

# A numerical algorithm for a class of BSDE via branching process

Pierre Henry-Labordere, Xiaolu Tan, Nizar Touzi

► **To cite this version:**

Pierre Henry-Labordere, Xiaolu Tan, Nizar Touzi. A numerical algorithm for a class of BSDE via branching process. 2013. <hal-00817180>

**HAL Id: hal-00817180**

**<https://hal.inria.fr/hal-00817180>**

Submitted on 24 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A numerical algorithm for a class of BSDE via branching process <sup>\*</sup>

Pierre Henry-Labordère<sup>†</sup>    Xiaolu Tan<sup>‡</sup>    Nizar Touzi<sup>§</sup>

February 4, 2013

## Abstract

We generalize the algorithm for semi-linear parabolic PDEs in Henry-Labordère [11] to the non-Markovian case for a class of Backward SDEs (BSDEs). By simulating the branching process, the algorithm does not need any backward regression. To prove that the numerical algorithm converges to the solution of BSDEs, we use the notion of viscosity solution of path dependent PDEs introduced by Ekren, Keller, Touzi and Zhang [5] and extended in Ekren, Touzi and Zhang [6, 7].

**Key words.** Numerical algorithm, BSDEs, branching process, viscosity solution, path dependent PDEs.

## 1 Introduction

Initially proposed by Pardoux and Peng [15], the theory of Backward Stochastic Differential Equation (BSDE) has been largely developed and has many applications in control theory, finance etc. In particular, BSDEs can be seen as providing a nonlinear Feynman-Kac formula for semi-linear parabolic PDEs in the Markovian case, i.e. the solution of a Markovian type BSDE can be given as the viscosity solution of a semi-linear PDE. We also remark that this connection has been extended recently to the non-Markovian case by Ekren et al. [5] with the notion of viscosity solution of path dependent PDEs (PPDEs).

Numerical methods for BSDE have also been largely investigated since then. The classical numerical schemes for BSDEs are usually given as a backward iteration, where every step consists in estimating the conditional expectations, see e.g. Bouchard and

---

<sup>\*</sup>The authors are grateful to Emmanuel Gobet for fruitful discussions.

<sup>†</sup>Société Générale, Global Market Quantitative Research, pierre.henry-labordere@sgcib.com

<sup>‡</sup>CEREMADE, University of Paris-Dauphine, xiaolu.tan@gmail.com

<sup>§</sup>Ecole Polytechnique Paris, Centre de Mathématiques Appliquées, nizar.touzi@polytechnique.edu. Research supported by the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, and the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

Touzi [2], Zhang [20]. Generally, we use the regression method to compute the conditional expectations, which is quite costly in practice and suffers from the curse of dimensionality.

A new numerical algorithm for a class of semi-linear PDEs was proposed recently by Henry-Labordère [11], using an extension of branching process. It is a classical result that the branching diffusion process gives a probabilistic representation of the so-called KPP (Kolmogorov-Petrovskii-Piskunov) semi-linear PDE (see e.g. Watanabe [19], McKean [14]):

$$\partial_t u(t, x) + \frac{1}{2} D^2 u(t, x) + \beta \left( \sum_{k=0}^{n_0} a_k u^k(t, x) - u(t, x) \right) = 0 \quad (1.1)$$

with a terminal condition  $u(T, x) = \psi(x)$ , where  $D^2$  is the Laplacian on  $\mathbb{R}^d$ , and  $(a_k)_{0 \leq k \leq n_0}$  is a probability mass sequence, i.e.  $a_k \geq 0$  and  $\sum_{k=0}^{n_0} a_k = 1$ . The above semi-linear PDE (1.1) characterizes a branching Brownian motion, where every particle in the system dies in an exponential time of parameter  $\beta$ , and creates  $k$  i.i.d. descendants with probability  $a_k$ . More precisely, let  $N_T$  denote the number of particles alive at time  $T$ , and  $(Z_T^i)_{i=1, \dots, N_T}$  denote the position of each particle, then up to integrability, the function

$$v(t, x) := \mathbb{E}_{t, x} [\Pi_{i=1}^{N_T} \psi(Z_T^i)]$$

solves the above equation (1.1), where the subscript  $t, x$  means that the system is started at time  $t$  with one particle at position  $x$ . This connection is also extended for a larger class of nonlinearity, typically  $u^\alpha$ ,  $\alpha \in [0, 2]$ , with the superdiffusion, for which we refer to Dynkin [4] and Etheridge [9]. Moreover, this representation allows to solve numerically the PDE (1.1) by simulating the corresponding branching process.

When the coefficients  $a_k$  are arbitrary in equation (1.1) and the Laplacian  $D^2$  is replaced by an Itô operator  $\mathcal{L}_0$  of the form

$$\mathcal{L}_0 u(t, x) := \mu(t, x) \cdot Du(t, x) + \frac{1}{2} \sigma \sigma^T(t, x) : D^2 u(t, x),$$

Henry-Labordère's [11] proposed to simulate a branching diffusion process with a probability mass sequence  $(p_k)_{k=0, \dots, M}$ , and by counting the weight  $\frac{a_k}{p_k}$ , he obtained a so-called "marked" branching diffusion method. A sufficient condition for the convergence of the algorithm is provided in [11]. In particular, the algorithm does not need to use the regression method, which is one of the main advantages comparing to the BSDE method.

For PDEs of the form (1.1), Rasulov, Raimov and Mascagni [21] introduced also a Monte-Carlo method using branching processes. Their method depends essentially on the representation of its solution by the fundamental solution of the heat equation.

The main objective of this paper is to generalize the algorithm in [11] to the non-Markovian case for a class of decoupled Forward Backward SDEs (FBSDEs) whose generators can be represented as the sum of a power series, which can be formally approximated by polynomials. Although the polynomial generators are only locally

Lipschitz, the solutions may be uniformly bounded under appropriate conditions, and hence they can be considered as standard decoupled FBSDEs with Lipschitz generators.

We shall use the branching process to give a numerical algorithm, where the branching process is constructed by countable number of independent Brownian motions and exponential random variables. To bring back the numerical solution to the BSDE context where there is only one Brownian motion with the Brownian filtration, we use the notion of viscosity solutions of path dependent PDEs introduced by Ekren, Keller, Touzi and Zhang [5] and next extended in Ekren, Touzi and Zhang [6, 7]. Namely, we shall prove that the numerical solution obtained by the branching diffusion is a viscosity solution to a corresponding semilinear PPDE, which admits also a representation by a decoupled FBSDE as illustrated in [5]. Then the numerical solution is the unique solution of the corresponding BSDE by the uniqueness of the solution to PPDEs.

The rest of the paper is organized as follows. In Section 2, we consider a class of decoupled FBSDEs whose generators can be represented as a convergent power series. We then introduce a branching diffusion process, which gives a representation of the solution of the FBSDE with polynomial generator. In particular, such a representation induces a numerical algorithm for the class of FBSDEs using branching process. Then in Section 3, we complete the proof of the regularity property of the value function represented by branching processes. Next, we complete the proof of the main representation theorem in Section 4. For this purpose, we introduce in Section 4.1 a notion of viscosity solutions to a class of semilinear PPDE, where there is no non-linearity on the derivatives of the solution function, following Ekren, Touzi and Zhang [6, 7]. Finally, we illustrate the efficiency of our algorithm by some numerical examples in Section 5.

## 2 A numerical algorithm for a class of BSDEs

In this section, we shall consider a class of decoupled FBSDEs whose generators can be represented as a convergent power series, which can be approximated by polynomials. Then for FBSDEs with polynomial generators, we provide a representation of their solutions by branching diffusion processes. In particular, the representation induces a natural numerical algorithm for the class of FBSDEs by simulating the branching diffusion process.

### 2.1 A class of decoupled FBSDEs

Let  $\Omega^0 := \{\omega \in C([0, T], \mathbb{R}^d) : \omega_0 = \mathbf{0}\}$  be the canonical space of continuous paths with initial value  $\mathbf{0}$ ,  $\mathbb{F}^0$  the canonical filtration and  $\Lambda^0 := [0, T] \times \Omega^0$ . For every  $(t, \omega) \in \Lambda^0$ , denote  $\|\omega\|_t := \sup_{0 \leq s \leq t} |\omega(s)|$ . Then the canonical process  $B(\omega) = \{B_t(\omega) := \omega_t, 0 \leq t \leq T\}$  for all  $\omega \in \Omega^0$ , defines a Brownian motion under the Wiener measure  $\mathbb{P}_0$ .

Let  $\mu : \Lambda^0 \rightarrow \mathbb{R}^d$  and  $\sigma : \Lambda^0 \rightarrow \mathbb{S}^d$  be  $\mathbb{F}^0$ -progressively measurable processes.

Suppose further that for every  $0 \leq t \leq t' \leq T$  and  $\omega, \omega' \in \Omega^0$ ,

$$|\mu(t, \omega) - \mu(t', \omega')| + |\sigma(t, \omega) - \sigma(t', \omega')| \leq C(\sqrt{|t - t'|} + \|\omega_{t \wedge \cdot} - \omega'_{t' \wedge \cdot}\|_T) \quad (2.1)$$

for some constant  $C > 0$ , and  $\sigma\sigma^T(t, \omega) \geq c_0 I_d$  for some constant  $c_0 > 0$ . We denote, for every  $(t, \mathbf{x}) \in \Lambda^0$ , by  ${}^{t, \mathbf{x}}X$  the solution of the following SDE under  $\mathbb{P}_0$ :

$$X_s = \mathbf{x}_s, \quad \forall s \leq t \quad \text{and} \quad X_s = \mathbf{x}_t + \int_t^s \mu(r, X.)dr + \int_t^s \sigma(r, X.)dB_r, \quad \forall s > t. \quad (2.2)$$

For later uses, we provide an estimate on the SDE (2.2).

**Lemma 2.1.** *There is a constant  $C$  such that for every  $t \in [0, T]$  and  $(t_1, \mathbf{x}_1), (t_2, \mathbf{x}_2) \in [t, T] \times \Omega^0$ ,*

$$\mathbb{E}^{\mathbb{P}_0} \left[ \left| {}^{t, \mathbf{x}_1}X_{t_1} - {}^{t, \mathbf{x}_2}X_{t_2} \right|^2 \right] \leq C(1 + \|\mathbf{x}_1\|_t^2 + \|\mathbf{x}_2\|_t^2)(|t_1 - t_2| + \|\mathbf{x}_1 - \mathbf{x}_2\|_t^2).$$

**Proof.** It follows by the same arguments as in Lemma 2 and Theorem 37 in Chapter V of Protter [16].  $\square$

Suppose that  $\psi : \Omega^0 \rightarrow \mathbb{R}$  is a non-zero, bounded Lipschitz continuous function, and  $F : (t, \mathbf{x}, y) \in \Lambda^0 \times \mathbb{R} \rightarrow \mathbb{R}$  is a function Lipschitz in  $y$  such that for every  $y$ ,  $F(\cdot, y)$  defined on  $\Lambda^0$  is  $\mathbb{F}^0$ -progressive. We consider the following BSDE:

$$Y_t = \psi({}^{0,0}X.) + \int_t^T F(s, {}^{0,0}X., Y_s)ds - \int_t^T Z_s dW_s, \quad \mathbb{P}_0 - a.s. \quad (2.3)$$

where the generator  $F$  has the following power series representation in  $y$ , locally in  $(t, \mathbf{x})$ :

$$F(t, \mathbf{x}, y) := \beta \left( \sum_{k=0}^{\infty} a_k(t, \mathbf{x}) y^k - y \right), \quad (t, \mathbf{x}) \in \Lambda^0, \quad (2.4)$$

for some constant  $\beta > 0$ , and some sequence  $(a_k)_{k \geq 0}$  of bounded scalar  $\mathbb{F}^0$ -progressive functions defined  $\Lambda^0$ . We also assume that every  $a_k$  is uniformly  $1/2$ -Hölder-continuous in  $t$  and Lipschitz-continuous in  $\omega$ .

Denoting by  $|\cdot|_0$  the  $L^\infty(\Lambda^0)$ -norm  $|a_k|_0$ , we now formulate conditions on the power series

$$\ell_0(s) := \sum_{k \geq 0} |a_k|_0 s^k \quad \text{and} \quad \ell(s) := \beta[|\psi|_0^{-1} \ell_0(s|\psi|_0) - s] \quad s \geq 0, \quad (2.5)$$

so as to guarantee the existence and uniqueness of a solution of the backward SDE (2.3).

**Assumption 2.2.** (i) *The power series  $\ell_0$  has a radius of convergence  $0 < R \leq \infty$ , i.e.  $\ell_0(s) < \infty$  for  $|s| < R$  and  $\ell_0(s) = \infty$  for  $|s| > R$ . Moreover, the function  $\ell$  satisfies either one of the following conditions:*

$$(\ell 1) \quad \ell(1) \leq 0,$$

$$(\ell 2) \quad \text{or, } \ell(1) > 0, \text{ and } \int_1^{\bar{s}} \ell^{-1}(s)ds = T, \text{ for some constant } \bar{s} > 1.$$

(ii) *The terminal function satisfies  $|\psi|_0 < R$ .*

**Proposition 2.3.** *Let Assumption 2.2 hold true, then the BSDE (2.3) has a unique solution  $(Y, Z)$  such that  $\sup_{0 \leq t \leq T} |Y_t| < R$   $\mathbb{P}_0$ -almost surely.*

**Remark 2.4.** *When  $\psi \equiv 0$ , the function  $\ell$  in (2.5) is not well defined. In order to provide a sufficient condition for the power series representation, we can consider the BSDE (2.3) with terminal condition  $Y_T = \varepsilon$ . Define the corresponding function  $q_\varepsilon(s) := \beta[\varepsilon^{-1}\ell_0(\varepsilon s) - s]$ . Suppose that for some  $\varepsilon > 0$ , Assumption 2.2 holds true with the corresponding function  $q_\varepsilon$ , then by comparison result of BSDEs, the BSDE (2.3) admits still a unique solution  $(Y, Z)$  such that  $Y$  is uniformly bounded.*

In preparation of the proof, let us consider first the ordinary differential equation (ODE) of  $\rho(t)$  on interval  $[0, T]$ :

$$\rho' = \ell(\rho), \quad \text{with initial condition } \rho(0) = 1. \quad (2.6)$$

**Lemma 2.5.** *Let  $|\psi|_0 < R$ , then the ODE (2.6) admits a unique bounded solution on the interval  $[0, T]$  if and only if Assumption 2.2 (i) holds true. Moreover, in this case, we have  $0 \leq \rho(t) \leq \frac{R_0}{|\psi|_0}$ ,  $\forall t \in [0, T]$  for some  $R_0 < R$ .*

**Proof.** First, since the function  $\ell$  is Lipschitz on  $[0, L]$  for every  $L < \frac{R}{|\psi|_0}$ , then it follows by Picard-Lindelöf theorem (see e.g. Chapter 2 of Teschl [17]) that there is  $T_{\max} > 0$  such that the ODE (2.6) admits a unique solution  $\rho$  on  $[0, T_{\max})$  and that  $\lim_{t \rightarrow T_{\max}} |\rho(t)| = \frac{R}{|\psi|_0} > 1$ . Further, we observe that  $\ell(0) \geq 0$ , which implies that  $\rho(t) \geq 0$  on  $[0, T_{\max})$ . Then it is enough to prove that  $T_{\max} > T$ .

