.g. [31, 32, 35, 52], it should be noted that all these references consider the linear case and most importantly, the terminal condition h is assumed to be smooth and the coefficients integrable enough.

In the case of semilinear PDEs, the authors in [25] study existence of local solutions for a wide class of non-smooth initial data and give sufficient conditions which guarantee the global existence of the solution to the PDE. Let us also mention the work [1] on general nonlinear non–degenerate parabolic equations. In both of these works, the authors assume that the initial data is in some suitable L^q space, with $q > 1$. For the case of nonlinear heat equation with integrable initial conditions, we refer the reader to [9, 22, 61, 62] and references therein. Using the notion of stable sets introduced in [57], many authors studied existence of global solution of semilinear heat and wave equations. For example, in [26], those stable stable and unstable sets were characterized by the asymptotic behavior of solutions (as $t \mapsto \infty$) of the semilinear PDE with non–smooth initial condition (corresponding to terminal condition in the present case). When the initial condition is a Radon measure there has been a lot of work studying sufficient conditions on the measure under which the PDE (degenerate or not) has a local or global solution. See for example [2, 27, 58, 59] and references therein. We also refer the reader to the work [28] with the non-linear term containing a distribution coefficient in a Besov space. Observe that the above works differ from ours in many ways: we do not assume any (weak) differentiability or continuity of the non linear terms in (t, x). In addition, since the terminal condition is measurable and bounded and not automatically integrable the above techniques are not directly applicable to our setup, and perhaps require additional technical localization arguments.

Some works also consider stochastic PDEs (SPDEs) with irregular coefficients. In [18] an SPDE corresponding to the PDE studied in the present paper, but with multiplicative Brownian noise is studied, and existence and uniqueness results are derived (see also [63] for the L^p-theory approach to the existence of such equations) assuming that the coefficients do not depend on $D_x v$. The above results were generalized in [19] to the quasilinear SPDEs and the authors prove existence, uniqueness and L^p–estimate for the weak solution. Their method relies on a version of Moser's iteration. In [14] quasilinear SPDEs for jump diffusions are investigated and comparison results are derived. In these works, the initial condition is assumed non-smooth. See also the work [20] for the case of SPDEs with obstacles. To the best of our knowledge the case of SPDEs with coefficients satisfying the conditions in the current paper has not been studied and is beyond the scope of this work. It is possible that combining our ideas with those introduced in [45] could allow to tackle this problem.

Having a non-smooth terminal condition pauses important difficulties as observed in the proofs above. Restricting x to a compact space would make the coefficients integrable, but in that case, boundary conditions should be added to the PDE (4.1) further complicating its analysis.

As the reader will have observed, PDEs still play an essential role in our argument since after mollification of the coefficients of the FBSDEs, we derive a decoupling field via the solution of the second order parabolic equation (3.2). A priori estimates on the gradient of the solution of this equation were crucial. Notice, however, that since the terminal condition h is not taken regular, the gradient of the decoupling field is not necessary bounded on the whole interval $[0, T]$. This prevents us from obtaining compactness of the sequence of decoupling fields in a Sobolev space. For instance, a solution to the PDE (4.1) would be hard to derive by such direct arguments. Therefore, it is interesting to point-out that our results should allow to prove existence of (4.1) (under conditions (A1)-(A4) and (B1) or (B2)) at least in the L^p-viscosity sense, and in one dimension.

4.2. Link with stochastic control

In this final subsection, we explain, at least informally, how our results apply to the study of optimal stochastic control of systems with rough coefficients.

Consider the control problem

$$\begin{cases} \inf_\alpha E\Big[g(X_T^\alpha) + \int_0^T f(u, X_u^\alpha, \alpha_u)\, du\Big] \\ dX_t^\alpha = b(u, X_u^\alpha, \alpha_u)\, du + \sigma\, dW_t, \quad X_0^\alpha = x, \end{cases}$$

where the infimum is over $\mathbb{R}^m$-valued square integrable progressive processes α and $b, f : [0,T] \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, and g are given Lipschitz continuous functions. In such a setting where the coefficients b, f, g are not smooth functions, the maximum principle was established by [6, 8, 48]. These authors showed that if an admissible control $\hat{\alpha}$ is optimal, then it holds

$$\hat{\alpha}_t \in \arg\min_{a \in \mathbb{R}^m} H(t, \hat{X}_t, \hat{Y}_t, a),$$

with $d\hat{X}_t = b(t, \hat{X}_t, \hat{\alpha}_t)\, dt + \sigma\, dW_t$, $\hat{X}_0 = x$, $(\hat{Y}, \hat{Z})$ solves the adjoint equation

