The Annals of Applied Probability
2015, Vol. 25, No. 5, 2535–2562
DOI: 10.1214/14-AAP1055

WEAK APPROXIMATION OF SECOND-ORDER BSDES

BY DYLAN POSSAMAÏ AND XIAOLU TAN

University of Paris-Dauphine

We study the weak approximation of the second-order backward SDEs (2BSDEs), when the continuous driving martingales are approximated by discrete time martingales. We establish a convergence result for a class of 2BSDEs, using both robustness properties of BSDEs, as proved in Briand, Delyon and Mémin [*Stochastic Process. Appl.* **97** (2002) 229–253], and tightness of solutions to discrete time BSDEs. In particular, when the approximating martingales are given by some particular controlled Markov chains, we obtain several concrete numerical schemes for 2BSDEs, which we illustrate on specific examples.

1. Introduction. Weak approximation is an important technique in stochastic analysis. A famous and classical result in this spirit is Donsker's theorem which stipulates the following. Let $(\zeta_k)_{k\geq 1}$ be a sequence of i.i.d. centered random variables such that $\mathrm{Var}(\zeta_1) = 1$, and define

$$S_t^n := \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \zeta_k,$$

then the process $S_\cdot^n$ converges weakly to a Brownian motion W. In particular, suppose that $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a bounded continuous function, we then have the following convergence:

$$\mathbb{E}[f(S_T^n)] \to \mathbb{E}[f(W_T)].$$

Similar result have been obtained for diffusion processes defined as solutions to stochastic differential equations (SDEs in the sequel); see, for example, Jacod and Shiryaev [15]. We also remind the reader that in this Markovian setting, the value $\mathbb{E}[f(W_T)]$ can be characterized using the heat equation from the Feynmann–Kac formula.

Backward stochastic differential equations (BSDEs in the sequel), which were introduced by Pardoux and Peng [20], as well as the more recent notion of G-expectation of Peng [21], are particular cases of so-called nonlinear expectations, and their weak approximation properties have attracted a lot of attention in the recent years. Hence, in Briand, Delyon and Mémin [6], the authors studied the

Received September 2013; revised May 2014.
MSC2010 subject classifications. Primary 60F05; secondary 93E15, 65C50.
Key words and phrases. Second-order BSDEs, weak approximation, numerical scheme, robustness of BSDE.

convergence of the solutions of the BSDE when the driving Brownian motion is approximated by a sequence of martingales. In particular, when the Brownian motion is approximated by some random walks, they obtained a weak convergence result similar to the above Donsker's theorem. More recently, Dolinsky, Nutz and Soner [9] studied the weak approximation of G-expectation. Since G-expectation can be considered as a sublinear expectation on the canonical space of continuous trajectories, by the analogue of Donsker's theorem, they approximated it by a sequence of sublinear expectations on the canonical space of discrete time paths. Extending BSDE and G-expectation, the second-order backward SDEs (2BSDEs) introduced by Soner, Touzi and Zhang [23], can be represented as the supremum of a family of nonlinear expectations on the canonical space of continuous trajectories. In particular, it generalizes the Feynmann–Kac formula to the fully nonlinear case. We are then motivated to extend the weak approximation property to 2BSDEs.

We notice that the weak approximation property should be an important property of the continuous time dynamic models, when it is the continuous limit of discrete time models. For example, in finance, it is convenient to use a Brownian motion to model the evolution of a risky asset, despite the fact that such a price only exists on discrete time instants. Therefore, it is important to confirm that as we take the limit of the discrete time model, it converges to the continuous time model.

Finally, weak approximation is also an important technique in numerical analysis; see, for example, Kushner and Dupuis [17] in the context of stochastic control problems, and Dolinsky [8] for pricing the financial "game" options. The main idea is to interpret the numerical scheme as a controlled Markov chain system, which converges weakly to the continuous time system. We notice also that another point of view is from the PDEs, which characterizes the solution of these dynamic problems in the Markovian case. A powerful numerical analysis method in this context is the monotone convergence theorem of Barles and Souganidis [1]. Comparing to the PDE numerical methods, the weak approximation method permits usually to relax regularity and integrability conditions, and also permits to study the non-Markovian problems as shown in Tan [26].

The main contribution of the paper is to prove a weak approximation property for a class of 2BSDEs, which can be considered as an extension of Donsker's theorem in this nonlinear context. Further, using some controlled Markov chains as approximating martingales, we obtain some numerical schemes for a class of 2BSDEs. In particular, these numerical schemes are coherent with the classical schemes proposed for the nonlinear PDEs in the Markovian cases. We also notice that these related numerical schemes have been largely tested in the previous literature; see, for example, Fahim, Touzi and Warin [12], Tan [25], Guo, Zhang and Zhuo [13], etc.

The rest of the paper is organized as follows. In Section 2, we introduce the class of 2BSDEs that is studied in the paper, and give first an equivalence result using two different classes of driving martingales. By considering a sequence of discrete time equations, we give a general weak approximation result, that is, the discrete time solution converges to the solution of a class of 2BSDE. Then in Section 3, by considering some particular controlled Markov chains, we can interpret the discrete time equations as numerical schemes, and the weak approximation result justifies the convergence of the numerical schemes. Section 3.3 is devoted to some numerical examples, highlighting the convergence of the proposed numerical schemes. In Section 4.1, we complete the proof of the equivalence theorem, and finally in Section 4.2, we report the proof of the weak approximation theorem.

Throughout the paper, we use the following notation. For every $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, we denote by $x \cdot y$ the usual scalar product of x and y, and for any $(x, y) \in \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d}$, we denote by $x : y := \mathrm{Tr}(xy)$. Similarly, x^T will denote the usual transposition and $|x|$ the Euclidean norm in the corresponding space.

2. The 2BSDE and its weak approximation.

In this section, we first introduce the class of second-order BSDEs that we next propose to approximate by the supremum of a family of BSDEs driven by approximating discrete time martingales. A convergence result is given under sufficient conditions, while the proof is postponed to other sections.

2.1. *A class of 2BSDEs.* Let $\Omega := \{\omega \in C([0, T], \mathbb{R}^d) : \omega_0 = 0\}$ denote the canonical space of continuous paths on $[0, T]$ which start at 0, B be the canonical process, $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ the canonical filtration and $\mathbb{P}_0$ the Wiener measure on Ω under which B is a standard Brownian motion. Denote by $\mathbb{F}^+ = (\mathcal{F}_t^+)_{0\le t\le T}$ the right-continuous filtration defined by $\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s$ for all $t < T$ and $\mathcal{F}_T^+ = \mathcal{F}_T$. For every probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}_T)$, we denote by $\overline{\mathbb{F}}^{\mathbb{P}}$ the $\mathbb{P}$-augmented filtration of $\mathbb{F}$ and $\overline{\mathbb{F}^+}^{\mathbb{P}}$ the $\mathbb{P}$-augmented filtration of $\mathbb{F}^+$. Moreover, for any $\mathbf{x} \in \Omega$, and for any $t \in [0, T]$, we note $\|x\|_t := \sup_{0\le s\le t} |x_s|$. A probability measure $\mathbb{P}$ on Ω such that B is a $\mathbb{P}$-local martingale will be called a local martingale measure.

We recall that by results of Bichteler [3] (see also Karandikar [16] for a simplified exposition) there are two $\mathbb{F}$-progressive processes on Ω given by

$$\langle B\rangle_t := B_t B_t^T - 2\int_0^t B_s\, dB_s^T \quad \text{and} \quad \widehat{a}_t := \limsup_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\big(\langle B\rangle_t - \langle B\rangle_{t-\varepsilon}\big),$$

such that $\langle B\rangle$ coincides with the $\mathbb{P}$-quadratic variation of B, $\mathbb{P}$-a.s., for all local martingale measures $\mathbb{P}$.

We consider next a set A such that

$$A \subset \mathbb{S}_d^+ \text{ is compact, convex and } a \ge \varepsilon_0 I_d, \forall a \in A, \tag{2.1}$$

where $\mathbb{S}_d^+$ is the set positive, symmetric $d \times d$ matrices and where $\varepsilon_0 > 0$ is a fixed constant. We denote by $\mathcal{P}_W$ the collection of all local martingale measures $\mathbb{P}$ such that $\widehat{a} \in A$, $d\mathbb{P} \times dt$-a.e., and by $\mathcal{P}_S \subset \mathcal{P}_W$ the subset consisting of all probability measures

$$\mathbb{P}^\alpha := \mathbb{P}_0 \circ \left(X^\alpha\right)^{-1} \qquad \text{where } X_t^\alpha := \int_0^t \alpha_s^{1/2}\, dB_s, \mathbb{P}_0\text{-a.s.}$$

for some $\mathbb{F}$-progressively measurable process α taking values in A.

Let now $\xi : \Omega \to \mathbb{R}$ be a random variable, $g : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d^+ \to \mathbb{R}^{d\times d}$ be a function which will play the role of our generator. Then for every $\mathbb{P} \in \mathcal{P}_W$, we consider the following generalized BSDE under $\mathbb{P}$:

$$\begin{aligned} \mathcal{Y}_t^{\mathbb{P}} &= \xi(B_\cdot) - \int_t^T g\big(s, B_\cdot, \mathcal{Y}_s^{\mathbb{P}}, \mathcal{Z}_s^{\mathbb{P}}, \widehat{a}_s\big) : d\langle B\rangle_s \\ &\quad - \int_t^T \mathcal{Z}_s^{\mathbb{P}} \cdot dB_s - \mathcal{N}_T^{\mathbb{P}} + \mathcal{N}_t^{\mathbb{P}}, \end{aligned} \tag{2.2}$$

whose solution is a triple of $\overline{\mathbb{F}^+}^{\mathbb{P}}$-progressive processes, denoted by $(\mathcal{Y}^{\mathbb{P}}, \mathcal{Z}^{\mathbb{P}}, \mathcal{N}^{\mathbb{P}})$, such that $\mathcal{N}^{\mathbb{P}}$ is a $\overline{\mathbb{F}^+}^{\mathbb{P}}$-martingale orthogonal to B and (2.2) holds true $\mathbb{P}$-a.s. We shall assume sufficient conditions (see Assumption 2.2 below) to guarantee the existence and uniqueness of the solution to (2.2) under every $\mathbb{P} \in \mathcal{P}_W$. In particular, whenever $\mathbb{P} \in \mathcal{P}_S$, (2.2) turns out to be a classical BSDE whose solution satisfies $\mathcal{N}^{\mathbb{P}} = 0$ and $\mathcal{Y}^{\mathbb{P}}$, $\mathcal{Z}^{\mathbb{P}}$ are $\overline{\mathbb{F}}^{\mathbb{P}}$-progressive. This is due to the fact that by Lemma 8.2 in [22], every probability measures in $\mathcal{P}_S$ satisfies the predictable martingale representation property and the Blumenthal 0–1 law. This also implies in this case that $\mathcal{Y}_0^{\mathbb{P}}$ is a deterministic constant.

The main purpose of the paper is to study the weak approximation of the following optimization problem:

$$Y_0 := \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}}. \tag{2.3}$$

REMARK 2.1. The above problem Y_0 in (2.3) is related to the solution of the following 2BSDE, in the sense that Y_0 is the initial value of the Y component of its solution

$$\begin{aligned} Y_t &= \xi(B_\cdot) - \int_t^T \big(g(s, B_\cdot, Y_s, Z_s, \widehat{a}_s) : \widehat{a}_s\big)\, ds \\ &\quad - \int_t^T Z_s \cdot dB_s + K_T - K_t, \qquad \mathcal{P}_S\text{-q.s.,} \end{aligned}$$

which has been introduced by Soner, Touzi and Zhang [23]. We also refer to their Section 3.3 for more details, and simply emphasize here that given the boundedness assumptions we make below, it is not necessary in our setting to work on the subset $\mathcal{P}_H^\kappa$ of $\mathcal{P}_S$ introduced in [23]. We would also like to comment on the fact that in [23], the solution (Y, Z) is $\mathbb{F}^+$-progressive, while we defined the solution to the BSDE (2.2) to be $\overline{\mathbb{F}^+}^{\mathbb{P}}$-progressive. However, thanks to Lemma 2.4

of [22], for any $\mathbb{P} \in \mathcal{P}_W$, any $\overline{\mathbb{F}^+}^{\mathbb{P}}$-progressive process X has a $\mathbb{P}$-version $\widetilde{X}$ which is $\mathbb{F}^+$-progressive, so that this is not a real difference.