In the case that  $\ell(1) \leq 0$ , we have  $\rho(t) \in [0, 1]$  for every  $t \in [0, T_{\max})$  and hence  $T_{\max} = \infty > T$ . Otherwise, when  $\ell(1) > 0$ , we have  $\rho(t) \geq 1$  for  $t \in [0, T_{\max})$  and it follows by (2.6) and the fact that  $\rho(R/|\psi|_0) = \infty$  that  $\int_0^\infty \frac{1}{\ell(s)} ds = T_{\max}$ . We hence deduce that  $T < T_{\max}$  by Assumption 2.2 (l2).  $\square$

**Remark 2.6.** *The ODE (2.6) can be rewritten as*

$$\rho(t) = \rho(0) + \int_0^t \ell(\rho(s)) ds.$$

Let  $\varphi(t) := \rho(t)|\psi|_0$ , then under Assumption 2.2 we have

$$e^{\beta t} \varphi(t) = \varphi(0) + \int_0^t e^{\beta s} \beta \left( \sum_{k=0}^{\infty} |a_k|_0 \varphi^k(s) \right) ds. \quad (2.7)$$

In other words, the existence and uniqueness of solution to (2.6) is equivalent to that of (2.7).

**Remark 2.7.** *Suppose that  $a_k \equiv 0$  for every  $k > n_0$  with some  $n_0 \in \mathbb{N}$ , then clearly  $\ell(s) := \beta \left( \sum_{k=0}^{n_0} |a_k|_0 |\psi|_0^{k-1} s^k - s \right)$ . Denote  $\ell_\varepsilon(s) := \beta \left( \sum_{k=0}^{n_0} |(1 + \varepsilon)a_k|_0 (1 + \varepsilon) |\psi|_0^{k-1} s^k - s \right)$ . Let Assumption 2.2 hold true, then for  $\varepsilon > 0$  small enough, we still have  $\ell_\varepsilon(1) \leq 0$ , or  $\ell_\varepsilon(1) > 0$  and  $\int_1^{\bar{s}_\varepsilon} \frac{1}{\ell_\varepsilon(s)} ds = T$  for some constant  $1 < \bar{s}_\varepsilon < \infty$ . It follows that the ODE:  $\rho'(t) = \ell_\varepsilon(\rho)$  with initial condition  $\rho(0) = 1$  admits a unique solution on  $[0, T]$  under Assumption 2.2.*

With the above existence and uniqueness result of ODE (2.6), we get the existence and uniqueness of the BSDE (2.3) in Proposition 2.3.

**Proof of Proposition 2.3.** By Lemma 2.5, the solution  $\rho$  of ODE (2.6) is uniformly bounded by  $\frac{C}{|\psi|_0}$  with some constant  $C = R_0 < R$ , where  $R$  is the convergence radius of the power series  $\sum_{k=0}^{\infty} |a_k|_0 x^k$ . Denote  $y_C := -C \vee (y \wedge C)$  for every  $y \in \mathbb{R}$ ,

$$F_C(s, \mathbf{x}, y) := F(s, \mathbf{x}, y_C)$$

and

$$\bar{f}_C(s, \mathbf{x}, y) := \beta \left( \sum_{k=0}^{\infty} |a_k|_0 |y_C|^k - y_C \right), \quad \underline{f}_C(s, \mathbf{x}, y) := -\beta \left( \sum_{k=0}^{\infty} |a_k|_0 |y_C|^k + |y_C| \right).$$

Then  $F_C$ ,  $\bar{f}_C$  and  $\underline{f}_C$  are all globally Lipschitz in  $y$ , and  $\underline{f}_C \leq F_C \leq \bar{f}_C$ . Moreover, if we replace the generator  $F$  by  $\bar{f}_C$  (resp.  $\underline{f}_C$ ), and the terminal condition  $\psi$  by  $|\psi|_0$  (resp.  $-|\psi|_0$ ) in BSDE (2.3), the solution is given by  $\bar{Z} := 0$  (resp.  $\underline{Z} := 0$ ) and

$$\bar{Y}_t := \rho(T-t)|\psi|_0 \quad (\text{resp. } \underline{Y}_t := -\rho(T-t)|\psi|_0).$$

Therefore, by comparison principle, it follows that the solution  $(Y_C, Z_C)$  of BSDE (2.3) with generator  $f_C$  satisfies  $\underline{Y} \leq Y_C \leq \bar{Y}$ , and hence  $|Y_C| \leq C$ . Further, since  $F(t, \mathbf{x}, y) = F_C(t, \mathbf{x}, y)$  for all  $|y| \leq C$ , it follows that  $(Y_C, Z_C)$  is the required solution of BSDE (2.3).  $\square$

## 2.2 A branching diffusion process

Let  $\beta > 0$ ,  $n_0 \geq 0$  and  $p = (p_k)_{0 \leq k \leq n_0}$  be such that  $\sum_{k \leq n_0} p_k = 1$  and  $p_k \geq 0$ ,  $k = 0, \dots, n_0$ . We now construct a branching diffusion process as follows: a particle starts at time  $t$ , from position  $x$ , performs a diffusion process given by (2.2), dies after a mean  $\beta$  exponential time and produces  $k$  i.i.d. descendants with probability  $p_k$ . Then the descendants go on to perform diffusion process defined by (2.2) driven by independent Brownian motions. Every descendant dies and reproduces i.i.d. descendants independently after independent exponential times, etc. In the following, we shall give a mathematical construction of this branching diffusion process in three steps.

**A birth-death process** Suppose that we are given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , in which  $(T^{i,j})_{i,j \geq 0}$  is a sequence of i.i.d. random variables drawn from the  $\mathcal{E}(\beta)$  exponential distribution with mean  $\beta > 0$ ,  $(I_n)_{n \geq 1}$  is a sequence of i.i.d random variables which are independent of  $(T^{i,j})_{i,j \geq 0}$ , of multi-nomial distribution  $\mathcal{M}(p)$ , i.e.  $\mathbb{P}(I_1 = k) = p_k$ ,  $\forall k = 0, 1, \dots, n_0$ . We shall construct a continuous time birth-death process associated with the coefficient  $\beta > 0$  and the probability density sequence  $(p_k)_{0 \leq k \leq n_0}$ .

The branching process starts with a particle at time 0,  $N_t$  denotes the number the particles in the system, every particle runs an independent exponential time and then branches into  $k$  i.i.d. particles with probability  $p_k$ . We denote by  $T_n$  the  $n$ -th branching time of the whole system, at which one of the existing particles branches into  $I_n$  particles. Between  $T_n$  and  $T_{n+1}$ , every particle is indexed by  $(k_1, \dots, k_n) \in$

$\{1, \dots, n_0\}^n$ , which means that its parent particle is indexed by  $(k_1, \dots, k_{n-1})$  between  $T_{n-1}$  and  $T_n$ . We also have a bijection  $c$  between  $\mathbb{N}$  and  $\cup_{n \geq 1} 2^{\{1, \dots, n_0\}^n}$  defined by

$$c((k_1, \dots, k_n)) := \sum_{i=1}^n k_i (n_0 + 1)^i. \quad (2.8)$$

Denote by  $\mathcal{K}_t$  the collection of the indexes of all existing particles in the system at time  $t$ . Then the initial setting of the system is given by

$$N_0 = 1, \quad T_0 = 0, \quad T_1 = T^{0,0}, \quad \mathcal{K}_t = \{(1)\}, \quad \forall t \in [0, T_1],$$

and we have the induction relationship

$$N_{T_{i+1}} = N_{T_i} + I_{i+1} - 1, \quad T_{i+1} = T_i + \min_{k \in \mathcal{K}_{T_i}} T^{i,c(k)} = T_i + T^{i,c(K_{i+1})},$$

where  $K_{i+1}$  denote the index of the particle which branches at time  $T_{i+1}$ . Let

$$\mathcal{K}_{T_{i+1}} := \{(K_{i+1}, m) : 1 \leq m \leq I_{i+1}\} \cup \{(k, 1) : k \in \mathcal{K}_{T_i} \setminus \{K_{i+1}\}\}.$$

In particular, if  $I_{i+1} = 0$ , then  $\mathcal{K}_{T_{i+1}} = \{(k, 1) : k \in \mathcal{K}_{T_i} \setminus \{K_{i+1}\}\}$ . Clearly, at a branching time  $T_i$ , the particle  $K_i$  branches into  $I_i$  particles which are indexed by  $(K_i, 1), \dots, (K_i, I_i)$ , and all the other particles with index  $k$  are re-indexed by  $(k, 1)$ . Let

$$N_t := N_{T_i} \quad \text{and} \quad \mathcal{K}_t := \mathcal{K}_{T_i} \quad \text{for all} \quad t \in [T_i, T_{i+1}).$$

Then  $(N_t)_{t \geq 0}$  is a continuous Markov process taking value in  $\mathbb{N}$ . Since it is possible that a particle dies with  $k = 0$  descendants, the branching system is subject to extinction in finite time horizon, i.e.  $\mathbb{P}[N_t = 0 \text{ for some } t > 0] > 0$ . Furthermore,  $(\mathcal{K}_t)_{t \geq 0}$  is a random process taking value in  $\cup_{n \in \mathbb{N}} 2^{\{1, \dots, n_0\}^n}$ , and  $N_t = 0$  whenever  $\mathcal{K}_t$  is empty.

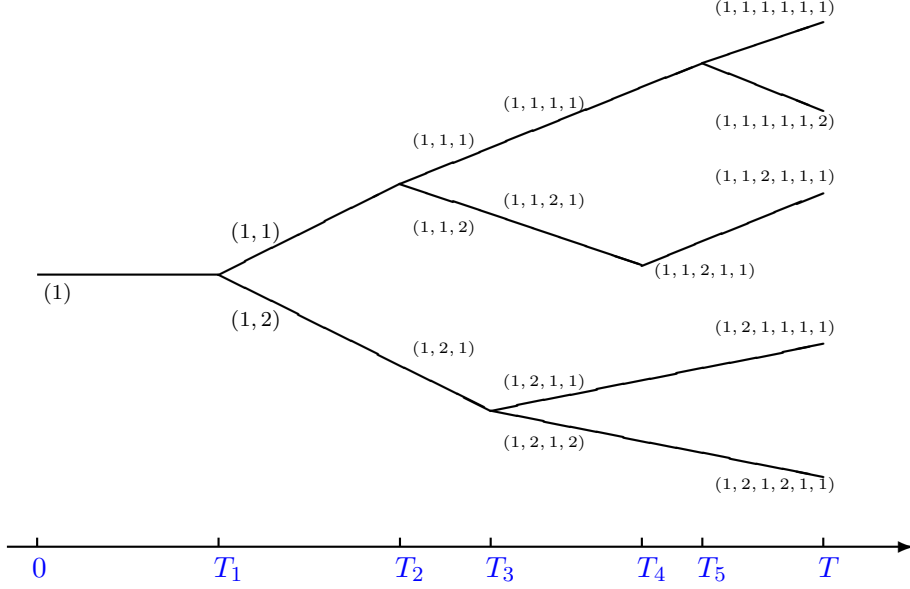
**Example 2.8.** *We give an example of the branching birth-death process, with graphic illustration below, where  $n_0 = 2$ . The process starts with one particle indexed by  $(1)$ , and branches at time  $T_1, \dots, T_5$ . The index of the branched particles are respectively  $(1)$ ,  $(1, 1)$ ,  $(1, 2, 1)$ ,  $(1, 1, 2, 1)$  and  $(1, 1, 1, 1, 1)$ . At terminal time  $T$ , the number of particles alive are  $N_T = 5$  and*

$$\mathcal{K}_T = \{(1, 1, 1, 1, 1), (1, 1, 1, 1, 1, 2), (1, 1, 2, 1, 1, 1), (1, 2, 1, 1, 1, 1), (1, 2, 1, 2, 1, 1)\}.$$

- At time  $T_1$ , particle  $(1)$  branches into two particles  $(1, 1)$  and  $(1, 2)$ .
- At time  $T_2$ , particle  $(1, 1)$  branches into  $(1, 1, 1)$  and  $(1, 1, 2)$ , particle  $(1, 2)$  is reindexed by  $(1, 2, 1)$ .
- At time  $T_3$ , particle  $(1, 2, 1)$  branches into  $(1, 2, 1, 1)$  and  $(1, 2, 1, 2)$ , the other two particles are reindexed by  $(1, 2, 1)$ .
- At time  $T_4$ , particle  $(1, 1, 2, 1)$  branches into one particle  $(1, 1, 2, 1, 1)$ , the other particles are reindexed.