$$d\hat{Y}_t = -D_x H(t, \hat{X}_t, \hat{Y}_t, \hat{\alpha})\, dt + \hat{Z}_t\, dW_t, \quad \hat{Y}_T = D_x g(\hat{X}_T)$$

and the function H is the Hamiltonian given by

$$H(t, x, y, a) := b(t, x, a) \cdot y + f(t, x, a).$$

Since the functions b, f and g are only Lipschitz continuous, the derivatives $D_x H$ and $D_x g$ are to be understood for almost every x. Under standard convexity conditions, it can be showed that $\hat{\alpha} = \Lambda_t(\hat{X}_t, \hat{Y}_t)$ for some Lipschitz continuous function Λ. Thus, it order for an optimal $\hat{\alpha}$ to exist, there must be a solution $(\hat{X}, \hat{Y}, \hat{Z})$ of the FBSDE

$$\begin{cases} d\hat{X}_t = b(t, \hat{X}_t, \Lambda_t(\hat{X}_t, \hat{Y}_t))\, dt + \sigma\, dW_t \\ d\hat{Y}_t = -D_x b(t, \hat{X}_t, \Lambda_t(\hat{X}_t, \hat{Y}_t)) \cdot \hat{Y}_t - D_x f(t, \hat{X}_t, \Lambda_t(\hat{X}_t, \hat{Y}_t))\, dt + \hat{Z}_t\, dW_t \\ \hat{Y}_T = D_x g(\hat{X}_T) \quad \hat{X}_0 = x. \end{cases} \tag{4.2}$$

When the drift, generator and terminal value of the above FBSDE are Lipschitz continuous, then a unique solution exists, see works by Delarue [15], Ma et al. [43]. Lipschitz continuity of the coefficients of the FBSDE require, in particular, the second derivatives of the functions b, f and g to exist (at least almost surely) and to be bounded. When the *second derivatives* of b and f are only assumed continuous, an existence result for FBSDEs as (4.2) is proved by Antonelli and Hamadène [4]. However, this paper assumes $\hat{Y}$ to be one-dimensional and the generator monotone in y. Our main result requires *only the first derivatives* if the coefficient of the control problem to exit and be bounded, $\hat{Y}$ can be *multi-dimensional* and non monotonicity assumptions are needed.

A. A priori estimations for quasi-linear PDEs

For the reader's convenience, in this appendix we collect some a priori estimations for quasi-linear PDEs. These are fundamental for the proofs of our main results. Different versions of these estimates can be found e.g. in [16, 40, 42] or [37]. The results we present here are taken from [16, 42].

Recall that the Sobolev space $\mathcal{W}^{1,2}_{p,\mathrm{loc}}((0,T) \times \mathbb{R}^d, \mathbb{R}^l)$ is the space of all functions $u : (0,T) \times \mathbb{R}^d \to \mathbb{R}^l$ such that for all $r > 0$,

$$\int\limits_{(0,T) \times B_r(0)} \Big(|u|^p + |\partial_t u|^p + |D_x u|^p + |D_{xx} u|^p\Big)\, dx\, dt < \infty$$

and consider the quasilinear parabolic PDE

$$\begin{cases} \partial_t v(t,x) + \mathcal{L}v(t,x) + g(t,x,v(t,x),D_x v(t,x)\sigma) = 0 \\ v(T,x) = h(x) \end{cases} \tag{A.1}$$

where $\mathcal{L}$ is the second order differential operator

$$\mathcal{L}v := b(t,x,v,D_x v\sigma)D_x v + \frac{1}{2}\text{trace}(\sigma\sigma^* D_{xx}v).$$

Theorem A.1. *([42, Theorem 3.1 & Lemma 6.2]) Assume that the conditions (A1)-(A4) are satisfied, and further assume that the functions b, g and h are bounded, smooth and with bounded derivatives. Let v be the unique classical solution of (A.1). Then for any $\delta > 0$ there is $\alpha \in (0,1)$ and constants C, C_δ and $C_{\delta,\alpha}$ depending on $k_1, k_2, k_3, \Lambda, T, l$, m, and the bound of b, g and which do not depend on the derivatives of b, g such that*

(i) $|D_x v(t,x)| \leq C_\delta$ *for all* $(t,x) \in [0, T-\delta] \times \mathbb{R}^d$.