We shall impose the following assumptions on the terminal function ξ and generator function g throughout the paper. For ease of notation, and since this function will be the main focus of our paper, we define the function $f : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d^+ \to \mathbb{R}$

$$f(t, \mathbf{x}, y, z, u) := g(t, \mathbf{x}, y, z, u) : u.$$

ASSUMPTION 2.2. (i) $\xi : \Omega \longrightarrow \mathbb{R}$ is a bounded Lipschitz continuous function.

(ii) The process $t \longmapsto f(t, X_\cdot, Y_t, Z_t, \nu_t)$ is progressively measurable given progressive processes (X, Y, Z, ν), and is uniformly continuous with modulus ρ in the sense that for every $s \le t$ and $\mathbf{x}, y, z, u$,

$$\left| f(t, \mathbf{x}_{s\wedge\cdot}, y, z, u) - f(s, \mathbf{x}_{s\wedge\cdot}, y, z, u)\right| \le \rho(t - s).$$

(iii) f is uniformly Lipschitz in $(\mathbf{x}, y, z)$, that is, for all $(t, \mathbf{x}_1, \mathbf{x}_2, y_1, y_2, z_1, z_2, u)$,

$$\left| f(t, \mathbf{x}_1, y_1, z_1, u) - f(t, \mathbf{x}_2, y_2, z_2, u)\right| \le \mu\left(\|\mathbf{x}_1 - \mathbf{x}_2\|_t + |y_1 - y_2| + |z_1 - z_2|\right),$$

for some constant $\mu > 0$.

(iv) The map $u \longmapsto f(t, \mathbf{x}, y, z, u)$ is convex and uniformly continuous for every $(t, \mathbf{x}, y, z) \in [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d$.

(v) We have the following integrability condition, for some constant $C > 0$:

$$\sup_{(t,\mathbf{x},u)\in[0,T]\times\Omega\times A} \left| f(t, \mathbf{x}, 0, 0, u)\right| \le C.$$

Let us give an existence and equivalence result on the above 2BSDE, whose proof is postponed to Section 4.1.

THEOREM 2.3. *Suppose that Assumption 2.2 holds true. Then for every $\mathbb{P} \in \mathcal{P}_W$, the BSDE (2.2) has a unique solution $(\mathcal{Y}^{\mathbb{P}}, \mathcal{Z}^{\mathbb{P}}, \mathcal{N}^{\mathbb{P}})$. Moreover, we have*

$$Y_0 := \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}} = \sup_{\mathbb{P}\in\mathcal{P}_W} \mathbb{E}^{\mathbb{P}}[\mathcal{Y}_0^{\mathbb{P}}]. \tag{2.4}$$

REMARK 2.4. Suppose that $\xi(\mathbf{x}) = \xi_0(\mathbf{x}_T)$ and $f(t, \mathbf{x}, y, z, u) = f_0(t, \mathbf{x}_t, y, z, u)$ for some deterministic functions $\xi_0 : \mathbb{R}^d \longrightarrow \mathbb{R}$ and $f_0 : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times A \longrightarrow \mathbb{R}$. In this Markovian case, the value function can be given as the viscosity solution $v(t, x)$ of the nonlinear equation

$$-\partial_t v - \sup_{a\in A}\left(\frac{1}{2} a : D^2 v - f_0(t, x, v, Dv, a)\right) = 0, \tag{2.5}$$

with terminal condition $v(T, x) = \xi_0(x)$. We refer the reader to the paper by Soner, Touzi and Zhang [23] for more information.

2.2. *Weak approximation of* 2*BSDEs*. Under every probability measure $\mathbb{P} \in \mathcal{P}_S$, the canonical process B is a continuous martingale, which drives the BSDE (2.2). When this martingale is approximated "weakly" by a sequence of martingales, it follows by the robustness property for BSDEs proved by Briand, Delyon and Mémin [6] that the corresponding solutions of the BSDEs driven by the approximating martingales converge to $\mathcal{Y}^{\mathbb{P}}$ (see their Theorem 12). In the context of 2BSDEs (2.3), the solution is given as the supremum of the family of solutions to BSDEs driven by the family of martingales $(B|_{\mathbb{P}})_{\mathbb{P}\in\mathcal{P}_S}$. Therefore, it is natural, in order to obtain weak approximation properties, to consider a sequence of families of BSDEs driven by approximating martingales. In particular, we shall consider a family of discrete time martingales, motivated by its application in the numerical approximation described in Section 3.

For every $n \geq 1$, we denote by $\Delta_n = (t_k^n)_{0\leq k\leq n}$ a discretization of $[0, T]$, such that $0 = t_0^n < t_1^n < \cdots < t_n^n = T$. Let $|\Delta_n| := \sup_{1\leq k\leq n}(t_k^n - t_{k-1}^n)$, and we suppose that $|\Delta_n| \longrightarrow 0$ as $n \longrightarrow \infty$. For ease of presentation, we shall simplify the notation of the time step size $\Delta t_k^n := t_k^n - t_{k-1}^n$ into Δt when there is no ambiguity. Similarly, we suppress the dependence in n of t_k^n and write instead t_k.

For every $n \geq 1$, let $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$ be a probability space containing n independent random variables $(U_k)_{1\leq k\leq n}$. Moreover, we consider a family of functions $(H_k^n)_{1\leq k\leq n, n\geq 1}$ such that every $H_k^n : A \times [0, 1] \longrightarrow \mathbb{R}^d$ is continuous in a and for some $\delta > 0$, we have for any a

$$\begin{aligned} \mathbb{E}\big[H_k^n(a, U_k)\big] = 0, \qquad \operatorname{Var}\big(H_k^n(a, U_k)\big) = a\Delta t, \\ \mathbb{E}\big[\big|H_k^n(a, U_k)\big|^{2+\delta}\big] \leq C\Delta t^{1+\delta/2}, \end{aligned} \tag{2.6}$$

where it is understood that the expectation is taken under $\mathbb{P}^n$.

Define the filtration $\mathbb{F}^n := (\mathcal{F}_{t_k}^n)_{1\leq k\leq n}$, with $\mathcal{F}_{t_k}^n := \sigma(U_1, \ldots, U_k)$ and denote by E_n the collection of all $\mathbb{F}^n$-predictable A-valued processes $e = (a_{t_1}^e, \ldots, a_{t_n}^e)$. Then for every $e \in E_n$, M^e is defined by

$$M_{t_k}^e := \sum_{i\leq k} H_i^n\big(a_{t_i}^e, U_i\big). \tag{2.7}$$

REMARK 2.5. An easy example is when U_k is a Gaussian random vector (d-dimension) with distribution $N(0, I_d)$ and $H_k^n(a, u) := au\Delta t$. More examples which induce several different numerical schemes will be given later in Section 3.

By abuse of notation, we define a continuous time filtration $\mathbb{F}^n = (\mathcal{F}_t^n)_{0\leq t\leq T}$, with $\mathcal{F}_t^n := \mathcal{F}_{t_k}^n$, $\forall t \in [t_k, t_{k+1})$ and a continuous time martingales $M_t^e := M_{t_k}^e$, for all $t \in [t_k, t_{k+1})$ on $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$. We next consider the completed filtration under $\mathbb{P}^n$, denoted by $\mathbb{G}^n := \overline{\mathbb{F}^n}^{\mathbb{P}^n}$. Clearly, $\mathbb{G}^n$ is right-continuous and complete under $\mathbb{P}^n$, and M^e is a right-continuous, piecewise constant in time, $\mathbb{G}^n$-martingale for every $e \in E_n$. We notice that the predictable quadratic variation of M^e is given by

$$\langle M^e \rangle_{t_k} = \sum_{i \leq k} \Delta \langle M^e \rangle_{t_k} = \sum_{i \leq k} a_i^e \Delta t_i.$$

For every $n \geq 1$, with the time discretization Δ_n, we introduce the truncated generator $f_n(t, \mathbf{x}, y, z, a) := g_n(t, \mathbf{x}, y, z, a) : a$ where

$$g_n(t, \mathbf{x}, y, z, a) := g(t_k, \mathbf{x}, y, z, a) \qquad \text{whenever } t \in [t_k, t_{k+1}).$$

Then for every $e \in E_n$ and $n \geq 1$, we consider the following BSDE:

$$\begin{aligned} \mathcal{Y}_t^e = \xi(\widehat{M}^e_\cdot) - \int_t^T g_n(s, \widehat{M}^e_\cdot, \mathcal{Y}^e_{s-}, \mathcal{Z}^e_s, a^e_s) : d\langle M^e \rangle_s \\ - \int_t^T \mathcal{Z}^e_s \cdot dM^e_s - \mathcal{N}^e_T + \mathcal{N}^e_t, \end{aligned} \tag{2.8}$$

whose solution is a triple of $\mathbb{G}^n$-progressive processes $(\mathcal{Y}^e, \mathcal{Z}^e, \mathcal{N}^e)$ such that $\mathcal{N}^e$ is a $\mathbb{G}^n$-martingale orthogonal to M^e, and where $\widehat{M}^e$ denotes the continuous interpolation of M^e on the interval $[0, T]$. We then have the following wellposedness result for the BSDE (2.8), which is a direct consequence of Proposition A.1 reported in Section 4.1 and the fact that by taking conditional expectation with respect to $\mathbb{G}^n$, the component the solution to (2.8) is given explicitly by the following scheme:

$$\begin{cases} \mathcal{Y}^e_{t_n} = \xi(\widehat{M}^e_\cdot), \\ \mathcal{Y}^e_{t_k} = \mathbb{E}^n_{t_k}[\mathcal{Y}^e_{t_{k+1}}] - f(t_k, \widehat{M}^e_\cdot, \mathcal{Y}^e_{t_k}, \mathcal{Z}^e_{t_k}, a^e_{t_k}) \Delta t, \\ \mathcal{Z}^e_{t_k} = \mathbb{E}^n_{t_k}\left[\dfrac{\Delta \mathcal{Y}^e_{t_{k+1}} (a^e_{t_k})^{-1} \Delta M^e_{k+1}}{\Delta t} \right], \\ \Delta \mathcal{N}^e_{t_{k+1}} = \mathcal{Y}^e_{t_{k+1}} - \mathbb{E}^n_{t_k}[\mathcal{Y}^e_{t_{k+1}}] - \mathcal{Z}^e_{t_k} \cdot \Delta M^e_{t_{k+1}}, \end{cases} \tag{2.9}$$

where $\mathbb{E}^n_{t_k}[\cdot]$ represents the conditional expectation w.r.t. $\mathcal{F}^n_{t_k}$.

LEMMA 2.6. *Suppose that Assumption 2.2 holds true. Then for every $n \geq 1$ and $e \in E_n$, there is a unique solution $(\mathcal{Y}^e, \mathcal{Z}^e, \mathcal{N}^e)$ to the BSDE (2.8) such that*

$$\mathbb{E}^{\mathbb{P}^e}\left[\sup_{0 \leq t \leq T} \left[|\mathcal{Y}^e_t|^2 + \int_0^t |(a^e_s)^{1/2} \mathcal{Z}^e_s|^2 \, ds + \langle \mathcal{N}^e_t \rangle \right] \right] \leq C,$$

for some constant C independent of e and n. In particular, $\mathcal{Y}^e_0$ is a deterministic constant.

PROOF. The existence and uniqueness is immediate by (2.9). Moreover, Proposition A.1 gives us the required estimate for $n \geq n_0$ for some n_0. Since only a finite number of values for n remains, the result is immediate by the fact that the solution given in (2.9) has the required integrability. $\square$

For every $n \geq 1$, denote now

$$Y_0^n := \sup_{e \in E_n} \mathcal{Y}_0^e. \tag{2.10}$$

The next assumption is a monotonicity condition for the discretized BSDEs.

ASSUMPTION 2.7. For every $e \in E_n$ and $n \geq 1$, the backward scheme in (2.9) is monotone, that is, let $(\mathcal{Y}^1, \mathcal{Z}^1)$, $(\mathcal{Y}^2, \mathcal{Z}^2)$ be two solutions of (2.9), then

$$\mathcal{Y}^1_{t_{k+1}} \leq \mathcal{Y}^2_{t_{k+1}} \quad \Longrightarrow \quad \mathcal{Y}^1_{t_k} \leq \mathcal{Y}^2_{t_k} \qquad \forall k = 0, \dots, n-1.$$

We now state our main result.