- At time  $T_5$ , particle  $(1, 1, 1, 1, 1)$  branches into  $(1, 1, 1, 1, 1, 1)$  and  $(1, 1, 1, 1, 1, 2)$ , the other particles are reindexed by  $(1, 1, 2, 1, 1, 1)$ ,  $(1, 2, 1, 1, 1, 1)$ ,  $(1, 2, 1, 2, 1, 1)$ .



**Lemma 2.9.** For every probability density sequence  $(p_k)_{0 \leq k \leq n_0}$ , we have  $\lim_{n \rightarrow \infty} T_n = \infty$ , a.s. In particular, the system is well defined from 0 to  $\infty$ .

**Proof.** Without loss of generality, we can consider the case when  $p_k = 0$ ,  $\forall k < n_0$  and  $p_{n_0} = 1$ . We first claim that  $N_t < \infty$  for all  $t \geq 0$ , it follows that  $\sup\{n : T_n \leq t\} < \infty$  for all  $t \geq 0$  and hence  $\lim_{n \rightarrow \infty} T_n = \infty$ . Then to conclude, it is enough to prove that  $N_t < \infty$ ,  $\forall t \geq 0$ , which means that the population of the particles never explodes. In fact, it is enough to use Example 2 of Kersting and Klebaner [12] to finish the proof.  $\square$

**The branching Brownian motion** Suppose that in the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , there is a sequence of independent  $d$ -dimensional standard Brownian motions  $(W^1, W^2, \dots)$ , which is also independent of the exponential random variables  $(T^{i,j})_{i,j \geq 0}$  and multi-nomial random variables  $(I_n)_{n \geq 1}$ . We can then construct a branching Brownian motion which starts at time  $t \geq 0$ .

For the first particle in the system indexed by  $k = (1)$ , we associate it with a Brownian motion on  $[t, \infty)$ , defined by  $B_{t+s}^{t,(1)} = W_s^1$ ,  $\forall 0 \leq s \leq T_1$ . Let  $k = (k_1, \dots, k_n) \in \mathcal{K}_{T_n}$  be the index of a living particle at time  $T_n$ , whose parent particle is indexed by  $(k_1, \dots, k_{n-1})$ , we associate it with a Brownian motion between  $[t, t + T_{n+1}]$ , defined by

$$B_{t+s}^{t,k} := \begin{cases} B_{t+s}^{t,(k_1, \dots, k_{n-1})}, & \forall s \in [0, T_n], \\ B_{t+T_n}^{t,(k_1, \dots, k_{n-1})} + W_{s-T_n}^{c(k)}, & \forall s \in [T_n, T_{n+1}]. \end{cases}$$

By the strong Markov property of the Brownian motion, it is clear that conditioned on  $(T^{i,j})_{i,j \geq 0}$  and  $(I_n)_{n \geq 0}$ , every process  $(B_r^{t,k})_{r \geq t}$  for  $k \in \mathcal{K}_T$  is a Brownian motion.

In particular, given two particles  $k^1 = (k_1^1, \dots, k_n^1)$  and  $k^2 = (k_1^2, \dots, k_n^2)$  such that  $k_j^1 = k_j^2$  for all  $j = 1, \dots, i$ , the associated Brownian motion  $B^{t,k^1}$  and  $B^{t,k^2}$  share the same path before time  $t + T_i$ .

**The branching diffusion process** To construct a branching diffusion process, we first remark that for every  $(t, \mathbf{x}) \in \Lambda^0$ , the SDE (2.2) with initial condition  ${}^{t,\mathbf{x}}X_s = \mathbf{x}_s$ ,  $0 \leq s \leq t$  has a unique strong solution  ${}^{t,\mathbf{x}}X$  adapted to the natural Brownian filtration, hence there is a progressively measurable function  $\Phi^{t,\mathbf{x}} : [t, T] \times C([t, T], \mathbb{R}^d) \rightarrow \mathbb{R}$  such that  ${}^{t,\mathbf{x}}X_s = \Phi^{t,\mathbf{x}}(s, B)$ ,  $\mathbb{P}_0$  - a.s..

Then a branching diffusion process  ${}^{t,\mathbf{x}}X^k$  is given by

$${}^{t,\mathbf{x}}X_{t+s}^k := \Phi^{t,\mathbf{x}}(t+s, B^{t,k}), \quad \forall s \in \mathbb{R}^+ \text{ and } k \in \mathcal{K}_s. \quad (2.9)$$

Moreover, for later uses, we extend  ${}^{t,\mathbf{x}}X^{(1)}$  on the whole interval  $[0, T]$  by

$${}^{t,\mathbf{x}}X_s^{(1)} := \mathbf{x}_s \quad \forall s \leq t \quad \text{and} \quad {}^{t,\mathbf{x}}X_s^{(1)} := \Phi^{t,\mathbf{x}}(s, B^{t,(1)}), \quad \forall s \geq t.$$

**Remark 2.10.** *By the flow property of the SDE (2.2), we have that for every  $(t, \mathbf{x}) \in \Lambda^0$ ,  $r \leq s$  and  $k \in \mathcal{K}_s$ ,*

$$\Phi^{t,\mathbf{x}}(t+s, (B_u^{t,k})_{u \geq t}) = \Phi^{t+r, {}^{t,\mathbf{x}}X^k}(t+s, (B_u^{t,k})_{u \geq t+r}), \quad \mathbb{P} - a.s. \quad (2.10)$$

To conclude this subsection, we equip the above system with two filtrations. First,  $\overline{\mathbb{F}} = (\overline{\mathcal{F}}_t)_{t \geq 0}$  with

$$\overline{\mathcal{F}}_t := \sigma((T_n, I_n, K_n)1_{T_n \leq t} + \partial 1_{T_n > t}, n \geq 1),$$

where  $\partial$  denotes a cemetery point. Intuitively,  $\overline{\mathbb{F}}$  is the filtration generated by the birth-death process. In particular,  $T_n$  is a  $\overline{\mathbb{F}}$ -stopping time and  $\overline{\mathcal{F}}_{T_n} = \sigma((T_k, I_k, K_k)_{1 \leq k \leq n})$ . Next, for every  $t \geq 0$ , let  $\overline{\mathbb{F}}^t = (\overline{\mathcal{F}}_{t+s}^t)_{s \geq 0}$  be the filtration on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  generated by the branching diffusion process, which is defined by

$$\overline{\mathcal{F}}_{t+s}^t := \overline{\mathcal{F}}_s \bigvee \sigma(\mathcal{K}_r, B_r^{(1)}, B_{t+r}^{t,k}, 0 \leq r \leq s, k \in \mathcal{K}_s). \quad (2.11)$$

### 2.3 Branching diffusion representation of backward SDE

Using the branching diffusion process defined above, we can provide a representation of the solution to the decoupled FBSDE (2.3).

Let  $(t, \mathbf{x}) \in \Lambda^0$ , we consider the branching diffusion process  $({}^{t,\mathbf{x}}X^k)_{k \in \mathcal{K}_T}$  on  $[t, T]$  defined in (2.9), where the probability sequence  $p = (p_k)_{0 \leq k \leq n_0}$  satisfies that  $p_k > 0$  whenever  $|a_k|_0 \neq 0$ . Denote

$$\mathcal{W}_{t,\mathbf{x}} := \prod_{n=1}^{M_{T-t}} \left( \frac{a_{I_n}(t+T_n, {}^{t,\mathbf{x}}X^{K_n})}{p_{I_n}} \right), \quad \text{where } M_{T-t} := \sup\{n : t+T_n \leq T\}$$

is the number of branchings occurred in the particles system before time  $T$ . Our main representation formula is the following function:

$$v(t, \mathbf{x}) := \mathbb{E}^{\mathbb{P}} [\Psi_{t,\mathbf{x}}] \quad \text{with} \quad \Psi_{t,\mathbf{x}} := \mathcal{W}_{t,\mathbf{x}} \prod_{k \in \mathcal{K}_{T-t}} \psi({}^{t,\mathbf{x}}X^k), \quad (2.12)$$

where the integrability of  $\Psi_{t,\mathbf{x}}$  is verified in the following result.

**Proposition 2.11.** *Suppose that Assumption 2.2 holds true. Then for every  $(t, \mathbf{x}) \in \Lambda^0$ , the random variable  $\Psi_{t, \mathbf{x}}$  given in (2.12) is integrable and the value function  $v$  is uniformly bounded. Moreover, for every  $M > 0$ , there is a constant  $C$  such that*

$$|v(t, \omega) - v(t', \omega')| \leq C(\sqrt{|t - t'|} + \|\omega_{t \wedge \cdot} - \omega'_{t' \wedge \cdot}\|_T),$$

whenever  $|(t, \omega)| \leq M$  and  $|(t', \omega')| \leq M$ .

The proof of Proposition 2.11 will be completed later in Section 3.

Our main result of the paper is the following representation theorem. Let  ${}^{0,0}X$  be the unique strong solution to the SDE (2.2) in the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , denote

$$Y_t^0 := v(t, {}^{0,0}X). \quad (2.13)$$

We also consider the BSDE (2.3) with generator

$$F_{n_0}(t, \mathbf{x}, y) := \beta \left( \sum_{k=0}^{n_0} a_k(t, \mathbf{x}) y^k - y \right). \quad (2.14)$$

We define  $\ell_{n_0}$  by

$$\ell_{n_0}(s) := \beta \left( \sum_{k=0}^{n_0} |a_k|_0 |\psi|_0^{k-1} s^k - s \right), \quad \forall s \geq 0.$$

It is clear that when Assumption 2.2 holds true for  $\ell$ , then  $\ell_{n_0}$  satisfies also Assumption 2.2. It follows from Proposition 2.3 that the BSDE (2.3) with generator  $F_{n_0}$  has a unique solution, denoted by  $(Y, Z)$ , such that  $Y$  is uniformly bounded.

**Theorem 2.12.** *Suppose that Assumption 2.2 holds true, and  $(Y, Z)$  is the unique solution of BSDE (2.3) with generator  $F_{n_0}$  (defined by (2.14)) such that  $Y$  is uniformly bounded by  $R_0$ , the constant introduced in Lemma 2.5. Then  $Y^0 = Y$ ,  $\mathbb{P}_0$ -a.s.*

The proof of this result will be provided in Section 4 using the notion of viscosity solutions to a path dependent PDE.

## 2.4 Numerical algorithm by branching process

The representation result in Theorem 2.12 induces immediately a numerical algorithm for BSDE (2.3) by simulating the branching diffusion process. For numerical implementation, the branching times can be simulated exactly since they follow the exponential law, and the diffusion process can be simulated by a Euler scheme.

Let  $\Delta = (t_0, \dots, t_n)$  be a discretization of the interval  $[0, T]$ , i.e.  $0 = t_0 < \dots < t_n = T$ . Denote  $|\Delta| := \max_{1 \leq k \leq n} (t_k - t_{k-1})$ . To give the Euler scheme, we introduce the frozen coefficients  $\mu^\Delta$  and  $\sigma^\Delta$  by

$$\mu^\Delta(t, \mathbf{x}) := \mu(t_k, \hat{\mathbf{x}}^\Delta) \quad \text{and} \quad \sigma^\Delta(t, \mathbf{x}) := \sigma(t_k, \hat{\mathbf{x}}^\Delta), \quad \forall t \in [t_k, t_{k+1}),$$

where  $\hat{\mathbf{x}}^\Delta$  denotes the linear interpolation of  $(\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n})$  on the interval  $[0, T]$ . Then clearly the process  $X^\Delta$  given by the SDE

$$X_t^\Delta = \int_0^t \mu^\Delta(s, X_s^\Delta) ds + \int_0^t \sigma^\Delta(s, X_s^\Delta) dB_s, \quad \mathbb{P}_0 - a.s. \quad (2.15)$$

can be simulated, which is also the Euler scheme of the SDE (2.2). By standard arguments using Gronwall's Lemma (see e.g. Kloeden Platen [13] or Graham and Talay [10] in the Markov case), we have the following error analysis result: Let  $X$  be the solution process of (2.2) with initial condition  $(t, \mathbf{x}) = (0, \mathbf{0})$ ,  $X^\Delta$  be the solution of (2.15) and  $\widehat{X}^\Delta$  denotes the linear interpolation of  $(X_{t_0}^\Delta, \dots, X_{t_n}^\Delta)$  on  $[0, T]$ .

**Lemma 2.13.** *There is a constant  $C$  independent of the discretization  $\Delta$  such that*

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} \left( |X_t - X_t^\Delta|^2 + |X_t^\Delta - \widehat{X}_t^\Delta|^2 \right) \right] \leq C |\Delta|.$$

Moreover, for the BSDE (2.3) with a general generator function  $F : [0, T] \times \Omega^0 \times \mathbb{R} \rightarrow \mathbb{R}$  which admits a representation (2.4), we can approximate it by some polynomial  $F_{n_0}$  of the form (2.14). Let  $F_{n_0}^\Delta(t, \mathbf{x}, y) := \beta \left( \sum_{k=0}^{n_0} a_k^\Delta(t, \mathbf{x}) y^k - y \right)$ , where

$$a_k^\Delta(t, \mathbf{x}) := a_k(t_i, \widehat{\mathbf{x}}^\Delta) \quad \text{for every } k = 0, \dots, n_0 \text{ and } t \in [t_i, t_{i+1}).$$

Further, under Assumption 2.2, by simulating the branching diffusion process  $(X^{\Delta, k})_{k \in \mathcal{K}_T}$ , the numerical solution

$$Y_0^\Delta := \mathbb{E} \left[ \prod_{n=1}^{M_T} \left( \frac{a_{I_n}^\Delta(T_n, \widehat{X}^{\Delta, K_n})}{p_{I_n}} \right) \prod_{k \in \mathcal{K}_T} \psi(\widehat{X}^{\Delta, k}) \right]$$

is the solution of the following BSDE

$$Y_0 = \psi(\widehat{X}^\Delta) + \int_0^T F_{n_0}^\Delta(t, \widehat{X}^\Delta, Y_t) dt - \int_0^T Z_t dB_t, \quad \mathbb{P}_0 - a.s.$$

Finally, we provide an error estimation of the numerical solution:

**Proposition 2.14.** *Under Assumption 2.2, there is a constant  $C$  independent of  $n_0$  and  $\Delta$  such that*

$$|Y_0^\Delta - Y_0| \leq C (|F - F_{n_0}|_{L^\infty(\Lambda^0 \times [-R_0, R_0])} + \sqrt{\Delta}).$$

**Proof.** This estimate follows from a direct application of the stability result of backward SDEs together with the error estimation in Lemma 2.13, see Proposition 2.1 and the subsequent remark in El Karoui, Peng and Quenez [8].  $\square$

### 3 Hölder and Lipschitz regularity of $v$

This section is devoted to the proof of Proposition 2.11. We first derive some estimates of the birth-death process defined in Section 2.2, then together with the tower property, we can complete the proof of Proposition 2.11.