(ii) for all $(t,x),(t',x') \in [0,T-\delta] \times \mathbb{R}^d$, *it holds that*

$$|D_x v(t,x) - D_x v(t',x')| \leq C_{\delta,\alpha}(|x-x'|^\alpha + |t-t'|^{\alpha/2}).$$

(iii) for every bounded domain $\mathcal{O} \subseteq \mathbb{R}^d$ *and* $p \geq 2$ *it holds*

$$\int_0^{T-\delta}\int_{\mathcal{O}} \Big[|D_x v(t,x)|^p + |D_{xx}v(t,x)|^p\Big]\,dx\,dt \leq C_\delta^p |\mathcal{O}|,$$

where $|\mathcal{O}|$ *is the Lebesgue measure of* $\mathcal{O}$.

If h is twice continuously differentiable with bounded first and second derivatives, then (i), (ii) and (iii) hold with $\delta = 0$ and C_0 may depend on $\|D_x h\|_\infty$ and $\|D_{xx}h\|_\infty$ as well.

Theorem A.2. *([16, Theorems 1.3 & 2.9]) Assume that the conditions (A1)-(A4) are satisfied and that h is α-Hölder continuous. Let v be a solution of (A.1) in the space $\mathcal{W}^{1,2}_{d+1,\mathrm{loc}}((0,T)\times\mathbb{R}^d, \mathbb{R}^l)$. Then there are constants $C > 0$ and $\alpha' \in (0,\alpha]$ depending only on $k_1, k_2, k_3, \Lambda, T, l$ and m such that*

$$|v(t,x) - v(t',x')| \leq C(|x-x'|^{\alpha'} + |t-t'|^{\alpha'/2})$$

for every $(t,x),(t',x') \in [0,T]\times\mathbb{R}^d$. If $\alpha = 1$, then it holds that

$$|D_x v(t,x)| \leq C \quad \text{for every } (t,x) \in [0,T]\times\mathbb{R}^d.$$

References

[1] H. Amann. Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. In H. Schmeisser and T. H. Triebel, editors, *Function spaces, differential operators and nonlinear analysis,*, volume 133, pages 9–126. Teubner-Texte zur Mathematik, 1993.

[2] D. Andreucci. Degenerate parabolic equations with initial data measures. *Trans. Amer. Math. Soc.*, 349:3911–3923, 1997.

[3] S. Ankirchner, P. Imkeller, and G. Dos Reis. Classical and variational differentiability of BSDEs with quadratic growth. *Elect. J. Probab.*, 12:1418–1453, 2007.

[4] F. Antonelli and S. Hamadène. Existence of solutions of backward-forward SDEs with continuous monotone coefficients. *Stat. Probab. Letters*, 76(14):1559–1569, 2006.

[5] K. Bahlali. Flows of homeomorphisms of stochastic differential equations with measurable drifts. *Stochastics*, 67(1):53–82, 1999.

[6] K. Bahlali, B. Djehiche, and B. Mezerdi. On the stochastic maximum principle in optimal control of degenerate diffusions with lipschitz coefficients. *Appl. Math. Optim.*, 56(364-378), 2007.

[7] K. Bahlali, M. Eddahbi, and Y. Ouknine. Quadratic BSDE with $\mathbb{L}^2$-terminal data: Krylov's estimates, Itô-Krylov's formula and existence results. *Ann. Probab.*, 45(4):2377–2397, 2017.

[8] S. Bahlali, B. Djehihe, and B. Mezerdi. The relaxed stochastic maximum principle in singular optimal control of diffusions. *SIAM J. Control Optim.*, 46(2):427–444, 2007.

[9] H. Brezis and T. Cazenave. A nonlinear heat equation with singular initial data. *J. Anal. Math.*, 68:277–304, 1996.

[10] R. Buckdahn and H.-J. Engelbert. A backward stochastic differential equation without strong solution. *Teor. Veroyatn. Primen.*, 50(2):390–396, 2005.

[11] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications. I*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. ISBN 978-3-319-56437-1; 978-3-319-58920-6. Mean field FBSDEs, control, and games.

[12] J. Cvitanic and J. Zhang. *Contract theory in continuos Time models*. Springer Finance, Heidelberg, 2012.

[13] G. Da Prato, P. Malliavin, and D. Nualart. Compact families of Wiener functionals. *C. R. Acad. Sci. Paris*, 315:1287–1291, 1992.

[14] K. A. Dareiotis and Y. Gyöngy. A comparison principle for stochastic integro-differential equations. *Potential Anal.*, 41: 1203–1222, 2014.

[15] F. Delarue. On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. *Stoch. Proc. Appl.*, 99:209–286, 2002.

[16] F. Delarue. Estimates of the solutions of a system of quasi-linear PDEs. A probabilistic scheme. *Séminaire de Probabilité (Strasbourg)*, XXXVII:290–332, 2003.