THEOREM 2.8. (i) *Suppose that Assumption 2.2 holds true. Then*

$$\liminf_{n \to \infty} Y_0^n \geq Y_0.$$

(ii) *Suppose in addition that Assumption 2.7 holds and f does not depend on z. Then*

$$\lim_{n \to \infty} Y_0^n = Y_0.$$

REMARK 2.9. We are not able to show (ii) when the generator depends on z. This is deeply linked to the fact that there are considerable difficulties to obtain any convergence results for the z part of the solution. Moreover, since we are working under many measures, the canonical process is no longer always a Brownian motion, which prevents us from recovering the strong regularity results of [27], for instance. We leave this open problem for future research.

In the case where $f = 0$, the solution of the 2BSDE is the so called G-expectation of Peng. Then, in particular, the above result generalizes the weak convergence result for G-expectation in Dolinsky, Nutz and Soner [9]. We shall report its proof later in Section 4.2.

REMARK 2.10. Let $(\mathcal{Y}^1, \mathcal{Z}^1)$, $(\mathcal{Y}^2, \mathcal{Z}^2)$ be two solutions of (2.9), we have then clearly

$$\begin{aligned} &(1 - L_{t_k,y} \Delta t)(\mathcal{Y}^1_{t_k} - \mathcal{Y}^2_{t_k}) \\ &\quad = \mathbb{E}^n_{t_k}\big[(\mathcal{Y}^1_{t_{k+1}} - \mathcal{Y}^2_{t_{k+1}})(1 + L_{t_k,z} \cdot (\alpha_{t_k})^{-1} \Delta M^e_{t_{k+1}})\big], \end{aligned} \tag{2.11}$$

where $L_{t_k,y}$ (resp., $L_{t_k,z}$) is a $\mathbb{R}$-valued (resp., $\mathbb{R}^d$-valued) and $\mathcal{F}^n_{t_k}$-measurable random variable bounded by the Lipschitz constant $L_{f,y}$ (resp., $L_{f,z}$). Then for Δt small enough, the monotonicity condition in Assumption 2.7 holds whenever

$$\big|L_{f,z} H^n_k(a_{t_k}, U_k)\big| \leq |a_{t_k}| \qquad \forall 1 \leq k \leq n.$$

In particular, when f is independent of z, Assumption 2.7 always holds true for Δt small enough.

3. Numerical schemes for 2BSDEs. As discussed in Remark 2.4, the solution of the Markovian 2BSDE (2.3) can be given as viscosity solution of a parabolic fully nonlinear PDE, for which a comparison principle holds. Several monotone numerical schemes have been proposed for PDEs in or closed to this form, for example, the generalized finite difference scheme of Bonnans, Ottenwaelter and Zidani [5], the semi-Lagrangian scheme of Debrabant and Jakobsen [7], and the probabilistic scheme of Fahim, Touzi and Warin [12], Guo, Zhang and Zhuo [13], where the convergence is ensured by the monotone convergence theorem of Barles and Souganidis [1].

Similar to Tan [26] in the context of non-Markovian control problems, we can interpret these schemes as a system of controlled Markov chains. Using these controlled Markov chains as the families of driving martingale $(M^e)_{e\in E_n}$ in (2.8), Theorem 2.8 also justifies the convergence of the corresponding numerical schemes. Moreover, it permits to extend these numerical schemes to the non-Markovian case. The aim of this section is to present a general abstract numerical scheme for 2BSDEs, which we then specialize in two particular examples. In particular, these schemes are coherent with the numerical methods proposed and tested in the previous literature, for which we can refer to [12, 13, 25], etc. We nonetheless start by studying the solution to the discrete-time BSDEs.

3.1. *An explicit scheme.* We notice that for every fixed $e \in E_n$ and $n \geq 1$, the backward iteration in (2.9) is in fact the so called implicit scheme for BSDEs. In practice, we consider also the following explicit scheme:

$$\begin{cases} \widetilde{\mathcal{Y}}^e_{t_n} = \xi(\widehat{M}^e_\cdot), \\ \widetilde{\mathcal{Y}}^e_{t_k} = \mathbb{E}^n_{t_k}[\widetilde{\mathcal{Y}}^e_{t_{k+1}}] - f(t_k, \widehat{M}^e_\cdot, \mathbb{E}^n_{t_k}[\widetilde{\mathcal{Y}}^e_{t_{k+1}}], \widetilde{\mathcal{Z}}^e_{t_{k+1}}, a^e_{t_k})\Delta t, \\ \widetilde{\mathcal{Z}}^e_{t_k} = \mathbb{E}^n_{t_k}\left[\dfrac{\Delta \mathcal{Y}^e_{t_{k+1}} (a^e_{t_k})^{-1} \Delta M^e_{k+1}}{\Delta t}\right]. \end{cases} \tag{3.1}$$

Denote

$$\widetilde{Y}^n_0 := \sup_{e\in E_n} \widetilde{\mathcal{Y}}^e_0. \tag{3.2}$$

The following lemma shows that the implicit and explicit schemes only differ by an amount proportional to Δ_n.

LEMMA 3.1. *There is a constant C independent of $n \geq 1$ such that*

$$|Y^n_0 - \widetilde{Y}^n_0| \leq C|\Delta_n|.$$

PROOF. It is enough to prove that there is some constant $C > 0$ independent of $n \geq 1$ and $e \in E_n$ such that

$$|\mathcal{Y}^e_0 - \widetilde{\mathcal{Y}}^e_0| \leq C|\Delta_n|.$$

First, by (2.9) and (3.1) and the Lipschitz property of the generator f, it is clear that for every $0 \le k \le n-1$, there are bounded $\mathcal{G}^n_{t_k}$-random variables α_k and β_k such that

$$\begin{aligned}(\mathcal{Y}^e_{t_k} - \widetilde{\mathcal{Y}}^e_{t_k}) &= \mathbb{E}^n_{t_k}[\mathcal{Y}^e_{t_{k+1}} - \widetilde{\mathcal{Y}}^e_{t_{k+1}}] + \alpha_k(\mathcal{Y}^e_k - \mathbb{E}^n_{t_k}[\widetilde{\mathcal{Y}}^e_{t_{k+1}}]) + \beta_k \cdot (\mathcal{Z}^e_{t_{k+1}} - \widetilde{\mathcal{Z}}^e_{t_{k+1}}) \\ &= (1+\alpha_k\Delta t)\mathbb{E}^n_{t_k}\big[(\mathcal{Y}^e_{t_{k+1}} - \widetilde{\mathcal{Y}}^e_{t_{k+1}})\big(1+(1+\alpha_k\Delta t)^{-1}\beta_k \cdot \Delta M_{k+1}\big)\big] \\ &\quad + f\big(t_k, \widehat{M}^e_\cdot, \mathcal{Y}^e_{t_k}, \mathcal{Z}^e_{t_{k+1}}, a^e_{t_k}\big)\Delta t^2.\end{aligned}$$

Then using the Young inequality $(a+b)^2 \le (1+\gamma h)a^2 + (1+\frac{1}{\gamma h})b^2$ and the Cauchy–Schwarz inequality, we get for some constant C independent of e and k,

$$\begin{aligned}(\mathcal{Y}^e_{t_k} - \widetilde{\mathcal{Y}}^e_{t_k})^2 &\le (1+\gamma\Delta t)(1+C\Delta t)\mathbb{E}^n_{t_k}\big[(\mathcal{Y}^e_{t_{k+1}} - \widetilde{\mathcal{Y}}^e_{t_{k+1}})^2\big] \\ &\quad + Cf^2\big(t_k, \widehat{M}^e_\cdot, \mathcal{Y}^e_{t_k}, \mathcal{Z}^e_{t_{k+1}}, a^e_{t_k}\big)\Delta t^2.\end{aligned}$$

Taking expectations en each side and using the Lipschitz property of f, we get

$$\begin{aligned}\mathbb{E}^n_0\big[(\mathcal{Y}^e_{t_k} - \widetilde{\mathcal{Y}}^e_{t_k})^2\big] &\le \mathbb{E}^n_0\big[(1+C\Delta t)(\mathcal{Y}^e_{t_{k+1}} - \widetilde{\mathcal{Y}}^e_{t_{k+1}})^2\big] \\ &\quad + C\Delta t^2\mathbb{E}^n_0\big[|\widehat{M}^e|^2 + |\mathcal{Y}^e|^2 + |\mathcal{Z}^e|^2\big].\end{aligned}$$

Finally, it is enough to conclude using the Gronwall lemma together with the estimates given by Lemma 2.6. □

For every $n \ge 1$, we can reformulate the problem (2.10) for Y^n_0 and (3.2) for $\widetilde{Y}^n_0$ as a numerical scheme defined on

$$\Lambda^n := \bigcup_{0\le k\le n} \{t_k\} \times \mathbb{R}^{d\times(k+1)}.$$

For every $n \ge 1$, $(t_k, \mathbf{x}) \in \Lambda^n$ and $a \in A$, we define $M^{t_k,\mathbf{x},a} \in \mathbb{R}^{d\times(k+2)}$ by

$$\begin{cases} M^{t_k,\mathbf{x},a}_{t_i} := \mathbf{x}_i, & \text{for every } i \le k, \\ M^{t_k,\mathbf{x},a}_{t_{k+1}} := M^{t_k,\mathbf{x},a}_{t_k} + H^n_{k+1}(a, U_{k+1}). \end{cases}$$

We then define $u^n : \Lambda_n \longrightarrow \mathbb{R}$ and $\tilde{u}^n : \Lambda_n \longrightarrow \mathbb{R}$ by the following backward iterations. The terminal conditions are given by

$$u^n(t_n, \mathbf{x}) := \tilde{u}^n(t_n, \mathbf{x}) := \xi(\hat{\mathbf{x}}) \qquad \forall \mathbf{x} \in \mathbb{R}^{d\times(n+1)},$$

and the backward iteration for u^n and $\tilde{u}^n$ are given by, for all $\mathbf{x} \in \mathbb{R}^{d\times(k+1)}$,

$$\begin{cases} u^n(t_k, \mathbf{x}) = \sup_{a\in A} u^n_a(t_k, \mathbf{x}), \\ u^n_a(t_k, \mathbf{x}) = \mathbb{E}\big[u\big(t_{k+1}, M^{t_k,\mathbf{x},a}\big)\big] - f\big(t_k, \hat{\mathbf{x}}, u^n_a(t_k, \mathbf{x}), Du^n_a(t_k, \mathbf{x}), a\big)\Delta t, \\ Du^n_a(t_k, \mathbf{x}) := \mathbb{E}\Big[\dfrac{u(t_{k+1}, M^{t_k,\mathbf{x},a})a^{-1}\Delta M^{t_k,\mathbf{x},a}_{k+1}}{\Delta t}\Big], \end{cases} \tag{3.3}$$

and

$$(3.4) \qquad \begin{cases} \tilde{u}^n(t_k, \mathbf{x}) = \sup_{a \in A} \big(\tilde{u}^n_a(t_k, \mathbf{x}) - f\big(t_k, \hat{\mathbf{x}}, \tilde{u}^n_a(t_k, \mathbf{x}), D\tilde{u}^n_a(t_k, \mathbf{x}), a\big)\Delta t \big), \\ \tilde{u}^n_a(t_k, \mathbf{x}) := \mathbb{E}\big[\tilde{u}\big(t_{k+1}, M^{t_k, \mathbf{x}, a}\big)\big], \\ D\tilde{u}^n_a(t_k, \mathbf{x}) := \mathbb{E}\bigg[\dfrac{\tilde{u}(t_{k+1}, M^{t_k, \mathbf{x}, a}) a^{-1} \Delta M^{t_k, \mathbf{x}, a}_{k+1}}{\Delta t}\bigg]. \end{cases}$$

We have the following dynamic programming result.

PROPOSITION 3.2. *Let Assumption 2.7 hold true, then*

$$\tilde{u}^n(0, \mathbf{0}) = \widetilde{Y}^n_0 \quad \textit{and} \quad u^n(0, \mathbf{0}) = Y^n_0.$$

PROOF. It is in fact a standard result from the dynamic programming principle; see, for example, Bertsekas and Shreve [2] for a detailed presentation on this subject. We also notice that the arguments are almost the same in Theorem 3.4 of Tan [26] for a similar problem. □

3.2. *Concrete numerical schemes of* 2*BSDE*. By constructing the driving martingales $(M^e)_{e \in E_n}$ as a family of controlled Markov chain, we can also compute the solution of (2.10) using a backward iteration, under some monotonicity conditions. In particular, it can be considered as a numerical scheme for the 2BSDE (2.3). For particular choices of functions $(H^n_k)_{1 \le k \le n, n \ge 1}$, we may obtain some numerical schemes, including a finite difference scheme and a probabilistic scheme.