### 3.1 Some estimates of the birth-death process

We recall that  $\overline{\mathbb{F}} = (\overline{\mathcal{F}}_t)_{0 \leq t \leq T}$  is the filtration generated by the birth-death process defined in the end of Section 2.2, and that the number of branchings occurred in the system before time  $t$  is denote by  $M_t := \sup \{n : T_n \leq t\}$ . We also introduce:

$$\eta(t) := \mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} \right].$$

**Lemma 3.1.** *For every  $0 \leq s \leq t$ ,*

$$\mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} \mid \overline{\mathcal{F}}_s \right] = \left( \prod_{n=1}^{M_s} \frac{|a_{I_n}|_0}{p_{I_n}} \right) (\eta(t-s))^{N_s},$$

and

$$\mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} 1_{T_1 \leq t} \mid \overline{\mathcal{F}}_{T_1} \right] = \frac{|a_{I_1}|_0}{p_{I_1}} (\eta(t-T_1))^{I_1} 1_{T_1 \leq t}.$$

**Proof.** Let  $Z$  be a random variable and  $A \in \mathcal{F}$ , then  $\mathcal{L}^{\mathbb{P}}(Z)$  denotes the law of  $Z$  and  $\mathcal{L}^{\mathbb{P}}(Z|A)$  denotes the distribution of  $Z$  conditioned on  $A$  under the probability  $\mathbb{P}$ . We notice that for every  $i, j \geq 1$  and  $s > 0$ ,

$$\mathcal{L}^{\mathbb{P}}(T^{i,j} | T^{i,j} > s) = \mathcal{L}^{\mathbb{P}}(T^{i,j}) = \mathcal{E}(\beta).$$

Let  $0 \leq s \leq t$ , the law of number of branches between  $s$  and  $t$  (which equals to  $M_t - M_s$ ) is completely determined by  $N_s$ ,  $(T^{i,j})_{i \geq M_s, j \geq 0}$  and  $(I_i)_{i \geq M_s+1}$ . It follows that

$$\begin{aligned} & \mathcal{L}^{\mathbb{P}}(M_t - M_s, (I_{M_s+i})_{i \geq 1} \mid N_s = 1, k \in \mathcal{K}_s, M_s = j) \\ &= \mathcal{L}^{\mathbb{P}}(M_t - M_s, (I_{M_s+i})_{i \geq 1} \mid N_s = 1, k \in \mathcal{K}_s, M_s = j, T^{M_s, c(k)} > s) \\ &= \mathcal{L}^{\mathbb{P}}(M_{t-s}, (I_i)_{i \geq 1}), \end{aligned}$$

and hence

$$\mathcal{L}^{\mathbb{P}}(M_t - M_s, (I_{M_s+i})_{i \geq 1} \mid N_s = 1) = \mathcal{L}^{\mathbb{P}}(M_{t-s}, (I_i)_{i \geq 1}).$$

Since  $N_t = N_s + \sum_{n=M_s+1}^{M_t} (I_n - 1)$ , we deduce that

$$\mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=M_s+1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} \mid N_s = 1 \right] = \eta(t-s).$$

Moreover, since every particle branches independently to each other, we deduce that

$$\mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=M_s+1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} \mid N_s = i \right] = (\eta(t-s))^i,$$

which implies that

$$\mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=M_s+1}^{M_t} \frac{|a_{I_n}|_0}{p_{I_n}} \right) |\psi|_0^{N_t} \mid N_s \right] = (\eta(t-s))^{N_s}.$$

And hence

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\left(\prod_{n=1}^{M_t}\frac{|a_{I_n}|_0}{p_{I_n}}\right)|\psi|_0^{N_t}\mid\bar{\mathcal{F}}_s\right] &= \mathbb{E}^{\mathbb{P}}\left[\left(\prod_{n=1}^{M_s}\frac{|a_{I_n}|_0}{p_{I_n}}\right)\left(\prod_{n=M_s+1}^{M_t}\frac{|a_{I_n}|_0}{p_{I_n}}\right)|\psi|_0^{N_t}\mid\bar{\mathcal{F}}_s\right] \\ &= \left(\prod_{n=1}^{M_s}\frac{|a_{I_n}|_0}{p_{I_n}}\right)(\eta(t-s))^{N_s}.\end{aligned}$$

For the second equality, we remark that  $(I_i)_{i\geq 2}$  and  $(T^{i,j})_{i\geq 2, j\geq 0}$  are all independent of  $(T_1, I_1)$ . Then consider a family of conditional probabilities  $(\mathbb{P}_{s,i})_{s\in\mathbb{R}^+, i\in\{0,\dots,n_0\}}$  of  $\mathbb{P}$  w.r.t. the  $\sigma$ -field generated by  $(T_1, I_1)$ . Under a probability  $\mathbb{P}_{s,i}$ , the law of  $(M_t, N_t)1_{s\leq t}$  depends only on  $(I_j)_{j\geq 2}$  and  $(T^{j,l})_{j\geq 2, l\geq 0}$ . Considering in particular  $i = 1$ , we have

$$\mathcal{L}^{\mathbb{P}_{s,1}}(M_t - M_s, (I_i)_{i\geq 2}) = \mathcal{L}^{\mathbb{P}}(M_{t-s}, (I_i)_{i\geq 1}).$$

And hence

$$\mathbb{E}^{\mathbb{P}_{s,1}}\left[\left(\prod_{n=M_s+1}^{M_t}\frac{|a_{I_n}|_0}{p_{I_n}}\right)|\psi|_0^{N_t}1_{s\leq t}\right] = \eta(t-s)1_{s\leq t}.$$

Moreover, by the independence of the evolution of  $i$  particles under  $\mathbb{P}_{s,i}$ , we get

$$\mathbb{E}^{\mathbb{P}_{s,i}}\left[\left(\prod_{n=M_s+1}^{M_t}\frac{|a_{I_n}|_0}{p_{I_n}}\right)|\psi|_0^{N_t}1_{s\leq t}\right] = (\eta(t-s))^i 1_{s\leq t}.$$

which concludes the proof.  $\square$

**Lemma 3.2.** *Suppose that for some  $t \geq 0$ ,  $\eta(t) < \infty$ . Then there is  $\delta > 0$  such that  $\eta(s) < \infty$ , for every  $s \in [t, t + \delta]$ .*

**Proof.** First, it follows from Lemma 3.1 that for every  $t$ ,  $\delta \geq 0$ ,

$$\eta(t + \delta) = \mathbb{E}^{\mathbb{P}}\left[\left(\prod_{n=1}^{M_\delta}\frac{|a_{I_n}|_0}{p_{I_n}}\right)(\eta(t))^{N_\delta}\right].$$

Let us consider another pure birth process  $(\tilde{N}_t, \tilde{\mathcal{K}}_t)$  on a probability space  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$  with the same constant characteristics  $\beta$  and  $(\tilde{p}_k)_{0\leq k\leq n_0}$  such that  $\tilde{p}_{n_0} = 1$ . We suppose without loss of generality that  $n_0 \geq 2$  and denote  $C := \max_{0\leq k\leq n_0}\frac{|a_k|_0}{p_k} + \eta(t)$ . Then clearly it is enough to prove that  $\mathbb{E}^{\tilde{\mathbb{P}}}[C^{\tilde{N}_\delta}] < \infty$  for some  $\delta > 0$  to conclude the proof. The distribution of  $\tilde{N}_\delta$  can be computed explicitly (see e.g. Athreya and Ney [1]) and verifies that for some constant  $\tilde{C} > 0$ ,

$$\mathbb{P}[\tilde{N}_\delta = n] \leq \tilde{C}\kappa_\delta^n \quad \text{with} \quad \kappa_\delta := (1 - e^{-\delta\beta(n_0-1)})^{1/(n_0-1)}.$$

Then for  $\delta > 0$  small enough,  $\kappa_\delta$  is small enough such that  $\mathbb{E}^{\tilde{\mathbb{P}}}[C^{\tilde{N}_\delta}] < \infty$ .  $\square$

The birth-death system is closely related to the ODE (2.6). Let us define

$$D_t := \mathbb{E}^{\mathbb{P}}\left[(1 \vee M_t)(1 \vee N_t)\left(\prod_{n=1}^{M_t}\frac{|a_{I_n}|_0}{p_{I_n}}\right)|\psi|_0^{N_t}\right]. \quad (3.1)$$

**Proposition 3.3.** *Suppose that Assumption 2.2 holds true. Then*

$$\sup_{0 \leq t \leq T} \eta(t) < \infty \quad \text{and} \quad \sup_{0 \leq t \leq T} D_t < \infty. \quad (3.2)$$

**Proof.** We first observe that  $\eta(0) = |\psi|_0$  by its definition, and it follows from Lemma 3.1 that

$$\begin{aligned} \eta(t) &= \mathbb{E}[|\psi|_0 \mathbf{1}_{T_1 > t}] + \mathbb{E}\left[\sum_{k=0}^{n_0} |a_k|_0 (\eta(t - T_1))^k \mathbf{1}_{T_1 \leq t}\right] \\ &= \eta(0)e^{-\beta t} + \int_0^t \beta e^{-\beta s} \left(\sum_{k=0}^{n_0} |a_k|_0 (\eta(t - s))^k\right) ds \\ &= e^{-\beta t} \left(\eta(0) + \int_0^t \beta e^{\beta s} \left(\sum_{k=0}^{n_0} |a_k|_0 (\eta(s))^k\right) ds\right). \end{aligned}$$

Suppose that  $T_0 := \inf\{s : \eta(s) = \infty\} \leq T$ , then it follows from Lemma 2.5 and Remark 2.6 that  $\eta(t) = \rho(t)|\psi|_0 < \infty$ , where  $\rho$  is the unique solution to the ODE (2.6). This contradicts Lemma 3.2. We then have  $T_0 > T$  and  $\eta(T) < \infty$ . Since  $\eta$  is increasing, this provides the first claim in (3.2).

We next denote

$$\eta_\varepsilon(t) := \mathbb{E}^{\mathbb{P}}\left[\left(\prod_{n=1}^{M_t} \frac{(1 + \varepsilon)a_{I_n}|_0}{p_{I_n}}\right) |(1 + \varepsilon)\psi|_0^{N_t}\right].$$

In spirit of Remark 2.7, we know that for  $\varepsilon > 0$  small enough,  $\eta_\varepsilon(T) < \infty$ . It follows that

$$\sup_{0 \leq t \leq T} D_t := \sup_{0 \leq t \leq T} \mathbb{E}^{\mathbb{P}}\left[(1 \vee M_t)(1 \vee N_t) \left(\prod_{n, T_n \leq t} \frac{|a_{I_n}|_0}{p_{I_n}}\right) |\psi|_0^{N_t}\right] < \infty,$$

since there is some constant  $C_\varepsilon > 0$  such that  $n < C_\varepsilon(1 + \varepsilon)^n$  for every  $n \geq 0$ . And we hence conclude the proof.  $\square$

## 3.2 Proof of Proposition 2.11

In preparation of the proof, we first provide a tower property of the branching diffusion process. Let  $(t, \mathbf{x}) \in \Lambda^0$  and  $\tau : \Omega^0 \rightarrow \mathbb{R}^+$  be a  $\mathbb{F}^0$ -stopping time such that  $\tau \geq t$ , then  $\hat{\tau} := \tau(\cdot, {}^{t, \mathbf{x}}X^{(1)})$  is a  $\bar{\mathbb{F}}^t$ -stopping time in the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Lemma 3.4.** *Suppose that Assumption 2.2 holds true, let  $(t, \mathbf{x}) \in \Lambda^0$ ,  $0 \leq s \leq T - t$  and  $\hat{\tau}$  be given above. Then we have*

$$\mathbb{E}^{\mathbb{P}}[\Psi_{t, \mathbf{x}} | \bar{\mathcal{F}}_{t+s}^t] = \left(\prod_{n=1}^{M_s} \frac{a_{I_n}(t + T_n, {}^{t, \mathbf{x}}X^{K_n})}{p_{I_n}}\right) \prod_{k \in \mathcal{K}_s} v(t + s, {}^{t, \mathbf{x}}X^k) \quad (3.3)$$

and

$$\begin{aligned} v(t, \mathbf{x}) &= \mathbb{E}^{\mathbb{P}}\left[v(\hat{\tau}, {}^{t, \mathbf{x}}X^{(1)}) \mathbf{1}_{t+T_1 > \hat{\tau}}\right. \\ &\quad \left. + \frac{a_{I_1}(t + T_1, {}^{t, \mathbf{x}}X^{(1)})}{p_{I_1}} v^{I_1}(t + T_1, {}^{t, \mathbf{x}}X^{(1)}) \mathbf{1}_{t+T_1 \leq \hat{\tau}}\right]. \quad (3.4) \end{aligned}$$

**Proof.** First, following the arguments of Lemma 3.1, we know

$$\begin{aligned} & \mathcal{L}^{\mathbb{P}}(M_t - M_s, (I_{M_s+j})_{j \geq 1}, (W^{M_s+l})_{l \geq 1} \mid N_s = 1, M_s = i) \\ &= \mathcal{L}^{\mathbb{P}}(M_{t-s}, (I_j)_{j \geq 1}, (W^l)_{l \geq 1}). \end{aligned}$$

Together with the flow property of SDE in (2.10), it follows that

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[ \left( \prod_{n=M_s+1}^{M_{T-t}} \frac{a_{I_n}(t+T_n, {}^{t,\mathbf{x}}X^{K_n})}{p_{I_n}} \right) \prod_{k \in \mathcal{K}_{T-t}} \psi(X^{t,\mathbf{x},k}) \right. \\ \left. \mid N_s = 1, M_s = i, k \in \mathcal{K}_s, ({}^{t,\mathbf{x}}X_r^k)_{t \leq r \leq t+s} \right] = v(t+s, {}^{t,\mathbf{x}}X^k). \end{aligned}$$

Then by the independence of evolution of every particle in  $\mathcal{K}_s$ , (3.3) holds true.