[17] F. Delarue and G. Guatteri. Weak existence and uniqueness for forward-backward SDEs. *Stoch. Proc. Appl.*, 116:1712–1742, 2006.

[18] L. Denis and L. Stoica. A general analytical result for non-linear s.p.d.e.'s and applications. *Elect. J. Probab.*, 9:674–709, 2004.

[19] L. Denis, A. Matoussi, and L. Stoica. L^p estimates for the uniform norm of solutions of quasilinear SPDE's. *Probab. Theory Related Fields*, 133:437–463, 2005.

[20] L. Denis, A. Matoussi, and J. Zhang. Maximum principle for quasilinear stochastic PDEs with obstacle. *Elect. J. Probab.*, 19: 1–32, 2014.

[21] A. Fromm, P. Imkeller, and D. J. Prömel. An FBSDE approach of the Skorokhod embedding problem for Gaussian processes with non-linear drift. *Elect. J. Probab.*, 20(127):1–38, 2015.

[22] Y. Giga. Solutions for semilinear parabolic equations in l^p and regularity of weak solutions of the Navier-Stokes system. *Journal of Differential Equations*, 62:186–212, 1986.

[23] G. Heyne, M. Kupper, and L. Tangpi. Portfolio optimization under nonlinear utility. *Int. J. Theor. Appl. Finance*, 19(5), 2016.

[24] U. Horst, Y. Hu, P. Imkeller, A. Réveillac, and J. Zhang. Forward backward systems for expected utility maximization. *Stoch. Proc. Appl.*, 124(5):1813–1848, 2014.

[25] H. Hoshino and Y. Yamada. Solvability and smoothing effect for semilinear parabolic equations. *Funkcialaj Ekvacioj*, 34: 475–494, 1991.

[26] R. Ikehata and T. Suzuki. Stable and unstable sets for evolution equations of parabolic and hyperbolic type. *Hiroshima Math. J.*, 26:475–491, 1996.

[27] K. Ishige, T. Kawakami, and S. Okabe. Existence of solutions for a higher-order semilinear parabolic equation with singular initial data. *Ann. Inst. Henri Poincare C, Ana. non linéaire*, 37:1185–1209, 2020.

[28] E. Issoglio. A non-linear parabolic PDE with a distributional coefficient and its applications to stochastic analysis. *Journal of Differential Equations*, 267:5976–6003, 2019.

[29] E. Issoglio and S. Jing. Forward-backward SDEs with distributional coefficients. *Stoch. Proc. Appl.*, 130:47–78, 2020.

[30] N. E. Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7:1–77, 1997.

[31] D. Kim and N. V. Krylov. Parabolic equations with measurable coefficients. *Potential Anal.*, 26(4):345–361, 2007.

[32] I. Kim, K.-H. Kim, and S. Lim. An $L_q(L_p)$-theory for the time fractional evolution equations with variable coefficients. *Adv. Math.*, 306(123-176), 2017.

[33] M. Kobylanski. Backward stochastic differential equations and partial differential equations with quadratic growth. *Ann. Probab*, 28(2):558–602, 2000.

[34] N. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, 131:154–196, 2005.

[35] N. V. Krylov. An analytic approach to SPDEs. In R. Carmona and B. Rozovsky, editors, *Stochastic Partial Differential Equations: Six Perspectives*. American Mathematical Society, 1999.

[36] M. Kupper, P. Luo, and L. Tangpi. Multidimensional Markovian FBSDEs with superquadratic growth. *Stoch. Proc. Appl., to appear*, 2018.

[37] O. Ladyzhenskaya, V. Solonnikov, and N. Ural'tseva. *Linear and Quasi-Linear Equations of Parabolic Type*. Translation of Mathematical Monographs. American Mathematical Society, 1968.

[38] M. Laurière and L. Tangpi. Convergence of large population games to mean field games with interaction through the controls. *Preprint*, 2020.

[39] G. Leoni and M. Morini. Necessary and sufficient conditions for the chain rule in $W^{1,1}_{loc}(\mathbb{R}^N;\mathbb{R}^d)$ and $BV_{loc}(\mathbb{R}^N;\mathbb{R}^d)$. *J. Eur. Math. Soc.*, 9:219–252, 2005.

[40] G. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific Publishing Co. Inc., 1996.

[41] P. Luo and L. Tangpi. Solvability of coupled FBSDEs with diagonally quadratic generators. *Stoch. Dyn.*, 17(6):1750043, 2017.

[42] J. Ma and J. Zhang. On weak solutions of forward-backward SDEs. *Probab. Theory Related Fields*, 151:475–507, 2011.

[43] J. Ma, P. Protter, and J. Yong. Solving forward-backward stochastic differential equations explicitly: a four step scheme. *Probab. Theory Related Fields*, 98:339–359, 1994.