3.2.1. *Finite difference scheme*. Let us stay in the one-dimensional case $d = 1$ for notational simplicity, where $\Delta x \in \mathbb{R}$ is the parameter of the space discretization. Denote $p_a := a\Delta t / \Delta x^2$, suppose that $p_a \le 1/2$ for all $a \in A$. Clearly, for every $n \ge 1$ and space discretization Δx, we can construct a function $H^n : A \times [0, 1] \longrightarrow \{-\Delta x, 0, \Delta x\}$ such that, for any uniformly distributed random variable U

$$\mathbb{P}^n\big[H^n(a, U) = \Delta x\big] = \mathbb{P}^n\big[H^n(a, U) = -\Delta x\big] = p_a,$$

and $\mathbb{P}^n[H^n(a, U) = 0] = 1 - 2p_a$. Let $H^n_k := H^n$, and denote $\mathbf{x}^{k,\pm} := (x_0, \dots, x_k, x_k \pm \Delta x)$ and $\mathbf{x}^{k,0} = (x_0, \dots, x_k, x_k)$ for every $\mathbf{x} = (x_0, \dots, x_k)$. Then it follows by a direct computation that the numerical iteration in (3.4) turns to be

$$(3.5) \qquad \tilde{u}(t_k, \mathbf{x}) := \tilde{u}\big(t_{k+1}, \mathbf{x}^{k,0}\big) + \sup_{a \in A} \bigg\{ \frac{1}{2} a D^2 \tilde{u} - f(\cdot, \tilde{u}_a, D\tilde{u}, a)(t_k, \mathbf{x}) \bigg\},$$

where $\tilde{u}_a(t_k, \mathbf{x}) = \tilde{u}(t_{k+1}, \mathbf{x}^{k,0}) + \frac{1}{2} a \Delta t D^2 \tilde{u}(t_k, \mathbf{x})$, with

$$D^2 \tilde{u}(t_k, \mathbf{x}) = \frac{\tilde{u}(t_{k+1}, \mathbf{x}^{k,+}) - 2\tilde{u}(t_{k+1}, \mathbf{x}) + \tilde{u}(t_{k+1}, \mathbf{x}^{k,-})}{\Delta x^2}$$

and

$$D\tilde{u}(t_k, \mathbf{x}) = \frac{\tilde{u}(t_{k+1}, \mathbf{x}^{k,+}) - \tilde{u}(t_{k+1}, \mathbf{x}^{k,-})}{2\Delta x}.$$

REMARK 3.3. (i) For the above choice of $(H_k^n)_{1\le k\le n}$, Assumption 2.7 holds true whenever $\Delta x \le L_{f,z}$.

(ii) To ensure that $p_a := a\Delta t/\Delta x^2 \le 1/2$, we should choose $\Delta x \sim \sqrt{\Delta t}$. Moreover, the family of functions $(H_k^n)_{0\le k\le n}$ associated with the finite difference scheme satisfies condition (2.6).

(iii) In the high dimensional case $d > 1$, the construction of finite difference scheme will be harder in general. We refer to Kushner and Dupuis [17] in the case where all $a \in A$ are diagonal dominant, and also to Bonnans, Ottenwaelter and Zidani [5] in general cases.

3.2.2. *Probabilistic scheme*. For parabolic nonlinear PDEs including (2.5), Fahim, Touzi and Warin [12] proposed a probabilistic scheme, which was reinterpreted and generalized in a non-Markovian stochastic control context in Tan [26]. We can easily adapt this probabilistic scheme in our context.

Let $a_0 \in \mathbb{S}_d^+$ be a fixed constant, denote $\sigma_0 = a_0^{1/2}$. Suppose that for all $a \in A$,

$$a \ge a_0 \quad \text{and} \quad 1 - \tfrac{1}{2}(a - a_0)a_0^{-1} \ge 0.$$

For every $n \ge 1$, denote $\rho_n : A \times \mathbb{R}^d \longrightarrow \mathbb{R}$ by

$$\rho_n(a, x) := \frac{1}{(2\pi\Delta t)^{d/2}|\sigma_0|^{1/2}} \exp\left(-\frac{1}{2}\Delta x^{-1} x^T a_0^{-1} x\right)\eta_n(a, x), \tag{3.6}$$

with

$$\eta_n(a, x) := \left(1 - \tfrac{1}{2} a \cdot a_0^{-1} + \tfrac{1}{2}\Delta t^{-1} a \cdot a_0^{-1} x x^T \left(a_0^T\right)^{-1}\right).$$

It is easy to verify that $x \longmapsto \rho_n(a, x)$ is a probability density function for every $a \in A$. Then following Tan [26], we can construct $H^n(a, x)$ which is continuous in a and such that $H^n(a, U)$ is a random variable of density function $\rho_n(a, x)$ whenever $U \sim \mathcal{U}[0, 1]$.

To make Assumption 2.7 hold true, we suppose in addition that f is independent of z (see Remark 2.10). Define then the family of functions $(H_k^n)_{1\le k\le n}$ by $H_k^n = H^n$. We can then rewrite $\tilde{u}_a^n$ in (3.4) by the following: let $\Delta W \sim N(0, \Delta t I_d)$,

$$\begin{aligned}
\tilde{u}_a^n(t_k, \mathbf{x}) &= \mathbb{E}\big[\tilde{u}\big(t_{k+1}, \big(\mathbf{x}, x_k + H^n(a, U)\big)\big)\big] \\
&= \mathbb{E}\big[\tilde{u}\big(t_{k+1}, (\mathbf{x}, x_k + a_0\Delta W)\big)\eta_n(a, a_0\Delta W)\big] \\
&= \mathbb{E}\big[\tilde{u}\big(t_{k+1}, (\mathbf{x}, x_k + a_0\Delta W)\big)\big] \\
&\quad + \frac{1}{2}\Delta t a \cdot \mathbb{E}\Big[\tilde{u}\big(t_{k+1}, (\mathbf{x}, x_k + a_0\Delta W)\big)\big(\sigma_0^T\big)^{-1} \\
&\qquad \times \frac{\Delta W_{k+1}\Delta W_{k+1}^T - \Delta t I_d}{\Delta t^2}\sigma_0^{-1}\Big].
\end{aligned}$$

Therefore, the explicit numerical scheme (3.4) can be rewritten in the following way: in a probability space $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$, let $X^0 := (a_0 W_{t_0}, \dots, a_0 W_{t_n}) \in \mathbb{R}^{d\times(n+1)}$, where W is a standard d-dimensional Brownian motion. Let $\widehat{X}$ denote continuous time process obtained by linear interpolation of the discrete time process X^0. The terminal condition is given by $\widetilde{Y}_{t_n} = \xi(\widehat{X}_\cdot)$, and the backward iteration:

$$\widetilde{Y}_{t_k} := \mathbb{E}_{t_k}[\widetilde{Y}_{t_{k+1}}] + \Delta t G\big(t_k, \widehat{X^0}_\cdot, \mathbb{E}_{t_k}[\widetilde{Y}_{k+1}], \Gamma_{t_k}\big), \tag{3.7}$$

with

$$\Gamma_{t_k} := \mathbb{E}_{t_k}\left[\widetilde{Y}_{t_{k+1}} (\sigma_0^T)^{-1} \frac{\Delta W_{k+1} \Delta W_{k+1}^T - \Delta t I_d}{\Delta t^2} \sigma_0^{-1}\right]$$

and

$$G(t, \mathbf{x}, y, \gamma) := \sup_{a\in U}\left(f\left(t, \mathbf{x}, y + \frac{1}{2} a\cdot\gamma\Delta t, a\right) + \frac{1}{2} a\cdot\gamma\right).$$

Notice that the above scheme is closely related to the scheme proposed by Fahim, Touzi and Warin [12] for nonlinear PDEs.

3.3. *Numerical examples*. We provide here some numerical tests on the schemes proposed in Section 3.2. Let $d = 1$ for simplicity, we shall consider two different equations with the following generators f_1 and f_2:

$$f_1(t, x, y, z, a) := \inf_{r\in K}\{rya\} \qquad \text{for some compact set } K \subset \mathbb{R}, \tag{3.8}$$

$$f_2(t, x, y, z, a) := \tfrac{1}{2}\big((\sqrt{a}z + b/\sqrt{a})^-\big)^2 - zb - \tfrac{1}{2} b^2/a, \tag{3.9}$$

and the terminal condition is given by

$$\xi(\mathbf{x}) := K_1 + \left(\int_0^T \mathbf{x}_t\,dt - K_1\right)^+ - \left(\int_0^T \mathbf{x}_t\,dt - K_2\right)^+, \tag{3.10}$$

for some constant $K_1 \le K_2$.

We would like to point out to the reader that the first example of second-order BSDE with generator (3.8) is motivated by a differential game type of problem

$$\sup_{a\in A}\inf_{r\in K} \mathbb{E}\left[\exp\left(\int_0^T r_s a_s\,ds\right)\xi(X_\cdot^a)\right],$$

while the second example with generator (3.9) is taken from the robust utility maximization problem studied by Matoussi, Possamaï and Zhou [19] (see the generator in their Theorem 4.1, when the set A_a is chosen to be $[0, +\infty)$). We also insist on the fact that the generator f_2 depends on the z variable and is of quadratic growth, so that our general convergence result does not apply in this setting. Nonetheless, as shown by the numerical results below, our numerical schemes still converge in this case, leading us to the natural conjecture that convergence also holds in this more general setting.

Moreover, with the above terminal condition (3.10), by adding the variable $M_\cdot := \int_0^\cdot \mathbf{x}_t\,dt$ in the diffusion system, we can also characterize the solution of the 2BSDE by the following degenerate PDE on $v:(t,,x,m) \in [0,T]\times\mathbb{R}^2 \longrightarrow \mathbb{R}$:

$$\partial_t + x\partial_m v + \sup_{a\in A}\left(\frac{1}{2}a\partial^2_{xx}v + f(t,x,v,\partial_x v,a)\right) = 0, \tag{3.11}$$

with terminal condition $v(T,x,m) := K_1 + (m-K_1)^+ - (m-K_2)^+$.

For each of the two 2BSDEs, we implemented the finite difference scheme given by (3.5) and the probabilistic scheme (3.7). As a comparison, we also implemented PDE (3.11) with a splitting finite difference scheme, that is to split it into two PDEs:

$$\partial_t + x\partial_m v = 0 \quad \text{and} \quad \partial_t + \sup_{a\in A}\left(\frac{1}{2}a\partial^2_{xx}v + f(t,x,v,\partial_x v,a)\right) = 0,$$

and then to solve the two PDEs sequentially with classical finite-difference scheme. Since each equation is one-dimensional, the associated classical finite-difference scheme is bound to be a good benchmark for our schemes. We implemented the numerical schemes on a computer with 2.4 GHz CPU and 4G memory.

In the following two low-dimensional examples, we choose $X_0 = 0.2$, $K = [-1,1]$, $K_1 = -0.2$, $K_2 = 0.2$ and $A = [0.04, 0.09]$, corresponding to a volatility uncertainty in $[0.2, 0.3]$. Using difference time-discretization with time step Δt, the numerical solutions of schemes (3.5) and (3.7) are quite stable and closed to the PDE numerical results w.r.t. the relative error. In Figures 1 and 2 below, we give the numerical solutions with different time discretization. The line PDE-FD denotes the splitting finite-difference method on the PDE (3.11), 2BSDE-FD denotes the finite-difference scheme (3.5) on the 2BSDE, and 2BSDE-Proba refers to the probabilistic scheme (3.7) on the 2BSDE. For the probabilistic scheme, we use a simulation-regression to estimate the conditional expectation arising in the backward iteration (3.7). When $\Delta t = 0.02$, a single computation takes 1.72 seconds for PDE-FD, 1.92 seconds for 2BSDE-FD, and 103.2 seconds for the 2BSDE-Proba method (using 2×10^5 simulations in the simulation-regression method). In this two-dimensional case, it is not surprising that the finite-difference scheme is much less time-consuming comparing to the probabilistic scheme.

4. Proof of the convergence result.

4.1. *Proof of Theorem* 2.3. (i) The wellposedness of the BSDE (2.2) is a already proved in Proposition A.1.