For the second equality, we consider a regular conditional probability distribution (r.c.p.d.)  $(\mathbb{P}_\omega)_{\omega \in \Omega^0}$  of  $\mathbb{P}$  w.r.t.  $\sigma(B_{\hat{\tau} \wedge \cdot}^{(1)})$  (see also Stroock Varadhan [18] for the notion of r.c.p.d.). Then using the fact that under  $\mathbb{P}_\omega$ , the law of  $T_1$  conditioned on  $T_1 > \hat{\tau} - t$  is the same as that of  $T_1$ , we get that

$$\mathbb{E}^{\mathbb{P}_\omega} [\Psi_{t,\mathbf{x}} 1_{t+T_1 > \hat{\tau}(\omega)} \mid t + T_1 > \hat{\tau}(\omega)] = v(\hat{\tau}(\omega), X^{\hat{\tau}, \mathbf{x}, (1)}).$$

Further, using similar arguments as in Lemma 3.1, by considering the distribution of  $\Psi_{t,\mathbf{x}} 1_{t+T_1 \leq \hat{\tau}}$  conditioned on  $\overline{\mathcal{F}}_{T_1}^1$ , we get

$$\mathbb{E}^{\mathbb{P}} [\Psi_{t,\mathbf{x}} 1_{t+T_1 \leq \hat{\tau}}] = \mathbb{E}^{\mathbb{P}} \left[ \frac{a_{I_1}(t+T_1, {}^{t,\mathbf{x}}X^{(1)})}{p_{I_1}} v^{I_1}(t+T_1, {}^{t,\mathbf{x}}X^{(1)}) 1_{t+T_1 \leq \hat{\tau}} \right],$$

which concludes the proof.  $\square$

**Proof of Proposition 2.11.** (i) First, it follows immediately from Proposition 3.3 that  $\Psi_{t,\mathbf{x}}$  is integrable and  $|v(t, \mathbf{x})| \leq \rho(T-t)|\psi|_0 \leq R_0$ .

(ii) Let  $t \in [0, T]$  and  $\mathbf{x}_1, \mathbf{x}_2 \in \Omega^0$ . By Lemma 2.1, for every  $s \in [t, T]$  and  $k \in \mathcal{K}_s$ , we have

$$\mathbb{E}^{\mathbb{P}} \left[ \sup_{t \leq r \leq s} |{}^{t,\mathbf{x}_1}X_r^k - {}^{t,\mathbf{x}_2}X_r^k| \right] \leq C(1 + \|\mathbf{x}_1\|_t + \|\mathbf{x}_2\|_t) \|\mathbf{x}_1 - \mathbf{x}_2\|_t,$$

for some constant  $C$  independent of  $\mathbf{x}_1, \mathbf{x}_2$ . Then using the fact that  $(a_k)_{0 \leq k \leq n_0}$  and  $\psi$  are all Lipschitz in  $\mathbf{x}$ ,

$$\begin{aligned} |v(t, \mathbf{x}_1) - v(t, \mathbf{x}_2)| &\leq \mathbb{E}^{\mathbb{P}} [|\Psi_{t,\mathbf{x}_1} - \Psi_{t,\mathbf{x}_2}|] \leq C D_t \mathbb{E}^{\mathbb{P}} [|\| {}^{t,\mathbf{x}_1}X - {}^{t,\mathbf{x}_2}X \|_T|] \\ &\leq C(1 + \|\mathbf{x}_1\|_t + \|\mathbf{x}_2\|_t) \|\mathbf{x}_1 - \mathbf{x}_2\|_t, \end{aligned}$$

where  $D_t$  is defined in (3.1).

(iii) Let  $0 \leq s \leq t \leq T$ , then it follows from Lemma 3.4 that

$$\begin{aligned} |v(s, \mathbf{x}) - v(t, \mathbf{x}_{s \wedge \cdot})| &\leq \left| \mathbb{E}^{\mathbb{P}} \left[ \prod_{n=1}^{M_{t-s}} \frac{a_{I_n}(T_n, {}^{s,\mathbf{x}}X^{K_n})}{p_{I_n}} \prod_{k \in \mathcal{K}_{t-s}} v(t, {}^{s,\mathbf{x}}X^k) \right] - v(t, \mathbf{x}_{s \wedge \cdot}) \right| \\ &\leq C \left( \sup_{s \leq r \leq t} D_r \right) \mathbb{E}^{\mathbb{P}} \left[ \sup_{r \in [s,t]} |\mathbf{x}_s - {}^{s,\mathbf{x}}X_r^k| \right] \\ &\quad + \left| \mathbb{E}^{\mathbb{P}} \left[ \prod_{n=1}^{M_{t-s}} \frac{a_{I_n}(t, \mathbf{x}_{s \wedge \cdot})}{p_{I_n}} (v(t, \mathbf{x}_{s \wedge \cdot}))^{N_{t-s}} \right] - v(t, \mathbf{x}_{s \wedge \cdot}) \right| \\ &\leq C(1 + \|\mathbf{x}\|_s) \sqrt{t-s} + |\phi(t) - \phi(s)|, \end{aligned}$$



where  $\phi$  is the unique solution of the ODE

$$\phi'(r) = \beta \left( \sum_{k=0}^{n_0} a_k(t, \mathbf{x}) \phi^k(r) - \phi(r) \right) \quad \text{with terminal condition} \quad \phi(t) = v(t, \mathbf{x}_{s\wedge \cdot}).$$

Moreover, by comparison principle of ODE,  $|\phi(r)| \leq \rho(r)$ ,  $\forall r \in [s, t]$ . Then  $|\phi(t) - \phi(s)| \leq C(t - s)$  for some constant  $C$  independent of  $(s, t, \mathbf{x})$ , which implies that  $v$  is locally  $(1/2)$ -Hölder in  $t$ .  $\square$

**Remark 3.5.** When  $(a_k)_{0 \leq k \leq n_0}$  and  $\psi$  are all constants, the value function  $v(t, \mathbf{x})$  is independent of  $\mathbf{x}$  and  $t \mapsto v(T - t, \mathbf{x}) |\psi|_0^{-1}$  is a solution of the ODE (2.6). Therefore, in spirit of Lemma 2.5, Assumption 2.2 is also a necessary condition for the integrability of  $\Psi_{0,0}$ .

## 4 The branching diffusion representation result

This section is devoted to the proof of Theorem 2.12.

We first consider a class of semi-linear parabolic path-dependent PDEs (PPDEs) and introduce a notion of viscosity solution, following Ekren, Keller, Touzi and Zhang [5] and Ekren, Touzi and Zhang [6, 7]. Our objective is to show that the value function  $v$ , defined by our branching diffusion representation, and the  $Y$ -component of the BSDE are viscosity solutions of the same path-dependent PDE. Then, our main result follows from a uniqueness argument.

### 4.1 Viscosity solutions of PPDEs and FBSDEs

We consider a PPDE which is linear in the first and second order derivatives of the solution function. This is a simpler context than that of [5, 6, 7]. As a consequence, following Remark 3.9 in [6], we use a simpler definition of viscosity solutions. We shall also provide an (easy) adaptation of the arguments in [6] which relaxes their boundedness conditions, thus allowing the terminal condition and the generator to have linear growth.

#### 4.1.1 Differentiability on the canonical space

For all  $t \in [0, T]$ , we denote by  $\Omega^t := \{\omega \in C([t, T], \mathbb{R}^d) : \omega_t = 0\}$  the shifted canonical space,  $B^t$  the shifted canonical process on  $\Omega^t$ ,  $\mathbb{F}^t$  the shifted canonical filtration generated by  $B^t$ ,  $\mathbb{P}_0^t$  the Wiener measure on  $\Omega^t$  and  $\Lambda^t := [t, T] \times \Omega^t$ .

For  $s \leq t$ ,  $\omega \in \Omega^s$  and  $\omega' \in \Omega^t$ , define the concatenation path  $\omega \otimes_t \omega' \in \Omega^s$  by

$$(\omega \otimes \omega')(r) := \omega_r 1_{s \leq r < t} + (\omega_t + \omega'_r) 1_{t \leq r \leq T}, \quad \forall r \in [s, T].$$

Let  $\xi \in \mathcal{F}_T^0$  and  $V$  be a  $\mathbb{F}^0$ -progressive process, then for every  $(t, \omega) \in \Lambda^0$ , we define  $\xi^{t, \omega} \in \mathcal{F}_T^t$  and  $(V_s^{t, \omega})_{t \leq s \leq T}$  by

$$\xi^{t, \omega}(\omega') := \xi(\omega \otimes_t \omega'), \quad V_s^{t, \omega}(\omega') := V_s(\omega \otimes_t \omega'), \quad \forall \omega' \in \Omega^t.$$

Following Ekren et al. [6, 7], we define some classes of processes in  $\Lambda^t$ ,  $t \geq 0$ . Let  $C^0(\Lambda^t)$  be the collection of all  $\mathbb{F}^t$ -progressive processes which are continuous under the norm  $d_\infty$ , where

$$d_\infty((s, \omega), (s', \omega')) := |s - s'| + \sup_{t \leq r \leq T} |\omega_{s \wedge r} - \omega'_{s' \wedge r}|, \quad \forall (s, \omega), (s', \omega') \in \Lambda^t.$$

Denote by  $C_b^0(\Lambda^t)$  (resp.  $UC(\Lambda^t)$ ) the collection of functions in  $C^0(\Lambda^t)$  which are uniformly bounded (resp. uniformly continuous), and  $UC_b(\Lambda^t) := UC(\Lambda^t) \cap C_b^0(\Lambda^t)$ .

Next, denote by  $X^{0,t,\mathbf{x}}$  the solution of the SDE on  $(\Omega^t, \mathcal{F}_T^t, \mathbb{P}_0^t)$ :

$$X_s = \mathbf{x}_s, \quad \forall s \leq t \quad \text{and} \quad X_s = \mathbf{x}_t + \int_t^s \mu(r, X.) dr + \int_t^s \sigma(r, X.) dB_r^t, \quad \forall s > t. \quad (4.1)$$

Clearly,  $X^{0,t,\mathbf{x}}$  under  $\mathbb{P}_0^t$  has the same law as that of  ${}^{t,\mathbf{x}}X$  introduced in (2.2) under  $\mathbb{P}_0$ . We denote the induced measure on the shifted space  $\Omega^t$  by:

$$\mathbb{P}_{t,\mathbf{x}} := \mathbb{P}_0^t \circ (X^{0,t,\mathbf{x}} - \mathbf{x}_t)^{-1} \quad \text{and} \quad \mathbb{P}_X := \mathbb{P}_{0,\mathbf{0}}. \quad (4.2)$$

**Remark 4.1.** Let  $(t, \mathbf{x}) \in \Lambda^0$ ,  $\tau \geq t$  be a  $\mathbb{F}^t$ -stopping time on  $\Omega^t$ ,  $\xi \in \mathcal{F}_T^t$  and  $(\mathbb{P}_\omega)_{\omega \in \Omega}$  be a regular conditional probability distribution (r.c.p.d., see Stroock-Varadhan [18]) of  $\mathbb{P}_{t,\mathbf{x}}$  w.r.t.  $\mathcal{F}_\tau^0$ , then clearly,  $\mathbb{E}^{\mathbb{P}_\omega}[\xi] = \mathbb{E}^{\mathbb{P}_{\tau(\omega),\omega}}[\xi^{\tau(\omega),\omega}]$  for  $\mathbb{P}_{t,\mathbf{x}}$ -a.s.  $\omega \in \Omega$ .

For every  $s \in [0, T]$  and  $u : \Lambda^s \rightarrow \mathbb{R}$ , we introduce the Dupire [3] right time-derivative of  $u$  defined by the following limit, if exists,

$$\partial_t u(t, \omega) := \lim_{h \downarrow 0} \frac{u(t+h, \omega_{\cdot \wedge t}) - u(t, \omega)}{h}, \quad t < T, \quad \text{and} \quad \partial_t u(T, \omega) := \lim_{t < T, t \rightarrow T} \partial_t u(t, \omega).$$

**Definition 4.2.** Let  $u$  be a process  $C^0(\Lambda^t)$ . We say  $u \in C^{1,2}(\Lambda^t)$  if  $\partial_t u \in C^0(\Lambda^t)$  and there exist  $\partial_\omega u \in C^0(\Lambda^t, \mathbb{R}^d)$ ,  $\partial_{\omega\omega}^2 u \in C^0(\Lambda^t, \mathbb{S}^d)$  such that, for any  $(s, \omega) \in \Lambda^t$ :

$$du_r^{s,\omega} = (\partial_t u)_r^{s,\omega} dr + (\partial_\omega u)_r^{s,\omega} \cdot dB_r^s + \frac{1}{2} (\partial_{\omega\omega}^2 u)_r^{s,\omega} : d\langle B^s \rangle_r, \quad \mathbb{P}_{t,\mathbf{x}} - \text{a.s.} \quad (4.3)$$

If, in addition,  $u \in C_b^0(\Lambda^t)$ , we then say  $u \in C_b^{1,2}(\Lambda^t)$ .