[44] J. Ma, J. Zhang, and Z. Zheng. Weak solutions for forward-backward SDEs - a martingale problem approach. *Ann. Probab.*, 36 (6):2092–2125, 2008.

[45] A. Matoussi, D. Possamaï, and Sabbagh. Probabilistic interpretation for solutions of fully nonlinear stochastic PDEs. *Probab. Theory Related Fields*, 174:177–233, 2019.

[46] O. Menoukeu-Pamen and L. Tangpi. Strong solutions of some one-dimensional SDE's with random unbounded drifts. *SIAM J. Math. Anal.*, 51(5):4105–4141, 2019.

[47] O. Menoukeu-Pamen, T. Meyer-Brandis, T. Nilssen, F. Proske, and T. Zhang. A variational approach to the construction and malliavin differentiability of strong solutions of SDE's. *Math. Ann.*, 357(2):761–799, 2013.

[48] B. Mezerdi. Necessary conditions for optimality for a diffusion with a non-smooth drift. *Stochastics*, 24(305-326), 1988.

[49] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stoch. Proc. Appl.*, 116(12):1815–1835, 2006.

[50] S. E. A. Mohammed, T. Nilssen, and F. Proske. Sobolev differentiable stochastic flows for sde's with singular coeffcients: Applications to the stochastic transport equation. *Ann. Probab.*, 43(3):1535–1576, 2015.

[51] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer Berlin, 2nd edition edition, 2006.

[52] D. K. Palagachev. Quasilinear elliptic equations with VMO coefficients. *Trans. A.M.S.*, pages 2481–2493, 1995.

[53] E. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In B. Rozuvskii and R. Sowers, editors, *Stochastic partial differential equations and their applications*, volume 176, pages 200–217. Springer, Berlin, New York, 1992.

[54] E. Pardoux and S. Tang. Forward-backward stochastic differential equations and quasilinear parabolic PDEs. *Probab. Theory Related Fields*, 114(2):123–150, 1999.

[55] S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.*, 28:966–979, 1990.

[56] S. Peng and Z. Wu. Fully coupled forward-backward stochastic differential equations and applications to optimal control. *SIAM J. Control Optim.*, 37(3):825–843, 1999.

[57] D. H. Sattingeri. On global solution of nonlinear hyperbolic equations. *Arch. Rat. Mech. Anal.*, 30:148–1721, 1968.

[58] H. Shang and F. Li. On the cauchy problem for the evolution p-Laplacian equations with gradient term and source and measures as initial data. *Nonlinear Analysis: Theory, Methods & Applications*, 72:3396–3411, 2010.

[59] J. Takahashi. Solvability of a semilinear parabolic equation with measures as initial data. In F. Gazzola, K. Ishige, C. Nitsch, and P. Salani, editors, *Geometric Properties for Parabolic and Elliptic PDE's*, pages 257–276. Springer, 2015.

[60] A. S. Üstünel and M. Zakai. *Transformation of Measure on Wiener Space*. Springer Monographs in Mathematics. Springer, 2000.

[61] F. B. Weissler. Semilinear evolution equations in Banach spaces. *J. Funct. Anal.*, 32:277–296, 1979.

[62] F. B. Weissler. Local existence and nonexistence for semilinear parabolic equations in L^p. *Indiana Univ. Math. J.*, 29:79–102, 1980.

[63] X. Zhang. L^p-theory of semi-linear SPDEs on general measure spaces and applications. *J. Funct. Anal.*, 239:44–75, 2006.

Acknowledgment: The authors thank two anonymous referees and an Associate Editor for their careful reading of the manuscript and their suggestions which improved the paper.

Peng Luo: School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China. *E-mail address:* peng.luo@sjtu.edu.cn
Financial support from the National Natural Science Foundation of China (Grant No. 12101400) is gratefully acknowledged

Olivier Menoukeu-Pamen: University of Liverpool Institute for Financial and Actuarial Mathematics, Department of Mathematical Sciences, L69 7ZL, United Kingdom and African Institute for Mathematical Sciences, Ghana. *E-mail address:* menoukeu@liverpool.ac.uk
Financial support from the Alexander von Humboldt Foundation, under the programme financed by the German Federal Ministry of Education and Research entitled German Research Chair No 01DG15010 is gratefully acknowledged.

Ludovic Tangpi: Department of Operations Research and Financial Engineering, Princeton University, Princeton, 08540, NJ; USA. *E-mail address:* ludovic.tangpi@princeton.edu
Financial suupport by NSF grant DMS-2005832 is gratefully acknowledged.