(ii) We fix a filtered probability space $(\Omega^0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}^0)$, where the filtration $\mathbb{F}^0$ satisfies the usual hypotheses. Let $\mathcal{P}_h$ denote the collection of all martingale prob-

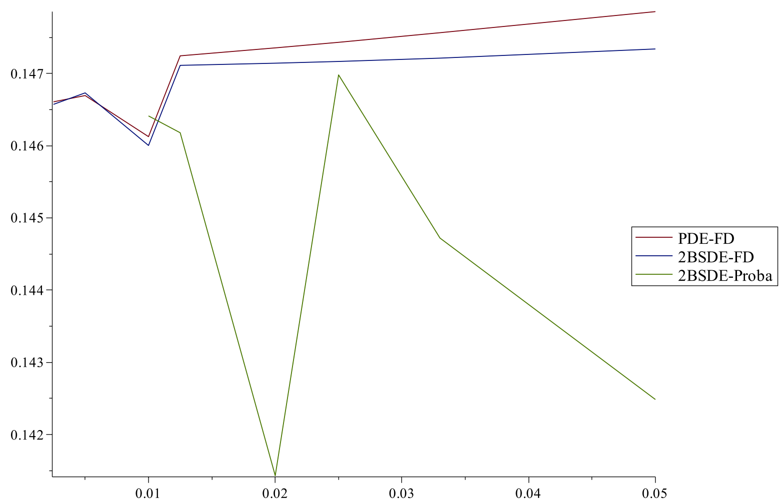

FIG. 1. *The comparison of numerical solutions for* 2*BSDE with generator* (3.8). *The faire value should be very closed to* 0.146, *and the probabilistic scheme seems more volatile comparing to the other schemes*.

ability measures $\mathbb{P} \in \mathcal{P}_W$ such that the density process $\hat{a}$ is piecewise constant, that is to say $\hat{a}_t = \sum_{k=1}^{n} a_{t_k} 1_{t\in[t_k,t_{k+1})}$, $d\mathbb{P} \times dt$-a.e., for some time discretization $0 = t_0 < \cdots < t_n = T$. Let M be a $\mathbb{F}^0$-martingale, whose distribution lies in $\mathcal{P}_W$.

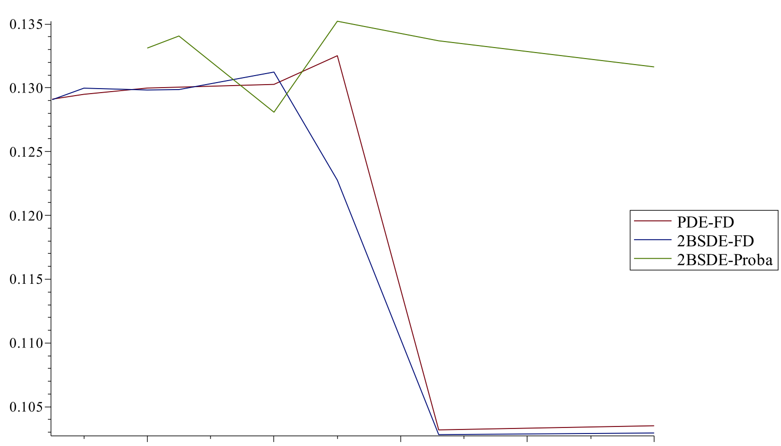

FIG. 2. *The comparison of numerical solutions for* 2*BSDE with generator* (3.9). *The faire value should be closed to* 0.129. *For finite-difference scheme*, *when* Δt *is greater than* 0.025, *we need to use a coarser space-discretization to ensure the monotonicity* (*similar to the classical CFL condition*), *which makes a big difference to the numerical solutions for the case* $\Delta t < 0.25$. *However*, *the convergence as* $\Delta t \to 0$ *is still obvious*.

We can approximate M by a sequence $\widehat{M}^n$ such that $\mathbb{P}^0 \circ (\widehat{M}^n)^{-1} \in \mathcal{P}_h$ and

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T |a_t^n - a_t|^2 \, dt\right] \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}\Big[\sup_{0\le t\le T} |\widehat{M}_t^n - M_t|^2\Big] \underset{n\to+\infty}{\longrightarrow} 0,$$

where $a_t^n := \frac{d\langle \widehat{M}^n\rangle_t}{dt}$ and $a_t := \frac{d\langle M\rangle_t}{dt}$. Then in the spirit of Proposition A.3, we have

$$\sup_{\mathbb{P}\in\mathcal{P}_h} \mathbb{E}[\mathcal{Y}_0^{\mathbb{P}}] = \sup_{\mathbb{P}\in\mathcal{P}_W} \mathbb{E}[\mathcal{Y}_0^{\mathbb{P}}].$$

Further, we claim that for every $\mathbb{P}_h \in \mathcal{P}_h$,

$$\mathbb{E}[\mathcal{Y}_0^{\mathbb{P}_h}] \le Y_0 := \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}}. \tag{4.1}$$

It follows that

$$\sup_{\mathbb{P}\in\mathcal{P}_W} \mathbb{E}[\mathcal{Y}_0^{\mathbb{P}}] = \sup_{\mathbb{P}\in\mathcal{P}_h} \mathbb{E}[\mathcal{Y}_0^{\mathbb{P}}] \le \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}}.$$

By the trivial inequality $Y_0 \le \sup_{\mathbb{P}\in\mathcal{P}_W} \mathbb{E}[\mathcal{Y}_0^{\mathbb{P}}]$, we get (2.4).

(iii) It remains now to prove the claim (4.1). We follow closely the randomization argument in Step 3 of the proof of Proposition 3.5 of Dolinsky, Nutz and Soner [9]. We emphasize that the proof in [9] only uses the fact that the set where the density of the quadratic variation of the canonical process is both convex and compact, which is the case for our set A. We notice that under $\mathbb{P}_h \in \mathcal{P}_h$, the canonical process B is a martingale such that the density of its quadratic variation is piecewise constant. Let us denote it by

$$\alpha_t := \sum_{k=0}^{n-1} \mathbf{1}_{[t_k,t_{k+1})}(t)\alpha(k),$$

where the $\alpha(k)$ are $\mathcal{F}_{t_k}$-measurable. Further, denote $W_t := \int_0^t \alpha_s^{-1/2} \, dB_s$, which is clearly a $\mathbb{P}_h$-Brownian motion. Then by exactly the same arguments as in the step 3 of the proof of Proposition 3.5 of [9], we can consider a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ equipped with a Brownian motion $\tilde{W}$ and i.i.d. uniformly distributed r.v. $(\tilde{U}_k)_{1\le k\le n}$, independent of $\tilde{W}$, and construct, using regular conditional probability measures, random variables $\tilde{\alpha}(k)$ which are $\sigma(\tilde{U}_j, 1 \le j \le k) \vee \sigma(\tilde{W}_s, 0 \le s \le t_k)$-measurable and such that the following equality holds:

the law of $\big(\tilde{W}, (\tilde{\alpha}(i))_{0\le i\le n-1}\big)$ under $\tilde{\mathbb{P}}$ = the law of $\big(W, (\alpha(i))_{0\le i\le n-1}\big)$ under $\mathbb{P}_h$.

Define next the martingale

$$\tilde{M}_t := \int_0^t \left(\sum_{k=0}^{n-1} \mathbf{1}_{[t_k,t_{k+1})}(s)\big(\tilde{\alpha}(k)\big)^{1/2}\right) d\tilde{W}_s, \qquad \tilde{\mathbb{P}}\text{-a.s.}$$

We deduce that $\tilde{\mathbb{P}} \circ \tilde{M}^{-1} = \mathbb{P}_h$. Let us moreover consider the family of conditional probability measures $(\tilde{\mathbb{P}}_c)_{c\in[0,1]^n}$ of $\tilde{\mathbb{P}}$ w.r.t. the sub-σ-field $\sigma(\tilde{U}_k, 1 \le k \le n)$ and define $\mathbb{P}_c := \tilde{\mathbb{P}}_c \circ \tilde{M}^{-1}$. We have that $\mathbb{P}_c \in \mathcal{P}_S$ for every $c \in [0, 1]^n$. It follows that

$$\mathbb{E}[\mathcal{Y}_0^{\mathbb{P}_h}] \le \sup_{c\in[0,1]^n} \mathcal{Y}_0^{\mathbb{P}_c} \le \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}},$$

which justifies the claim (4.1), and we hence complete the proof.

4.2. *Proof of the convergence theorem.* To prove Theorem 2.8, we shall first provide some technical lemmas.

LEMMA 4.1. *Let the functions H_k^n satisfy* (2.6), *then there are some constants $\delta > 0$ and $C > 0$ such that for every $e \in E_n$, $n \ge 1$,*

$$\mathbb{E}\big[\big|M_\cdot^e\big|^{2+\delta}\big] \le C \quad \text{and} \quad \big|a_k^e\big| \le C \qquad \forall 1 \le k \le n, \mathbb{P}^n\text{-a.s.} \tag{4.2}$$

In particular this implies that for every $e \in E_n$,

$$\langle M^e\rangle_t - \langle M^e\rangle_s \le C\big((t-s) + |\Delta_n|\big) I_d, \qquad \mathbb{P}^n\text{-a.s.} \tag{4.3}$$

Moreover, any sequence $(M^{e_n})_{n\ge1}$, with $e_n \in E_n$, $\forall n \ge 1$, is relatively compact and any limit of the sequence lies in $\mathcal{P}_W$.

PROOF. Let $n \ge 1$ and $0 \le s < t \le T$, we can suppose without loss of generality that $t - s > |\Delta_n|$ by (2.6). Then for every $e \in E_n$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{s\le r\le t} \big(M_r^e - M_s^e\big)^{2+\delta}\Big] \\
&\quad \le C\mathbb{E}\big[\big([M^e]_t - [M^e]_s\big)^{1+\delta/2}\big] \\
&\quad \le C\mathbb{E}\Big[\Big(\sum_{s\le t_i\le t} \big|H_i^n(e_i, U_i)\big|^2\Big)^{1+\delta/2}\Big] \\
&\quad = C(t-s)^{1+\delta/2}\mathbb{E}\Big[\Big(\frac{\Delta t}{(t-s)} \sum_{s\le t_i\le t} \big|H_i^n(e_i, U_i)/\sqrt{\Delta t}\big|^2\Big)^{1+\delta/2}\Big] \\
&\quad \le C(t-s)^{1+\delta/2}\mathbb{E}\Big[\frac{\Delta t}{(t-s)} \sum_{s\le t_i\le t} \big|H_i^n(e_i, U_i)/\sqrt{\Delta t}\big|^{2+\delta}\Big] \\
&\quad \le C(t-s)^{1+\delta/2},
\end{aligned}$$

where the first inequality follows from BDG inequality, the second from Jensen's inequality and the last from (2.6). It follows that (4.2) and (4.3) hold true, and hence any sequence $(M^{e_n})_{n\ge1}$ such that $e_n \in E_n$ is relatively compact (see, e.g., Stroock and Varadhan [24]). Finally, let $\mathbb{P}$ be a limit probability measure, it follows by exactly the same argument as in Lemma 3.3 of Dolinsky, Nutz and Soner [9] that $\mathbb{P} \in \mathcal{P}_W$, which completes the proof. $\square$

LEMMA 4.2. *Let u^n be defined in* (3.4) *and* (3.3), *then there is constant C independent of n such that*

$$|u^n(t_k, \mathbf{x}_1) - u^n(t_{k+1}, \mathbf{x}_2^{t_k})| \le C(|\mathbf{x}_1 - \mathbf{x}_2|_k + \sqrt{|\Delta_n|}). \tag{4.4}$$

PROOF. (i) Suppose that $u^n(t_{k+1}, \mathbf{x})$ is Lipschitz in $\mathbf{x}$ with Lipschitz constant L_{k+1}, let $\mathbf{x}^1, \mathbf{x}^2 \in \mathbb{R}^{d\times k}$, then using the same argument as in (2.11), we have for every $a \in A$,

$$\begin{aligned}&|u_a^n(t_k, \mathbf{x}^1) - u_a^n(t_k, \mathbf{x}^2)| \\ &\quad \le (1 - L_{f,y}\Delta t)^{-1}\left(1 + L_{f,z}\frac{\mathbb{E}[|\Delta M|]\Delta t}{|a|} + L_{f,\mathbf{x}}\Delta t\right)|\mathbf{x}^1 - \mathbf{x}^2|.\end{aligned}$$

It follows that for some constant C independent of n,

$$|u^n(t_k, \mathbf{x}^1) - u^n(t_k, \mathbf{x}^2)| \le L_{k+1}(1 + C\Delta t),$$

which implies that u^n is Lipschitz in $\mathbf{x}$ uniformly for all $(t_k)_{0\le k\le n}$ and all $n \ge 1$.