It is clear, for  $s \leq t$ ,  $\omega \in \Omega^0$  and  $u \in C^{1,2}(\Lambda^s)$ , we have  $u^{t,\omega} \in C^{1,2}(\Lambda^t)$ .

Finally, for all  $t \in [0, T]$ , we denote by  $\mathcal{T}^t$  the collection of all  $\mathbb{F}^t$ -stopping times  $\tau$  such that  $\{\omega : \tau(\omega) > s\}$  is an open set in  $(\Omega^t, \|\cdot\|_T)$  for all  $s \in [t, T]$ , and by  $\mathcal{T}_+^t$  the collection of stopping times  $\tau \in \mathcal{T}^t$  such that  $\tau > t$ . The set  $\Lambda^t(\tau) := \{(t, \omega) \in \Lambda^t : t < \tau(\omega)\}$  is the corresponding localized canonical space, and we define similarly the spaces  $C^0(\Lambda^t(\tau))$ ,  $C^{1,2}(\Lambda^t(\tau))$ , etc...

#### 4.1.2 A path-dependent PDE

In this section, we do not need the restriction of the generator to have a power series representation in  $y$  as in (2.4). We then consider a slightly more general generator  $\widehat{F} : \Lambda^0 \times \mathbb{R} \rightarrow \mathbb{R}$  such that  $(t, \omega) \mapsto \widehat{F}(t, \omega, y)$  is  $\mathbb{F}^0$ -progressive for every  $y \in \mathbb{R}$ . Consider the second order path-dependent differential operator:

$$\mathcal{L}\varphi := \mu \cdot \partial_\omega \varphi + \frac{1}{2} \sigma \sigma^T : \partial_{\omega\omega}^2 \varphi. \quad (4.4)$$

Given a  $\mathcal{F}_T$ -measurable r.v.  $\xi : \Omega^0 \rightarrow \mathbb{R}$ , we consider the path-dependent PDE:

$$\{-\mathcal{L}u - \widehat{F}(\cdot, u)\}(t, \omega) = 0, \quad t < T, \quad \omega \in \Omega^0, \quad (4.5)$$

$$u(T, \omega) = \xi(\omega), \quad \omega \in \Omega^0. \quad (4.6)$$

**Assumption 4.3.** *There is a constant  $C$  such that  $\sup_{t \leq T} |\widehat{F}(t, \mathbf{0}, 0)| \leq C$ , and*

$$|\widehat{F}(t, \omega, y) - \widehat{F}(t, \omega', y')| + |\xi(\omega) - \xi(\omega')| \leq C(|y - y'| + \|\omega - \omega'\|_T),$$

for every  $t \in [0, T]$ ,  $(\omega, y), (\omega', y') \in \Omega^0 \times \mathbb{R}$ .

We next introduce two classes of test functions for every  $\mathbb{F}^0$ -adapted process  $u$ :

$$\begin{aligned} \underline{\mathcal{A}}u(t, \omega) &:= \left\{ \varphi \in C^{1,2}(\Lambda^t) : \exists \mathbb{H} \in \mathcal{T}_+^t, (\varphi - u^{t,\omega})_t(\mathbf{0}) = \min_{\tau \in \mathcal{T}_+^s} \mathbb{E}^{\mathbb{P}^{t,\omega}} [(\varphi - u^{t,\omega})_{\tau \wedge \mathbb{H}}] \right\}, \\ \overline{\mathcal{A}}u(t, \omega) &:= \left\{ \varphi \in C^{1,2}(\Lambda^t) : \exists \mathbb{H} \in \mathcal{T}_+^t, (\varphi - u^{t,\omega})_t(\mathbf{0}) = \max_{\tau \in \mathcal{T}^t} \mathbb{E}^{\mathbb{P}^{t,\omega}} [(\varphi - u^{t,\omega})_{\tau \wedge \mathbb{H}}] \right\}. \end{aligned}$$

The next definition requires the following notation for the path-dependent second order differential operator on the shifted canonical space:

$$(\mathcal{L}^{t,\omega}\varphi)(s, \omega') := (\mu^{t,\omega} \cdot \partial_x \varphi)(s, \omega') + \frac{1}{2}((\sigma \sigma^T)^{t,\omega} : \partial_{xx}^2 \varphi)(s, \omega'), \quad (s, \omega') \in \Lambda^t.$$

**Definition 4.4.** *Let  $u : \Lambda^0 \rightarrow \mathbb{R}$  be a locally bounded  $\mathbb{F}^0$ -progressive process.*

(i) *We say that  $u$  is a viscosity subsolution (resp. supersolution) of PPDE (4.5) if, for any  $(t, \omega) \in [0, T] \times \Omega^0$  and any  $\varphi \in \underline{\mathcal{A}}u(t, \omega)$  (resp.  $\varphi \in \overline{\mathcal{A}}u(t, \omega)$ ), it holds that*

$$\{-\partial_t \varphi - \mathcal{L}^{t,\omega} \varphi - \widehat{F}^{t,\omega}(\cdot, u^{t,\omega})\}(t, \mathbf{0}) \leq \quad (\text{resp. } \geq) \quad 0.$$

(ii) *We say that  $u$  is a viscosity solution of PPDE (4.5) if it is both a viscosity subsolution and a viscosity supersolution.*

**Remark 4.5.** (i) *In Definition 4.4, we can only use, without loss of generality, the test functions  $\varphi \in \underline{\mathcal{A}}$  (resp.  $\overline{\mathcal{A}}$ ) such that  $(\varphi - u^{t,\omega})_t(\mathbf{0}) = 0$ .*

(ii) *Similar to Remark 3.9 of Ekren, Keller, Touzi and Zhang [5], we can easily verify that under Assumption 4.3, for every  $\lambda \in \mathbb{R}$ ,  $u$  is a viscosity solution to (4.5) if and only if  $\tilde{u}(t, \omega) := e^{\lambda(T-t)}u(t, \omega)$  is a viscosity solution of*

$$-\partial_t \tilde{u} - \mathcal{L}\tilde{u} - \widehat{F}_\lambda(\cdot, \tilde{u}) = 0, \quad \text{where } \widehat{F}_\lambda(t, \omega, y) := -\lambda y + e^{\lambda t} \widehat{F}(t, \omega, e^{-\lambda t} y).$$

(iii) *Similar to Remark 2.11 of [5], we point out that in the Markovian setting, where the PPDE (4.5) decreases to a classical PDE, a viscosity solution in sense of Definition 4.4 is consistent to the viscosity solution in standard sense, by the uniqueness result proved below.*

### 4.1.3 The existence and uniqueness of solutions to PPDE

This section follows closely the arguments of [5, 6, 7]. However, their results do not apply to our context, because of the possible unboundedness of  $\mu$  and  $\sigma$ . Moreover,

the PPDE in our context is linear in the gradient and the Hessian components, which significantly simplifies the approach, see Remark 3.9 of [6].

The above viscosity solution to PPDE (4.5) is closely related to the following decoupled FBSDE. For every  $(t, \mathbf{x}) \in \Lambda^0$ , let  $X^{0,t,\mathbf{x}}$  be the solution of (4.1),  $(\widehat{Y}^{0,t,\mathbf{x}}, \widehat{Z}^{0,t,\mathbf{x}})$  be the solution of the BSDE on  $(\Omega^t, \mathcal{F}_T^t, \mathbb{P}_0^t)$ ,

$$\widehat{Y}_s = \xi(X^{0,t,\mathbf{x}}) + \int_s^T \widehat{F}(r, X^{0,t,\mathbf{x}}, \widehat{Y}_r) dr - \int_s^T \widehat{Z}_r \cdot dB_r^t. \quad (4.7)$$

By the Blumenthal 0-1 law,  $\widehat{Y}_t^{0,t,\mathbf{x}}$  is a constant and we then define

$$\hat{u}(t, \mathbf{x}) := \widehat{Y}_t^{0,t,\mathbf{x}}. \quad (4.8)$$

For later use, we observe that, since the diffusion matrix  $\sigma$  is nondegenerate, the above BSDE (4.7) is equivalent to the following BSDE on  $(\Omega^t, \mathcal{F}_T^t, \mathbb{P}_{t,\mathbf{x}})$ :

$$\tilde{Y}_s = \xi^{t,\mathbf{x}}(B^t) + \int_s^T \widehat{F}^{t,\mathbf{x}}(r, B^t, \tilde{Y}_r) dr - \int_s^T \tilde{Z}_r \cdot (dB_r^t - \mu^{t,\mathbf{x}}(r, B^t) dr),$$

whose solutions are denoted by  $(\tilde{Y}^{0,t,\mathbf{x}}, \tilde{Z}^{0,t,\mathbf{x}})$ , i.e.  $\widehat{Y}_t^{0,t,\mathbf{x}} = \tilde{Y}_t^{0,t,\mathbf{x}} = \hat{u}(t, \mathbf{x})$  for every  $(t, \mathbf{x}) \in \Lambda^0$ . Moreover, by equation (4.6) of [6], we have the dynamic programming principle

$$\begin{aligned} \tilde{Y}_s^{0,t,\mathbf{x}} &= \hat{u}^{t,\mathbf{x}}(\tau, B^t) + \int_s^\tau \widehat{F}^{t,\mathbf{x}}(r, B^t, \tilde{Y}_r^{0,t,\mathbf{x}}) dr \\ &\quad - \int_s^\tau \tilde{Z}_r^{0,t,\mathbf{x}} \cdot (dB_r^t - \mu^{t,\mathbf{x}}(r, B^t) dr), \quad \mathbb{P}_{t,\mathbf{x}} - \text{a.s.}, \end{aligned} \quad (4.9)$$

for all  $(t, \mathbf{x}) \in \Lambda^0$  and  $\tau \in \mathcal{T}^t$ .

**Theorem 4.6.** *Let Assumption 4.3 hold true.*

(i) *There is a constant  $C > 0$  such that  $\forall (t, \omega), (t', \omega') \in \Lambda^0$ ,*

$$|\hat{u}(t, \omega) - \hat{u}(t', \omega')| \leq C(\|\omega\|_t + \|\omega'\|_{t'}) (\sqrt{|t - t'|} + \|\omega_{t \wedge \cdot} - \omega'_{t' \wedge \cdot}\|_T).$$

(ii)  *$\hat{u}$  is a viscosity solution to PPDE (4.5).*

**Proof.** (i) is proved in Proposition 4.5 of Ekren et al. [6], since our BSDE (4.7) is a particular case to their equation (4.4). It is in fact an immediate consequence of Proposition 2.1 in El Karoui, Peng and Quenez [8] together with the estimation in our Lemma 2.1.

(ii) We adapt the arguments in the proof of Proposition 4.5 [6] to our context. We only show that  $\hat{u}$  is a viscosity subsolution. Assume  $\hat{u}$  is not a viscosity subsolution, then there exist  $(t, \omega) \in \Lambda^0$  and  $\varphi \in \underline{A}\hat{u}(t, \omega)$  such that  $c := \mathcal{L}^{t,\omega}\varphi(t, \mathbf{0}) > 0$ . Without loss of generality, we may also assume that  $\varphi(t, \mathbf{0}) = \hat{u}(t, \omega)$ . Denote, for  $s \in [t, T]$ ,

$$\tilde{Y}'_s := \varphi(s, B^t), \quad \tilde{Z}'_s := \partial_\omega \varphi(s, B^t), \quad \delta Y_s := \tilde{Y}'_s - \tilde{Y}_s^{0,t,\omega}, \quad \delta Z_s := \tilde{Z}'_s - \tilde{Z}_s^{0,t,\omega}.$$

Applying Itô's formula together with Definition 4.2, we have for some  $\mathbb{F}^t$ -progressive process  $\alpha$  bounded by some constant  $L_0$ ,

$$d(\delta Y_s) = - \left[ (\mathcal{L}^{t,\omega}\varphi)(s, B^t) + \alpha_s \delta Y_s \right] ds + \delta Z_s \cdot (dB_s^t - \mu^{t,\omega}(s, B^t) ds), \quad \mathbb{P}_{t,\omega} - \text{a.s.}$$

Observing that  $\delta Y_t = 0$ , we define

$$\mathbb{H} := T \wedge \inf \{s > t : \mathcal{L}^{t,\omega} \varphi(s, B^t) - L_0 |\delta Y_s| \leq c/2\}.$$

Then by the continuity of  $\delta Y$  as well as  $\mathcal{L}^{t,\omega} \varphi$ ,  $\mathbb{H} \in \mathcal{T}_+^t$  and

$$\mathcal{L}^{t,\omega} \varphi(s, B^t) + \alpha_s \delta Y_s \geq c/2, \quad \text{for all } s \in [t, \mathbb{H}].$$

Now for any  $\tau \in \mathcal{T}^t$  such that  $\tau \leq \mathbb{H}$ , we have

$$\begin{aligned} 0 = \delta Y_t &= \delta Y_\tau + \int_t^\tau \left[ (\mathcal{L}^{t,\omega} \varphi)(s, B^t) + \alpha_s \delta Y_s \right] - \int_t^\tau \delta Z_s \cdot (dB_s^t - \mu^{t,\omega}(s, B^t) ds) \\ &\geq \varphi(\tau, B^t) - \hat{u}^{t,\omega}(\tau, B^t) + c(\tau - t)/2 - \int_t^\tau \delta Z_s \cdot (dB_s^t - \mu^{t,\omega}(s, B^t) ds), \end{aligned}$$

$\mathbb{P}_{t,\omega}$  - a.s., where we also used the dynamic programming principle in (4.9). Then  $\mathbb{E}_t^{\mathbb{P}_{t,\omega}}[(\varphi - \hat{u}^{t,\omega})(\tau, B^t)] < 0$ , contradicting the fact that  $\varphi \in \underline{A}\hat{u}(t, \omega)$ .  $\square$

For the proof of our main result, we also need the following comparison result.

**Theorem 4.7.** *Let Assumption 4.3 hold true,  $u^1, u^2$  be two  $\mathbb{F}$ -progressive càdlàg processes on  $\Omega^0$  with corresponding jumps  $\Delta u^1 \geq 0 \geq \Delta u^2$ . Assume that  $u^1$  (resp.  $u^2$ ) is a viscosity subsolution (resp. supersolution) of PPDE (4.5), and  $u^1(T, \cdot) \leq \xi(\cdot) \leq u^2(T, \cdot)$ . Then  $u^1 \leq u^2$  on  $\Lambda^0$ .*

In preparation of the comparison principle, we first introduce two extended spaces  $\underline{C}_{t,\omega}^{1,2}(\Lambda^t)$  and  $\overline{C}_{t,\omega}^{1,2}(\Lambda^t)$  of  $C^{1,2}(\Lambda^0)$  and derive a partial comparison principle as in [6, 7].

**Definition 4.8.** *Let  $(t, \omega) \in \Lambda^0$ ,  $u : \Lambda^t \rightarrow \mathbb{R}$  be  $\mathbb{F}^t$ -adapted.*

- (i) *We say  $u \in \underline{C}_{t,\omega}^{1,2}(\Lambda^t)$  if there exist an increasing sequence of  $\mathbb{F}^t$ -stopping times in  $\mathcal{T}^t$ :  $t = \tau_0 \leq \tau_1 \leq \dots \leq T$  such that,*
  - a)  $\tau_i < \tau_{i+1}$  whenever  $\tau_i < T$ , and for all  $\tilde{\omega} \in \Omega^t$ , the set  $\{i : \tau_i(\tilde{\omega}) < T\}$  is finite;
  - b) For each  $i \geq 0$  and  $\tilde{\omega} \in \Omega^t$ ,  $\tau_{i+1}^{\tau_i(\tilde{\omega}), \tilde{\omega}} \in \mathcal{T}^{\tau_i(\tilde{\omega})}$  and  $u^{\tau_i(\tilde{\omega}), \tilde{\omega}} \in C_b^{1,2}(\Lambda^{\tau_i(\tilde{\omega})}(\tau_{i+1}^{\tau_i(\tilde{\omega}), \tilde{\omega}}))$ ;
  - c)  $u$  has non-negative jumps ( $\Delta u \geq 0$ ), and

$$\mathbb{E}^{\mathbb{P}_{t,\omega}} \left[ \sum_{i \geq 0} \int_{\tau_i}^{\tau_{i+1}} \left( |\partial_t u + \mathcal{L}^{t,\omega}|^2 + |\sigma^{t,\omega} \partial_\omega u|^2 \right) (s, B^t) ds \right] < \infty. \quad (4.10)$$

- (ii) *We say  $u \in \overline{C}_{t,\omega}^{1,2}(\Lambda^t)$  if  $-u \in \underline{C}_{t,\omega}^{1,2}(\Lambda^t)$ .*

**Lemma 4.9.** *Suppose that Assumption 4.3 holds true. Let  $u^1$  be a viscosity subsolution and  $u^2$  be a viscosity supersolution of PPDE (4.5) such that  $u^1(T, \cdot) \leq u^2(T, \cdot)$ . If  $u^1 \in \underline{C}_{0,0}^{1,2}(\Lambda^0)$  or  $u^2 \in \overline{C}_{0,0}^{1,2}(\Lambda^0)$ , then  $u^1 \leq u^2$  on  $\Lambda^0$ .*

**Proof.** We follow the lines of Proposition 4.1 of Ekren, Touzi and Zhang [7]. Suppose that  $u^1 \in \underline{C}_{0,0}^{1,2}(\Lambda^0)$ . First, let us show that, for every  $i \geq 0$  and  $\omega \in \Omega^0$ ,

$$(u^1 - u^2)_{\tau_i(\omega)}^+(\omega) \leq \mathbb{E}^{\mathbb{P}_{\tau_i(\omega), \omega}} \left[ \left( (u^1)_{\tau_{i+1}(\omega)}^{\tau_i(\omega), \omega} - (u^2)_{\tau_{i+1}(\omega)}^{\tau_i(\omega), \omega} \right)^+ \right]. \quad (4.11)$$

Without loss of generality, it is enough to consider the case  $i = 0$ , where  $\mathbb{P}_{\tau_0(\omega), \omega} = \mathbb{P}_X$  for all  $\omega \in \Omega$ . Assume to the contrary that

$$2Tc := (u_0^1 - u_0^2)^+ - \mathbb{E}^{\mathbb{P}^X}[(u_{\tau_1}^1 - u_{\tau_1}^2)^+] > 0,$$

we set

$$X_t := (u_t^1 - u_t^2)^+ + ct, \quad Y_t := \sup_{\tau \in \mathcal{T}^t} \mathbb{E}_t[X_{\tau \wedge \tau_1}], \quad \tau^* := \inf\{t > 0 : X_t = Y_t\} \leq \tau_1,$$

where  $\mathbb{E}_t[\zeta](\omega) := \mathbb{E}^{\mathbb{P}^{t, \omega}}[\zeta^{t, \omega}] = \mathbb{E}^{\mathbb{P}^X}[\zeta | \mathcal{F}_t](\omega)$ . In particular,  $\mathbb{E}_0[\cdot] = \mathbb{E}^{\mathbb{P}^X}[\cdot]$ . Then  $(Y_t)_{t \geq 0}$  is a supermartingale,  $(Y_{t \wedge \tau^*})_{t \geq 0}$  is a martingale and  $\tau^*$  is an optimal stopping time for the problem  $\sup_{\tau \in \mathcal{T}^0} \mathbb{E}_0[X_\tau]$ . It follows that

$$\mathbb{E}_0[X_{\tau^*}] = \mathbb{E}_0[Y_{\tau^*}] = Y_0 \geq X_0 = 2Tc + \mathbb{E}_t[(u_{\tau_1}^1 - u_{\tau_1}^2)^+] \geq Tc + \mathbb{E}_0[X_{\tau_1}].$$

Then there exists  $\omega^* \in \Omega^0$  such that  $t^* := \tau^*(\omega^*) < \tau_1$ . And therefore

$$(u^1 - u^2)^+(t^*, \omega^*) + ct^* = X_{t^*}(\omega^*) = Y_{t^*}(\omega^*) \geq \mathbb{E}_{t^*}[X_{\tau_1}] > ct^*,$$

which implies that  $0 < (u^1 - u^2)^+(t^*, \omega^*)$ . Set  $\varphi(t, \omega) := (u^1)^{t^*, \omega^*}(t, \omega) + c(t^*)$ . Then  $\varphi \in C^{1,2}(\Lambda^{t^*}(\tau_1))$  since  $u^1 \in C^{1,2}(\Lambda(\tau_1))$ . Moreover, let

$$\mathbb{H} := \inf\{t > t^* : u_t^1 - u_t^2 \leq 0\} \wedge \tau_1 \in \mathcal{T}_+^{t^*}.$$

Then for every  $\tau \in \mathcal{T}^{t^*}$ ,

$$\begin{aligned} (\varphi - (u^2)^{t^*, \omega^*})(t^*, \mathbf{0}) &= X_{t^*}(\omega^*) \geq \mathbb{E}_{t^*}[Y_{\tau \wedge \mathbb{H}}](\omega^*) \\ &\geq \mathbb{E}_{t^*}[X_{\tau \wedge \mathbb{H}}](\omega^*) = \mathbb{E}^{\mathbb{P}^{t^*, \omega^*}}[(\varphi - (u^2)^{t^*, \omega^*})_{\tau \wedge \mathbb{H}}], \end{aligned}$$

which implies that  $\varphi \in \underline{\mathcal{A}}u^2(t^*, \omega^*)$ . It follows that

$$0 \leq -\bar{\mathcal{L}}\varphi(t^*, \omega^*) \leq -c - \bar{\mathcal{L}}u^1(t^*, \omega^*),$$

which contradicts the fact that  $u^1$  is a subsolution and we hence prove (4.11). Further, since  $(\mathbb{P}_{\tau_i(\omega), \omega})_{\omega \in \Omega}$  induces a r.c.p.d. of  $\mathbb{P}_X$  w.r.t.  $\mathcal{F}_{\tau_i}$  (see Remark 4.1), it follows by (4.11) that for every  $i \geq 0$ ,

$$(u^1 - u^2)_0 \leq \mathbb{E}_0[(u^1 - u^2)_{\tau_i}^+].$$

By sending  $i \rightarrow \infty$ , we get that  $(u^1 - u^2)_0 \leq \mathbb{E}_0[(u^1 - u^2)_T^+] = 0$ , which completes the proof of  $u_0^1 \leq u_0^2$ .  $\square$

**Proof of Theorem 4.7.** We follow the lines of the proof of Theorem 7.4 of Ekren, Touzi and Zhang [6], where a comparison principle for PPDE (4.5) was proved in case  $\sigma \equiv I_d$ . In spirit of Remark 4.5, we suppose without loss of generality that  $\hat{F}$  decreases in  $y$ .

For every  $\varepsilon > 0$ , we denote

$$O_\varepsilon := \{x \in \mathbb{R}^d : |x| < \varepsilon\}, \quad \bar{O}_\varepsilon := \{x \in \mathbb{R}^d : |x| \leq \varepsilon\}, \quad \partial O_\varepsilon := \{x \in \mathbb{R}^d : |x| = \varepsilon\};$$

$$\mathcal{O}_t^\varepsilon := [t, T] \times O_\varepsilon, \quad \bar{\mathcal{O}}_t^\varepsilon := [t, T] \times \bar{O}_\varepsilon, \quad \partial\mathcal{O}_t^\varepsilon := ([t, T] \times \partial O_\varepsilon) \cup (\{T\} \times O_\varepsilon).$$

Let  $t_0 = 0$ ,  $x_0 = 0$ ,  $(t_i)_{i \geq 1}$  an increasing sequence in  $(0, T]$  with  $t_i = T$  when  $i$  is large enough, and  $(x_i)_{i \geq 1}$  a sequence in  $\mathbb{R}^d$ . Set  $\pi := (t_i, x_i)_{i \geq 0}$  and  $\pi_n := (t_i, x_i)_{0 \leq i \leq n}$ . Given  $\pi_n$  and  $(t, x) \in \mathcal{O}_{t_n}^\varepsilon$ , define

$$\mathbb{H}_0^{t,x,\varepsilon} := \inf\{s \geq t : |B^t + x| = \varepsilon\} \wedge T, \quad \mathbb{H}_{i+1}^{t,x,\varepsilon} := \inf\{s \geq \mathbb{H}_i^{t,x,\varepsilon} : |B_s^t - B_{\mathbb{H}_i^{t,x,\varepsilon}}^t| = \varepsilon\} \wedge T.$$

For  $t \in (t_n, T]$ , let  $\widehat{B}^{\varepsilon, \pi_n, t, x}(\omega)$  denote the linear interpolation of  $(t_i, \sum_{j=0}^i x_j)_{0 \leq i \leq n}$  and  $(\mathbb{H}_i^{t,x,\varepsilon}(\omega), \sum_{j=0}^n x_j + x + B_{\mathbb{H}_i^{t,x,\varepsilon}}^t(\omega))_{i \geq 0}$ . Define

$$\theta_n^\varepsilon(\pi_n; (t, x)) := \mathcal{Y}_t^{\varepsilon, \pi_n, t, x},$$

where, omitting the superscripts  $\varepsilon, \pi_n, t, x$ ,  $\mathcal{Y}$  is defined  $\mathbb{P}_{t, \widehat{B}}$ ,

$$\mathcal{Y}_s = \xi(\widehat{B}) + \int_s^T \widehat{F}\left(r, \sum_{i \geq -1} \widehat{B}_{\cdot \wedge \mathbb{H}_i^{t,x,\varepsilon}} \cdot 1_{[\mathbb{H}_i^{t,x,\varepsilon}, \mathbb{H}_{i+1}^{t,x,\varepsilon})} \mathcal{Y}_r\right) dr - \int_s^T \mathcal{Z}_r(dB_r - \mu(r, \widehat{B})dr),$$

with  $\mathbb{H}_{-1}^{t,x,\varepsilon} := t$ . Then clearly, for every  $n$  and  $\pi_n$ , the deterministic function  $\theta_n^\varepsilon := \theta_n^\varepsilon(\pi_n; \cdot)$  is the viscosity solution of the standard PDE on  $\mathcal{O}_{t_n}^\varepsilon$ :

$$-\partial_t \theta_n^\varepsilon - \mu(s, \widehat{\omega}^{\pi_n}) D \theta_n^\varepsilon - \frac{1}{2} \sigma \sigma^T(s, \widehat{\omega}^{\pi_n}) : D^2 \theta_n^\varepsilon - \widehat{F}(s, \widehat{\omega}^{\pi_n}, \theta_n^\varepsilon) = 0 \text{ on } \mathcal{O}_{t_n}^\varepsilon, \quad (4.12)$$

with terminal condition  $\theta_n^\varepsilon(\pi_n; t, x) = \theta_{n+1}^\varepsilon(\pi_n, (t, x); t, 0)$  on  $\partial\mathcal{O}_{t_n}^\varepsilon$ , where  $\widehat{\omega}^{\pi_n} := \widehat{B}_{\cdot \wedge t_n}^{\varepsilon, \pi_n, t, x}$  is deterministic, and  $\theta_n^\varepsilon(\pi_n; T, x) = \xi(\widehat{\omega}^{\pi_n})$  when  $t_n = T$ . Further, it follows from Proposition 7.2 of [6] that for every  $\delta > 0$ , there is  $\bar{\theta}_n^{\varepsilon, \delta} \in C^{1,2}(\mathcal{O}_{t_n}^\varepsilon)$  which is a classical supersolution of (4.12) such that  $\bar{\theta}_n^{\varepsilon, \delta}(\pi_n; t, x) \geq \theta_{n+1}^\varepsilon(\pi_n, (t, x); t, 0)$  on  $\partial\mathcal{O}_{t_n}^\varepsilon$  and  $|\bar{\theta}_n^{\varepsilon, \delta} - \theta_n^\varepsilon| \leq \delta$  on  $\bar{\mathcal{O}}_{t_n}^\varepsilon$ . Let  $\delta_n = \varepsilon/2^n$ ,  $\mathbb{H}_i^\varepsilon := \mathbb{H}_i^{0,0,\varepsilon}$ , and  $\widehat{B}^\varepsilon$  be the linear interpolation of  $(\mathbb{H}_i^\varepsilon, B_{\mathbb{H}_i^\varepsilon})_{i \geq 0}$ . Define

$$\psi^\varepsilon(t, \omega) := \sum_{n=0}^{\infty} \left( \delta_n + \bar{\theta}_n^{\varepsilon, \delta_n}((\mathbb{H}_i^\varepsilon, B_{\mathbb{H}_i^\varepsilon})_{0 \leq i \leq n}; t, B_t - B_{\mathbb{H}_n^\varepsilon}) \right) 1_{[\mathbb{H}_n^\varepsilon, \mathbb{H}_{n+1}^\varepsilon)},$$

and denote

$$\tilde{B}^\varepsilon := \sum_{i \geq -1} \widehat{B}_{\cdot \wedge \mathbb{H}_i^{t,x,\varepsilon}} \cdot 1_{[\mathbb{H}_i^{t,x,\varepsilon}, \mathbb{H}_{i+1}^{t,x,\varepsilon})}.$$