(ii) By the Lipschitz property of u^n, we have immediately

$$Du_a^n(t_k, \mathbf{x}) := \mathbb{E}\left[\frac{u(t_{k+1}, M^{t_k,\mathbf{x},a})a^{-1}\Delta M_{t_{k+1}}^{t_k,\mathbf{x},a}}{\Delta t}\right],$$

is uniformly bounded, which implies that $f(t_k, \mathbf{x}, u_a^n(t_k, \mathbf{x}), Du_a^n(t_k, \mathbf{x}), a)$ is uniformly bounded. It follows by the expression (3.3) that

$$|u^n(t_k, \mathbf{x}) - u^n(t_{k+1}, \mathbf{x}^{t_k})| \le C\sqrt{\Delta t}. \qquad \square$$

PROPOSITION 4.3. *Let Assumption* 2.7 *hold. We have the following properties*:

(i) *For every $n \ge 1$, there is $e_n^* \in E_n$ such that the solution $(\mathcal{Y}^{e_n^*}, \mathcal{Z}^{e_n^*}, \mathcal{N}^{e_n^*})$ of* (2.8) *satisfies $\mathcal{Y}_{t_k}^{e_n^*} = u^n(t_k, M^{e_n^*})$, $\mathbb{P}^n$-a.s.*

(ii) *The sequence $(\mathcal{Y}^{e_n^*})_{n\ge1}$ is tight, and $(\mathcal{Z}^{e_n^*})_{n\ge1}$ is uniformly bounded.*

PROOF. (i) Let $n \ge 1$ be fixed, using the continuity of H_k^n in a and the dominated convergence theorem, $a \longmapsto u_a^n(t_k, \mathbf{x})$ is continuous, where u_a^n is defined by (3.3). Since A is compact, there is always an optimal a for the maximization problem (3.3). It is then enough to use a classical measurable selection theorem to construct the required optimal $e_n^* \in E_n$.

(ii) Notice that since we assumed that Assumption 2.7 holds, we can apply Proposition 3.2. Therefore, by (3.3) and using (4.4), it follows immediately that

$$\mathcal{Z}_{t_k}^{e_n^*} = \mathbb{E}_{t_k}^n\left[\frac{(u^n(t_{k+1}, M_{t_{k+1}}^{e_n^*}) - u^n(t_k, M_{t_k}^{e_n^*}))(a_{t_k}^{e_n^*})^{-1}\Delta M_{t_{k+1}}^{e_n^*}}{\Delta t_{k+1}}\right],$$

is uniformly bounded. Further, using the expression (3.3) with direct computation, we can easily verify that

$$\mathbb{E}^n_{t_k}\big[\big(\Delta \mathcal{Y}^{e^*_n}_{t_k}\big)^2\big] \le C\,\Delta t,$$

for some constant C independent of n, which implies, since $\mathcal{Y}^{e^*_n}$ is a pure jump process that

$$\langle \mathcal{Y}^{e^*_n}\rangle_t \le C t_k, \qquad t_{k-1} \le t \le t_k.$$

Finally, we notice that the deterministic nondecreasing process

$$G^n(s) := C \sum_{k=0}^{n-1} t_{k+1} \mathbf{1}_{[t_k, t_{k+1})}(s),$$

converges weakly to the deterministic process $s \longmapsto Cs$ as $n \longrightarrow \infty$. Then it is enough to apply Theorem 2.3 of Jacod, Mémin and Métivier [14] for the tightness of $(\mathcal{Y}^{e^*_n})_{n\ge 1}$, where their condition $C1$ holds for the nondecreasing process G^n. $\square$

REMARK 4.4. In the context of BSDE, Ma, Protter, San Martin and Torres [18] gave a similar tightness result for their numerical solutions, which is also a key step to prove the convergence of their numerical scheme.

Finally, we are ready to provide the proof of Theorem 2.8 in two steps.

PROOF OF THEOREM 2.8. *Part* (i). Let us consider the BSDE (2.2) under some probability measure $\mathbb{P} \in \mathcal{P}_S$. In this case, we know that $\overline{\mathbb{F}}^{\mathbb{P}} = \overline{\mathbb{F}^{W^{\mathbb{P}}}}^{\mathbb{P}}$, for some $\mathbb{P}$-Brownian motion $W^{\mathbb{P}}$ and thus that thanks to the predictable representation property, we can write for some $\overline{\mathbb{F}^{W}}^{\mathbb{P}}$-predictable process $\tilde{a}$

$$B_t = \int_0^t \tilde{a}_s^{1/2}\, dW^{\mathbb{P}}_s, \qquad \mathbb{P}\text{-a.s.}$$

We may now always approach the process $\tilde{a}$ by a sequence $(\tilde{a}^p)_{p\ge 0}$ of piecewise-constant processes, over a grid $(t^p_k)_{0\le k\le p}$, whose mesh goes to 0, in the sense that

$$\tilde{a}^p_{t_k} \in \overline{\mathcal{F}^{W^{\mathbb{P}}}_{t_k}}^{\mathbb{P}} \qquad \text{for each } 0 \le k \le p \quad \text{and}$$

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T |\tilde{a}_s^{1/2} - (\tilde{a}^p_s)^{1/2}|^2\, ds\right] \underset{p\to+\infty}{\longrightarrow} 0.$$

Next, since there is a priori no reason that the applications $\omega \longmapsto \tilde{a}^p_{t_k}(\omega)$ has any regularity, we further approximate (by classical density arguments) the random variables $(\tilde{a}^p_{t_k})_{0\le k\le p}$ by Lipschitz-continuous functionals $(\tilde{a}^{p,n}_{t_k})_{0\le k\le p}$ such that the following convergence holds true:

$$\mathbb{E}^{\mathbb{P}}\big[|\tilde{a}^p_{t_k}-\tilde{a}^{p,n}_{t_k}|^2\big]\underset{n\to+\infty}{\longrightarrow}0.$$

Let us finally denote by $a^n_\cdot := \tilde{a}^{n,n}_\cdot$. For every $n\ge 1$, let now $(\Omega^n,\mathcal{F}^n,\mathbb{P}^n)$ be a probability space containing n independent random variables $(U_k)_{1\le k\le n}$, and consider the following discrete-time martingale defined exactly as in Section 2, with functions H^n_k satisfying (2.6):

$$M^n_0:=0\quad\text{and}\quad M^n_{t_{k+1}}:=M^n_{t_k}+H^n_k\big(a^n_{t_k}(W^{\mathbb{P},n}_\cdot),U\big),$$

where $W^{\mathbb{P},n}$ is a discretized version of $W^{\mathbb{P}}$ defined by

$$W^{\mathbb{P},n}_0:=0\quad\text{and}\quad W^{\mathbb{P},n}_{t_{k+1}}:=W^{\mathbb{P},n}_{t_k}+\big(a^n_{t_k}\big)^{-1/2}(W^{\mathbb{P},n}_\cdot)H^n_k\big(a^n_{t_k}(W^{\mathbb{P},n}_\cdot),U\big).$$

Consider now the following BSDE under $\mathbb{P}^n$

$$y^n_t=\xi\big(\widehat{M}^n_\cdot\big)-\int_t^T g_n\big(s,\widehat{M}^n_\cdot,y^n_{s-},a^n_sz^n_s,a^n_s\big):d\langle M^n_s\rangle-\int_t^T a^n_sz^n_s.dW^n_s-N^n_T+N^n_t,$$

which is clearly in the same form of BSDE (2.8), and hence $y^n_0\le Y^n_0$.

We know that $W^{\mathbb{P},n}$ converges weakly to $W^{\mathbb{P}}$. Using Skrorohod theorem and changing the probability space under which we are working, it is clear with Lemma 4.1 that we may assume without loss of generality that $W^{\mathbb{P},n}$ actually converges to $W^{\mathbb{P}}$ strongly in $\mathcal{S}^2$ (see also Corollary 14 in Briand, Delyon and Mémin [6] for similar arguments). Moreover, since the filtrations are Brownian filtrations, we know from [6] (see their Proposition 3) that the corresponding filtrations also converge.[1] Then, using the uniform continuity of g in t, we can apply Theorem 12 in [6] to obtain that

$$\lim_{n\to\infty}y^n_0=\mathcal{Y}^{\mathbb{P}}_0.$$

Therefore, we get

$$\liminf_{n\to\infty}Y^n_0\ge\liminf_{n\to\infty}y^n_0=\mathcal{Y}^{\mathbb{P}}_0,$$

which implies the first assertion of Theorem 2.8. $\square$

To prove the second part of Theorem 2.8, we shall consider the weak limit of the triplet $(M^{e^*_n},\mathcal{Y}^{e^*_n},a^{e^*_n})_{n\ge1}$ introduced in Proposition 4.3. Let us first introduce the associated canonical space. For the process $(M^{e^*_n},\mathcal{Y}^{e^*_n})$, it is natural to consider the spaces of all càdlàg paths on $[0,T]$ equipped with Skorokhod topology $C([0,T],\mathbb{R}^d)$ and $D([0,T],\mathbb{R})$ (let us refer to Billinsley [4] for a presentation of

[1] In the sense that, if $(\mathcal{F}^n_t)_{0\le t\le T}$ denotes the natural filtration of W^n and $(\mathcal{F}_t)_{0\le t\le T}$ that of $W^{\mathbb{P}}$, then for every $A\in\mathcal{F}_T$, $\mathbb{E}[\mathbf{1}_A|\mathcal{F}^n_t]$ converges u.c.p. to $\mathbb{E}[\mathbf{1}_A|\mathcal{F}_t]$.

this canonical space). For $(a^{e_n^*})_{n\geq 1}$, we follow Kushner and Dupuis [17] in their numerical analysis to use the canonical space of measure valued processes (see also El Karoui, Huu Nguyen and Jeanblanc-Picqué [10], or El Karoui and Tan [11] for a more detailed presentation). More precisely, since $a^{e_n^*}$ take values in compact set A, we define $\mathbb{M}$ as the space of all measures m on $[0,T]\times A$ such that the marginal distribution of m on $[0,T]$ is the Lebesgue measure. By disintegration, m can be write as $m(dt,da) = m_t(da)\,dt$, where every m_t is a probability measure on A, which can be viewed as measure-valued processes. We then take $\overline{\Omega} := C([0,T],\mathbb{R}^d)\times D([0,T],\mathbb{R})\times\mathbb{M}$ as canonical space, with canonical process $(\overline{M},\overline{\mathcal{Y}},\bar{m})$ and the canonical filtration $\overline{\mathbb{F}} = (\overline{\mathcal{F}}_t)_{0\leq t\leq T}$ generated by the canonical process. For every $\varphi\in C_b^2(\mathbb{R}^d\times\mathbb{R})$, we define a process on $\overline{\Omega}$

$$\mathcal{C}_t^{\varphi}(\overline{M},\bar{m}) := \varphi(\overline{M}_t) - \int_0^t\int_A \frac{1}{2}a : D^2\varphi(\overline{M}_s)\bar{m}_s(da,ds),$$

and another process

$$\mathcal{D}_t(\overline{M},\overline{\mathcal{Y}},\bar{m}) := Y_t + \int_0^t\int_A f(s,\overline{M}_\cdot,\overline{\mathcal{Y}}_s,a)\bar{m}_s(da,ds),$$

as well as

$$\mathcal{D}_t^n(\overline{M},\overline{\mathcal{Y}},\bar{m}) := Y_t + \int_0^t\int_A f_n(s,\overline{M}_\cdot,\overline{\mathcal{Y}}_s,a)\bar{m}_s(da,ds),$$

for every $n\geq 1$. Notice that for every fixed $t>0$, the two random variables $\mathcal{C}_t^{\varphi}$ and $\mathcal{D}_t$ are both bounded continuous in $(\overline{M},\overline{\mathcal{Y}},\bar{m})$.

PROOF OF THEOREM 2.8. *Part* (ii). Let us take the sequence $(e_n^*)_{n\geq 1}$ introduced in Proposition 4.3, we denote $(M^{e_n^*},\mathcal{Y}^{e_n^*},a^{e_n^*})$ by $(M^n,\mathcal{Y}^n,a^n)$ to simplify the presentation. Then

$$\limsup_{n\to\infty}\mathcal{Y}_0^n = \limsup_{n\to\infty} Y_0^n.$$

Denote

$$m^n(dt,da) := \sum_{k=0}^{n-1}\delta_{a_{t_k}^n}(da)\,dt\,1_{t\in[t_k,t_{k+1})}.$$

Let $\overline{\mathbb{P}}^n$ denote the law on $\overline{\Omega}$ induced by $(\widehat{M}^n,\mathcal{Y}^n,m^n)$ in probability space $(\Omega^n,\mathcal{F}^n,\mathbb{P}^n)$, where $\widehat{M}^n$ is the linear interpolation of $(M_{t_k}^n)_{0\leq k\leq n}$. Since $(\widehat{M}^n,\mathcal{Y}^n)_{n\geq 1}$ is tight (by Proposition 4.3 which uses Assumption 2.7) and A is a compact set, then $(\overline{\mathbb{P}}^n)_{n\geq 1}$ is relatively compact. Let $\overline{\mathbb{P}}^{\infty}$ be a limit probability measure, we claim that

$$\mathcal{C}^{\varphi}(\overline{M},\bar{m}) \text{ and } \mathcal{D}(\overline{M},\overline{\mathcal{Y}},\bar{m}) \text{ are both } \overline{\mathbb{F}}\text{-martingales under } \overline{\mathbb{P}}^{\infty}. \tag{4.5}$$

Let $0 \le s < t \le T$ and $\Psi : C([0,T],\mathbb{R}^d) \times D([0,T],\mathbb{R}) \times \mathbb{M} \longrightarrow \mathbb{R}$ be a bounded continuous function which is $\overline{\mathcal{F}}_s$-measurable. Then by the definition of $(M^n, \mathcal{Y}^n)$ in (2.8), it is clear that

$$\mathbb{E}^{\overline{\mathbb{P}}^n}\big[\Psi(\overline{M},\overline{\mathcal{Y}},\bar{m})\big(\mathcal{C}_t^\varphi(\overline{M},\bar{m}) - \mathcal{C}_s^\varphi(\overline{M},\bar{m})\big)\big] = 0$$

and

$$\mathbb{E}^{\overline{\mathbb{P}}^n}\big[\Psi(\overline{M},\overline{\mathcal{Y}},\bar{m})\big(\mathcal{D}_t^n(\overline{M},\overline{\mathcal{Y}},\bar{m}) - \mathcal{D}_s^n(\overline{M},\overline{\mathcal{Y}},\bar{m})\big)\big] = 0.$$

Since the functionals Ψ, $\mathcal{C}_t^\varphi$ and $\mathcal{D}_t$ are all bounded continuous, by taking the limit $n \longrightarrow \infty$, it follows that

$$\mathbb{E}^{\overline{\mathbb{P}}^\infty}\big[\Psi(\overline{M},\overline{\mathcal{Y}},\bar{m})\big(\mathcal{C}_t^\varphi(\overline{M},\bar{m}) - \mathcal{C}_s^\varphi(\overline{M},\bar{m})\big)\big] = 0$$

and

$$\mathbb{E}^{\overline{\mathbb{P}}^\infty}\big[\Psi(\overline{M},\overline{\mathcal{Y}},\bar{m})\big(\mathcal{D}_t(\overline{M},\overline{\mathcal{Y}},\bar{m}) - \mathcal{D}_s(\overline{M},\overline{\mathcal{Y}},\bar{m})\big)\big] = 0,$$

which implies claim (4.5) by the arbitrariness of Ψ and $s \le t$.

It follows that there exists some probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ containing the processes $(M^*, \mathcal{Y}^*, m^*)$ whose distribution is $\overline{\mathbb{P}}^\infty$. Let $\mathbb{F}^* = (\mathcal{F}_t^*)_{0 \le t \le T}$ be the right-limit of the filtration generated by $(M^*, \mathcal{Y}^*, m^*)$, completed under $\mathbb{P}^*$ and let $a_s^* := \int_A a m_s^*(da)$ (notice that a^* also takes values in A, since this set is assumed to be convex). Then M^* is a martingale w.r.t. $\mathbb{F}^*$ with quadratic variation $\int_0^t a_s^*\,ds$ and $\mathcal{D}(M^*, \mathcal{Y}^*, m^*)$ is a martingale w.r.t. $\mathbb{F}^*$ by claim (4.5). Further, by the convexity of f in a, we have

$$\int_A f(s,\mathbf{x},y,a) m_s^*(da) \ge f\big(s,\mathbf{x},y,a_s^*\big).$$

It follows that $\mathcal{Y}_t^* - \int_0^t f(s, M_\cdot^*, \mathcal{Y}_s^*, a_s^*)\,ds$ is a bounded $\mathbb{F}^*$-submartingale.

Next, since this is a bounded submartingale, applying Doob–Meyer decomposition and the orthogonal decomposition for the $\mathbb{F}^*$-martingales gives us the existence of a $\mathbb{F}^*$-predictable process $\mathcal{Z}^*$, a càdlàg $\mathbb{F}^*$-martingale $\mathcal{N}^*$, orthogonal to M^* and a nondecreasing process $\mathcal{K}^*$ such that

$$\mathcal{Y}_t^* = \xi - \int_t^T f\big(s, M_\cdot^*, \mathcal{Y}_s^*, a_s^*\big)\,ds - \int_t^T \mathcal{Z}_s^*\,dM_s^* - \int_t^T d\mathcal{N}_s^* - \int_t^T d\mathcal{K}_s^*, \quad \mathbb{P}^*\text{-a.s.}$$

Consider now $(\widetilde{\mathcal{Y}}^*, \widetilde{\mathcal{Z}}^*, \widetilde{\mathcal{N}}^*)$ the unique solution of the following BSDE under $\mathbb{P}^*$:

$$\widetilde{\mathcal{Y}}_t^* = \xi - \int_t^T f\big(s, M_\cdot^*, \widetilde{\mathcal{Y}}_s^*, a_s^*\big)\,ds - \int_t^T \widetilde{\mathcal{Z}}_s^*\,dM_s^* - \int_t^T d\widetilde{\mathcal{N}}_s^*, \quad \mathbb{P}^*\text{-a.s.} \tag{4.6}$$

We now claim that we necessarily have

$$\mathbb{E}^{\mathbb{P}^*}[\widetilde{\mathcal{Y}}_0^*] \geq \mathbb{E}^{\mathbb{P}^*}[\mathcal{Y}_0^*]. \tag{4.7}$$

This implies that

$$\limsup_{n\to\infty} Y_0^n = \limsup_{n\to\infty} \mathcal{Y}_0^{e_n} = \mathbb{E}^{\mathbb{P}^*}[\mathcal{Y}_0^*] \leq \mathbb{E}^{\mathbb{P}^*}[\widetilde{\mathcal{Y}}_0^*] \leq \sup_{\mathbb{P}\in\mathcal{P}_W} \mathbb{E}^{\mathbb{P}}[\mathcal{Y}_0^{\mathbb{P}}] = \sup_{\mathbb{P}\in\mathcal{P}_S} \mathcal{Y}_0^{\mathbb{P}},$$

which proves the desired property.

It remains now to prove the claim (4.7). It follows from a classical linearization argument, which we give for completeness. Using the fact that f is uniformly Lipschitz in y, we may define bounded $\mathbb{F}^*$-progressively measurable process λ such that, $\mathbb{P}^*$-a.s.

$$\delta\mathcal{Y}_t^* = -\int_t^T \lambda_s \delta\mathcal{Y}_s^*\,ds - \int_t^T \delta\mathcal{Z}_s^*\,dM_s^* - \int_t^T d(\delta\mathcal{N}_s^*) + \int_t^T d\mathcal{K}_s^*,$$

where

$$\delta\mathcal{Y}_t^* := \widetilde{\mathcal{Y}}_t^* - \mathcal{Y}_t^*, \qquad \delta\mathcal{Z}_t^* := \widetilde{\mathcal{Z}}_t^* - \mathcal{Z}_t^*, \qquad \delta\mathcal{N}_t^* := \widetilde{\mathcal{N}}_t^* - \mathcal{N}_t^*.$$

Then denote $\Lambda_t := \exp(-\int_0^t \lambda_s\,ds)$. Applying Itô's formula to $\Lambda_t\delta\mathcal{Y}_t^*$ and remembering that M^* is orthogonal to $\mathcal{N}^*$ and $\widetilde{\mathcal{N}}^*$, we deduce that

$$\mathbb{E}^{\mathbb{P}^*}[\delta\mathcal{Y}_0] = \mathbb{E}^{\mathbb{P}^*}\left[\int_0^T \Lambda_s\,d\mathcal{K}_s\right] \geq 0,$$

which completes the proof. $\square$

APPENDIX

We provide here some classical results on BSDEs which are used in the paper. Let us start by stating a general wellposedness result for BSDEs in an abstract setting, which will encompass all the cases considered in this paper.

PROPOSITION A.1. *Let $(\Omega_0, \mathcal{F}, \mathbb{P})$ be a complete probability space carrying a square integrable continuous martingale M, adapted to a complete and right-continuous filtration $\mathbb{F}^0 := (\mathcal{F}_t^0)_{0\leq t\leq T}$ and a sequence of square-integrable càdlàg martingales M^n adapted to some filtration $\mathbb{F}^n := (\mathcal{F}_t^n)_{0\leq t\leq T}$ which are complete and right-continuous for each n. Let f_0 and f_n be functions from $[0,T]\times\Omega_0\times\mathbb{R}\times\mathbb{R}^d$ to $\mathbb{R}$ and assume furthermore that:*

(i) *$\langle M\rangle$ is absolutely continuous with respect to the Lebesgue measure, with a density $(a_s)_{0\leq s\leq T}$ taking values in A.*

(ii) *There exists a deterministic sequence $(a_n)_{n\geq 0}$ converging to 0 such that*

$$\langle M^n\rangle_t - \langle M^n\rangle_s \leq C(t-s+a_n)I_d, \qquad 0\leq s\leq t\leq T, \mathbb{P}\text{-a.s.},$$

for some $C>0$.

(iii) *For each* (y,z), $f_0(\cdot, M_\cdot, y, z)$ [*resp.*, $f_n(\cdot, M^n_\cdot, y, z)$] *is progressively measurable with respect to* $\mathbb{F}^0$ (*resp.*, $\mathbb{F}^n$).

(iv) *There is a constant* $\mu > 0$ *such that for each* $n \geq 0$ *and each* (t, y, y', z, z')

$$\left|f_0(t, M_\cdot, y, z) - f_0(t, M_\cdot, y', z')\right| \leq \mu\left(|y - y'| + |z - z'|\right),$$
$$\left|f_n(t, M^n_\cdot, y, z) - f_n(t, M^n_\cdot, y', z')\right| \leq \mu\left(|y - y'| + |z - z'|\right).$$

(v) *For all* (y,z), f_0 *and* f_n *are continuous in* t.

Then, for n large enough, the following BSDEs under $\mathbb{P}$

$$\mathcal{Y}_t = \xi - \int_t^T f_0(s, M_\cdot, \mathcal{Y}_s, \mathcal{Z}_s)\, d\langle M\rangle_t - \int_t^T \mathcal{Z}_s \cdot dM_s - \mathcal{N}_T + \mathcal{N}_t, \tag{A.1}$$

$$\mathcal{Y}^n_t = \xi - \int_t^T f_n(s, M^n_\cdot, \mathcal{Y}^n_{s-}, \mathcal{Z}^n_s)\, d\langle M^n\rangle_t - \int_t^T \mathcal{Z}^n_s \cdot dM_s - \mathcal{N}^n_T + \mathcal{N}^n_t, \tag{A.2}$$

where $\mathcal{N}$ (*resp.*, $\mathcal{N}^n$) *is a càdlàg* $\mathbb{F}^0$*-martingale* (*resp.*, $\mathbb{F}^n$*-martingale*) *orthogonal to* M (*resp.*, M^n), *have a unique solution such that*

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{0\leq t\leq T} |\mathcal{Y}_t|^2 + \int_0^T |a_s^{1/2}\mathcal{Z}_s|^2\, ds + \langle \mathcal{N}\rangle_T\right] \leq C,$$
$$\mathbb{E}^{\mathbb{P}}\left[\sup_{0\leq t\leq T} |\mathcal{Y}^n_t|^2 + \int_0^T \mathcal{Z}^n_s (\mathcal{Z}^n_s)^T : d\langle M^n\rangle_s + \langle \mathcal{N}^n\rangle_T\right] \leq C$$

for some constant $C > 0$ *independent of* n.

PROOF. This is actually a direct consequence of the proof of existence via fixed point arguments in [6]. Indeed, the assumptions above imply directly that their assumptions H1, H2 and H3 hold, with the exception that we do not assume that M^n converges to M and that our martingale M can be written as

$$M_t = \int_0^t a_s^{1/2}\, dW_s,$$

where W is $(\mathbb{P}, \mathbb{F}^0)$-Brownian motion.

However, by looking carefully at their proofs of Theorem 9 and Corollary 10, it is easy to see that they can be carried out with the exact same arguments in our setting to obtain the desired results for the BSDE (A.2) for n large enough. Moreover, since the martingale M satisfies their assumption (H1)(ii) with a constant $C := \sup_{a\in A} |a|$ and a deterministic sequence $a_n = C|\Delta_n|$, we can once again follow their proof of existence to obtain easily that existence, uniqueness and the desired estimates also hold for (A.1). □

We will now provide a particular robustness result for BSDEs. We go back to the canonical space $(\Omega, \mathcal{F}_T)$ and fix a measure $\mathbb{P} \in \mathcal{P}_W$. We let W be a $\overline{\mathbb{F}^+}^{\mathbb{P}}$-Brownian motion under $\mathbb{P}$, $(a_s)_{0\le s\le T}$ be a $\mathbb{F}$-progressively measurable process and $(a^n)_{n\ge 0}$ a sequence of $\mathbb{F}$-progressively measurable processes such that

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T |a_s^n - a_s|^2\, ds\right] \underset{n\to+\infty}{\longrightarrow} 0. \tag{A.3}$$

We next define the following $\overline{\mathbb{F}^+}^{\mathbb{P}}$-martingales under $\mathbb{P}$:

$$M_t := \int_0^t a_s^{1/2}\, dW_s \quad \text{and}$$
$$\widehat{M}_t^n := \int_0^t (a_s^n)^{1/2}\, dW_s, \qquad \mathbb{P}\text{-a.s.}$$

Notice that we than have immediately that $\widehat{M}^n$ converges to M in the sense that

$$\mathbb{E}^{\mathbb{P}}\Big[\sup_{0\le t\le T} |M_t - \widehat{M}_t^n|^2\Big] \underset{n\to+\infty}{\longrightarrow} 0. \tag{A.4}$$

We would like to approximate the BSDE

$$\mathcal{Y}_t = \xi - \int_t^T f(s, M_\cdot, \mathcal{Y}_s, \mathcal{Z}_s, a_s)\, ds - \int_t^T (a_s)^{1/2} \mathcal{Z}_s \cdot dW_s - \mathcal{N}_T + \mathcal{N}_t, \qquad \mathbb{P}\text{-a.s.} \tag{A.5}$$

by the following one for $n \ge 0$:

$$\begin{aligned} \widehat{Y}_t^n = \xi_n &- \int_t^T f\big(s, \widehat{M}_\cdot^n, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s^n\big)\, ds \\ &- \int_t^T (a_s^n)^{1/2} \widehat{Z}_s^n \cdot dW_s - \widehat{N}_T^n + \widehat{N}_t^n, \qquad \mathbb{P}\text{-a.s.,} \end{aligned} \tag{A.6}$$

for some $\mathcal{F}_T$-measurable random variable ξ_n converging to ξ in $L^2(\mathbb{P})$.

REMARK A.2. Notice that existence and uniqueness for these BSDEs are once again guaranteed by Proposition A.1.

We have the following result, which can be proved using classical stability arguments for BSDEs. We nonetheless give the proof for completeness.

PROPOSITION A.3. *Let Assumptions 2.2 hold. Then we have*

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{0\le t\le T} |\widehat{Y}_t^n - \mathcal{Y}_t|^2 + \int_0^T |\widehat{Z}_t^n - \mathcal{Z}_t|^2\, ds + \langle \widehat{N}^n - N\rangle_T\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

PROOF. Let us apply Itô's formula to $e^{\eta t}(\widehat{Y}_t^n - \mathcal{Y}_t)^2$, for some constant η to be fixed later. We obtain, using the fact that $\widehat{N}^n$ and N are orthogonal to W

$$
\begin{aligned}
& e^{\eta t}(\widehat{Y}_t^n - \mathcal{Y}_t)^2 + \int_t^T e^{\eta s}\big|(a_s^n)^{1/2}\widehat{Z}_s^n - a_s^{1/2}\mathcal{Z}_s\big|^2\,ds + \int_t^T e^{\eta s}\,d\langle \widehat{N}^n - N\rangle_s \\
&\quad \le e^{\eta T}|\xi_n - \xi|^2 - 2\int_t^T e^{\eta s}(\widehat{Y}_s^n - \mathcal{Y}_s) \\
&\qquad\qquad \times \big(f(s, \widehat{M}_\cdot^n, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s^n) - f(s, M_\cdot, \mathcal{Y}_s, \mathcal{Z}_s, a_s)\big)\,ds \qquad (\text{A.7})\\
&\qquad - \eta\int_t^T e^{\eta s}|\widehat{Y}_s^n - \mathcal{Y}_s|^2\,ds - 2\int_t^T (\widehat{Y}_s^n - \mathcal{Y}_s)\big((a_s^n)^{1/2}\widehat{Z}_s^n - a_s^{1/2}\mathcal{Z}_s\big)\cdot dW_s \\
&\qquad - \int_t^T e^{\eta s}(\widehat{Y}_{s^-}^n - \mathcal{Y}_{s^-})\,d(\widehat{N}_s^n - N_s).
\end{aligned}
$$

Next, using the uniform continuity of f in u and its Lipschitz continuity in $(\mathbf{x}, y, z)$, we have for some modulus of continuity ρ and using the trivial inequality $ab \le \varepsilon a^2 + \frac{1}{\varepsilon}b^2$ for any $\varepsilon > 0$

$$
\begin{aligned}
& \left|\int_t^T e^{\eta s}(\widehat{Y}_s^n - \mathcal{Y}_s)\big(f(s, \widehat{M}_\cdot^n, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s^n) - f(s, M_\cdot, \mathcal{Y}_s, \mathcal{Z}_s, a_s)\big)\,ds\right| \\
&\quad \le \int_t^T e^{\eta s}|\widehat{Y}_s^n - \mathcal{Y}_s|\big|f(s, \widehat{M}_\cdot^n, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s^n) - f(s, M_\cdot, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s)\big|\,ds \\
&\qquad + \int_t^T e^{\eta s}|\widehat{Y}_s^n - \mathcal{Y}_s|\big|f(s, M_\cdot, \widehat{Y}_s^n, \widehat{Z}_s^n, a_s) - f(s, M_\cdot, \mathcal{Y}_s, \mathcal{Z}_s, a_s)\big|\,ds \qquad (\text{A.8})\\
&\quad \le C\left(\|\widehat{M}^n - M\|_T^2 + \int_t^T \rho^2(a_s^n - a_s)\,ds\right) \\
&\qquad + \left(C + \frac{1}{\varepsilon}\right)\int_t^T e^{\eta s}|\widehat{Y}_s^n - \mathcal{Y}_s|^2\,ds \\
&\qquad + \varepsilon\int_t^T e^{\eta s}|\widehat{Z}_s^n - \mathcal{Z}_s|^2\,ds.
\end{aligned}
$$

Using the fact that a^n and a are uniformly bounded, if we take the expectation in (A.7) and use the estimate (A.8), we obtain by choosing η large enough and $\varepsilon < 1$

$$
\begin{aligned}
& \mathbb{E}^{\mathbb{P}}\left[|\widehat{Y}_t^n - \mathcal{Y}_t|^2 + \int_0^T |\widehat{Z}_s^n - \mathcal{Z}_s|^2\,ds + \langle \widehat{N}^n - N\rangle_T\right] \\
&\qquad \le C\mathbb{E}^{\mathbb{P}}\left[|\xi_n - \xi|^2 + \|\widehat{M}^n - M\|_T^2 + \int_0^T \rho^2(a_s^n - a_s)\,ds\right].
\end{aligned}
$$

By the dominated convergence theorem and using the fact that ξ_n converges to ξ and $\widehat{M}^n$ to M, the right-hand side above goes to 0. Now the proof can be finished by taking the supremum in t in (A.7) and using the BDG inequality. Since this part is classical, we refrain from writing its proof. $\square$

Acknowledgements. We are grateful to Nizar Touzi, Jianfeng Zhang and Chao Zhou for fruitful discussions. We also would like to thank an anonymous referee and an Associate Editor, whose advices helped to improve an earlier version of the paper.

REFERENCES

[1] BARLES, G. and SOUGANIDIS, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptot. Anal.* **4** 271–283. MR1115933

[2] BERTSEKAS, D. P. and SHREVE, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case. Mathematics in Science and Engineering* **139**. Academic Press, New York. MR0511544

[3] BICHTELER, K. (1981). Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.* **9** 49–89. MR0606798

[4] BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York. MR0534323

[5] BONNANS, J. F., OTTENWAELTER, É. and ZIDANI, H. (2004). A fast algorithm for the two dimensional HJB equation of stochastic control. *M2AN Math. Model. Numer. Anal.* **38** 723–735. MR2087732

[6] BRIAND, P., DELYON, B. and MÉMIN, J. (2002). On the robustness of backward stochastic differential equations. *Stochastic Process. Appl.* **97** 229–253. MR1875334

[7] DEBRABANT, K. and JAKOBSEN, E. R. (2013). Semi-Lagrangian schemes for linear and fully non-linear diffusion equations. *Math. Comp.* **82** 1433–1462. MR3042570

[8] DOLINSKY, Y. (2010). Applications of weak convergence for hedging of game options. *Ann. Appl. Probab.* **20** 1891–1906. MR2724424

[9] DOLINSKY, Y., NUTZ, M. and SONER, H. M. (2012). Weak approximation of G-expectations. *Stochastic Process. Appl.* **122** 664–675. MR2868935

[10] EL KAROUI, N., HUU NGUYEN, D. and JEANBLANC-PICQUÉ, M. (1987). Compactification methods in the control of degenerate diffusions: Existence of an optimal control. *Stochastics* **20** 169–219. MR0878312

[11] EL KAROUI, N. and TAN, X. (2013). Capacities, measurable selection and dynamic programming, part II: Application in stochastic control problems. Preprint.

[12] FAHIM, A., TOUZI, N. and WARIN, X. (2010). A probabilistic numerical method for fully nonlinear parabolic PDEs. *Ann. Appl. Probab.* **21** 1322–1364.

[13] GUO, W., ZHANG, J. and ZHUO, J. (2013). A monotone scheme for high dimensional fully nonlinear PDEs. Preprint. Available at arXiv:1212.0466.

[14] JACOD, J., MÉMIN, J. and MÉTIVIER, M. (1983). On tightness and stopping times. *Stochastic Process. Appl.* **14** 109–146. MR0679668

[15] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften* **288**. Springer, Berlin. MR0959133

[16] KARANDIKAR, R. L. (1995). On pathwise stochastic integration. *Stochastic Process. Appl.* **57** 11–18. MR1327950

[17] KUSHNER, H. J. and DUPUIS, P. G. (1992). *Numerical Methods for Stochastic Control Problems in Continuous Time. Applications of Mathematics (New York)* **24**. Springer, New York. MR1217486

[18] MA, J., PROTTER, P., SAN MARTIN, J. and TORRES, S. (2002). Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* **12** 302–316. MR1890066

[19] MATOUSSI, A., POSSAMAÏ, D. and ZHOU, C. (2014). Robust utility maximization in nondominated models with 2BSDE: The uncertain volatility model. *Math. Finance*. To appear.

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

[21] PENG, S. (2007). G-expectation, G-Brownian motion and related stochastic calculus of Itô type. In *Stochastic Analysis and Applications*. *Abel Symp*. **2** 541–567. Springer, Berlin. MR2397805

[22] SONER, H. M., TOUZI, N. and ZHANG, J. (2011). Quasi-sure stochastic analysis through aggregation. *Electron*. *J*. *Probab*. **16** 1844–1879. MR2842089

[23] SONER, H. M., TOUZI, N. and ZHANG, J. (2012). Wellposedness of second order backward SDEs. *Probab. Theory Related Fields* **153** 149–190. MR2925572

[24] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften* **233**. Springer, Berlin. MR0532498

[25] TAN, X. (2013). A splitting method for fully nonlinear degenerate parabolic PDEs. *Electron*. *J*. *Probab*. **18** 1–24. MR3035743

[26] TAN, X. (2014). Discrete-time probabilistic approximation of path-dependent stochastic control problems. *Ann*. *Appl*. *Probab*. **24** 1803–1834. MR3226164

[27] ZHANG, J. (2004). A numerical scheme for BSDEs. *Ann*. *Appl*. *Probab*. **14** 459–488. MR2023027

CEREMADE
UNIVERSITE PARIS-DAUPHINE
PLACE DU MARECHAL DE LATTRE DE TASSIGNY
75775, PARIS CEDEX 16
FRANCE
E-MAIL: possamai@ceremade.dauphine.fr
tan@ceremade.dauphine.fr