One can check straightforwardly that  $-\psi^\varepsilon$  satisfies the conditions of Definition 4.8 (c),  $\psi^\varepsilon(T, \omega) \geq \xi(\tilde{B}^\varepsilon)$ , and

$$-\partial_t \psi^\varepsilon - \mu(s, \tilde{B}^\varepsilon) \cdot \partial_\omega \psi^\varepsilon - \sigma(s, \tilde{B}^\varepsilon) : \partial_\omega \psi^\varepsilon - \widehat{F}(s, \tilde{B}^\varepsilon, \psi^\varepsilon(s, \tilde{B}^\varepsilon)) \geq 0. \quad (4.13)$$

Then  $\tilde{Y} := \psi^\varepsilon$ ,  $\tilde{Z} := \partial_\omega \psi^\varepsilon$  satisfy the BSDE

$$\tilde{Y}_s = \tilde{Y}_{\mathbb{H}_{i+1}} + \int_s^T \widehat{F}(r, \tilde{B}^\varepsilon, \tilde{Y}_r) dr - \int_s^T \tilde{Z}_r(dB_r - \mu(r, \tilde{B}^\varepsilon)dr), \quad \mathbb{P}_X - a.s.$$

on every interval  $[H_i, H_{i+1})$  such that  $\sup_{0 \leq t \leq T} |\mathcal{Y}_t - \tilde{Y}_t| \leq \varepsilon$ , which implies that (4.10) holds true for  $\psi^\varepsilon$  and hence  $\psi^\varepsilon \in \overline{C}_{0,0}^{1,2}(\Lambda^0)$ . Notice that  $\|\tilde{B}^\varepsilon - B\|_T \leq \varepsilon$ , then for some constant  $C$ ,

$$|\xi(\tilde{B}^\varepsilon) - \xi(B)| \leq C\varepsilon, \quad |\widehat{F}(s, \tilde{B}^\varepsilon, y) - \widehat{F}(s, B, y)| \leq C\varepsilon.$$

Set

$$\overline{\psi} := \psi^\varepsilon + 2C\varepsilon[1 + T - t],$$

one can verify that  $\overline{\psi} \in \overline{C}_{0,0}^{1,2}(\Lambda^0)$  is a viscosity supersolution of (4.5), and it follows by the partial comparison principle in Lemma 4.9 that  $u^1(0, \mathbf{0}) \leq \overline{\psi}(0, \mathbf{0})$ . Similarly, we can construct a viscosity subsolution  $\underline{\psi} \in \underline{C}_{0,0}^{1,2}(\Lambda^0)$  such that  $u^2(0, \mathbf{0}) \geq \underline{\psi}(0, \mathbf{0})$  and  $|\overline{\psi} - \underline{\psi}| \leq 4C[2 + T]\varepsilon$ . By sending  $\varepsilon \rightarrow 0$ , we conclude the proof.  $\square$

## 4.2 Proof of Theorem 2.12

Finally, we can complete the proof of our main result which gives a representation of BSDE by branching process.

**Proof of Theorem 2.12.** By Theorems 4.6 and 4.7, we only need to show that  $v$  is a viscosity solution of (4.5) with terminal condition  $\psi$  and generator  $F_{n_0}$  defined in (2.14) following Definition 4.4. We shall only show the subsolution part. Moreover, we recall that in the branching process, the process  ${}^{t,\mathbf{x}}X^{(1)}$  associated with the first particle is extended after its default time  $T_1$  by  ${}^{t,\mathbf{x}}X_s^{(1)} := \Phi^{t,\mathbf{x}}(s, B^{t,(1)})$  for all  $s \in [t, T]$ , where  $B^{t,(1)}$  is defined by  $B_{t+s}^{t,(1)} := W_s^1$  for all  $s \in [t, T]$ .

Suppose that  $v$  is not a viscosity subsolution of (4.5), then by Definition 4.4 and Remark 4.5, there is  $(t_0, \omega_0) \in \Lambda^0$  and  $\varphi \in \underline{\mathcal{A}}v(t_0, \omega_0)$  such that  $v(t_0, \omega_0) = \varphi(t_0, \mathbf{0})$  and

$$-\overline{\mathcal{L}}\varphi(t_0, \omega_0) = -\mathcal{L}\varphi(t_0, \omega_0) - \beta\mathcal{G}\varphi(t_0, \omega_0) = c > 0, \quad (4.14)$$

where  $\mathcal{L}$  is defined by (4.4) and

$$\mathcal{G}\varphi(t, \omega) := \sum_{k=0}^{n_0} a_k(t, \omega) \varphi^k(t, \omega) - \varphi(t, \omega).$$

Without loss of generality, we suppose that  $t_0 = 0$ . Then  ${}^{0, \omega_0}X^{(1)} = {}^{0, \mathbf{0}}X^{(1)}$ . Further, it follows by the continuity of functions  $\varphi$  and  $v$  in Proposition 2.11 that for every  $\varepsilon > 0$ , there is  $\mathbb{H} \in \mathcal{T}_+^0$  such that for every  $t \in [0, \bar{\tau}]$  (with  $\bar{\tau} := \mathbb{H}({}^{0, \mathbf{0}}X^{(1)})$ ),

$$|v(t, {}^{0, \mathbf{0}}X^{(1)}) - v(0, \mathbf{0})| + |\mathcal{G}\varphi(t, {}^{0, \mathbf{0}}X^{(1)}) - \mathcal{G}v(t, {}^{0, \mathbf{0}}X^{(1)})e^{-\beta t}| \leq \varepsilon.$$

Clearly,  $\bar{\tau}$  is a  $\overline{\mathbb{F}}^0$ -stopping time (see (2.11)) in probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Denote  $\mathbb{H}_h = \mathbb{H} \wedge h$  the  $\mathbb{F}^0$ -stopping time on  $(\Omega^0, \mathcal{F}^0, \mathbb{P}_0)$ ,  $\bar{\tau}_h := \bar{\tau} \wedge h$  and  $X_t := {}^{0, \mathbf{0}}X_t^{(1)}$  the



$\bar{\mathbb{F}}^0$ –stopping time and process on  $(\Omega, \mathcal{F}, \mathbb{P})$ , it follows from equation (3.4) of Lemma 3.4, together with (4.14), that

$$\begin{aligned}
& \mathbb{E}^{\mathbb{P}^x} \left[ \varphi(H_h, B^0) - v(H_h, B^0) \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[ \varphi(\bar{\tau}_h, X.) - v(\bar{\tau}_h, X.) \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[ \varphi(\bar{\tau}_h, X.) - \varphi(0, \mathbf{0}) + v(0, \mathbf{0}) - v(\bar{\tau}_h, X.) \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[ \varphi(\bar{\tau}_h, X.) - \varphi(0, \mathbf{0}) + \mathcal{G}v(T_1, X.) 1_{\bar{\tau}_h \geq T_1} \right] + \mathbb{E}^{\mathbb{P}} \left[ (v(T_1, X.) - v(\bar{\tau}_h, X.)) 1_{\bar{\tau}_h \geq T_1} \right] \\
&\leq -c \mathbb{E}^{\mathbb{P}}[\bar{\tau}_h] - \mathbb{E} \left[ \int_0^{\bar{\tau}_h} \left( \beta \mathcal{G}\varphi(t, X.) - \mathcal{G}v(t, X.) \beta e^{-\beta t} \right) dt \right] \\
&\quad + \mathbb{E}^{\mathbb{P}} \left[ (v(T_1, X.) - v(\bar{\tau}_h, X.)) 1_{\bar{\tau}_h \geq T_1} \right] \\
&\leq (-c + \varepsilon \beta + 2\varepsilon) \mathbb{E}^{\mathbb{P}}[\bar{\tau}_h] < 0
\end{aligned}$$

for  $\varepsilon$  small enough, which is in contradiction with Definition 4.4 of the viscosity sub-solution.  $\square$

## 5 Numerical examples

In this section, we provide a numerical illustration of our representation result, and the corresponding numerical implications. Consider the two following PDEs:

$$\partial_t v_1 + x \partial_a v_1 + \frac{1}{2} \sigma^2 x^2 \partial_{xx}^2 v_1 + \beta(v_1^2 - v_1) = 0, \quad v_1(T, x, a) = \psi(x, a) : \text{PDE1} \quad (5.1)$$

$$\partial_t v_2 + x \partial_a v_2 + \frac{1}{2} \sigma^2 x^2 \partial_{xx}^2 v_2 + \beta(-v_2^2 - v_2) = 0, \quad v_2(T, x, a) = \psi(x, a) : \text{PDE2} \quad (5.2)$$

with  $\psi(x, a) = (a - 1)^+$ . These PDEs correspond to the forward backward SDEs

$$dX_t = \sigma X_t dB_t, \quad X_0 = 1 \quad (5.3)$$

$$dA_t = X_t dt, \quad A_0 = 0 \quad (5.4)$$

$$dY_t = -\beta(F(Y_t) - Y_t) dt + Z_t dB_t, \quad Y_T = \psi(X_T, A_T) \quad (5.5)$$

with the non-linearities  $F_1(y) = y^2$  and  $F_2(y) = -y^2$ . In our numerical experiments, we have taken a diffusion coefficient  $\sigma = 0.2$  and a Poisson intensity  $\beta = 0.1$ , and the maturity  $T = 2$  or  $T = 5$  years. For  $T = 2$  years (resp. 5 years), the probability of default is around 0.18 (resp. 0.39).

In comparison with the KPP type PDE with  $F_1(y) = y^2$ , the replacement of the non-linearity  $y^2$  by  $-y^2$  has added the term  $(-)^{N_T-1}$  in the multiplicative functional (see Equation (2.12)), without changing the complexity of the branching diffusion algorithm. More precisely, we have:

$$v_1(0, X_0, A_0) = \mathbb{E}_{0,x} \left[ \prod_{i=1}^{N_T} \psi(X_T^i, A_T^i) \right], \quad (5.6)$$

$$v_2(0, X_0, A_0) = \mathbb{E}_{0,x} \left[ \prod_{i=1}^{N_T} (-1)^{N_T-1} \psi(X_T^i, A_T^i) \right]. \quad (5.7)$$

N	Fair(PDE1)	Stdev(PDE1)	Fair(PDE2)	Stdev(PDE2)	CPU (seconds)
12	5.69	0.16	5.36	0.16	0.1
14	5.61	0.08	5.23	0.08	0.6
16	5.50	0.04	5.15	0.04	1.5
18	5.52	0.02	5.16	0.02	5.9
20	5.53	0.01	5.16	0.01	23.6
22	5.54	0.00	5.17	0.01	94.1

Table 1: MC price quoted in percent as a function of the number of MC paths  $2^N$ . PDE pricer(PDE1) = **5.54**. PDE pricer(PDE2) = **5.17** (CPU PDE: 10 **seconds**). Maturity= 2 years. Non-linearities for PDE1 (resp. PDE2)  $F_1(u) = u^2$  (resp.  $F_2(u) = -u^2$ ). For completeness, the price with  $\beta = 0$  (which can be obtained using a classical Monte-Carlo pricer) is 6.52.

Our branching diffusion algorithm has been checked against a two-dimensional PDE solver with an ADI scheme (see Tables 1, 2). The degenerate PDEs have been converted into elliptic PDEs by introducing the process  $\tilde{A}_t = \int_0^t X_s ds + (T - t)X_t$ , satisfying  $d\tilde{A}_t = (T - t)dX_t$ . The computational experiments was done using a PC with 2.99 Ghz Intel Core 2 Duo CPU.

Note that our algorithm converges to the exact PDE result as expected and the error is properly indicated by the Monte-Carlo standard deviation estimator (see column Stdev). In order to illustrate the impact of the non-linearity  $F$  on the price  $v$ , we have indicated the price corresponding to  $\beta = 0$ .

We would like to highlight that replacing the generator  $\frac{1}{2}x^2\sigma^2\partial_x^2 + x\partial_a$  by a multi-dimensional operator  $\mathcal{L}$  can be easily handled in our framework by simulating the branching particles with a diffusion process associated to  $\mathcal{L}$ . This is out-of-reach with finite-difference scheme methods and not such an easy step for the BSDE approach which require computing conditional expectations.

## References

- [1] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [2] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Process. Appl., 111(2):175-206, 2004.
- [3] B. Dupire, *Functional Itô calculus*, ssrn, 2009, <http://ssrn.com/abstract=1435551>.
- [4] E. B. Dynkin, *Superdiffusions and positive solutions of nonlinear partial differential equations*, Appendix A by J.-F. Le Gall and Appendix B by I. E. Verbitsky. University Lecture Series, 34. American Mathematical Society, 2004.

N	Fair(PDE1)	Stdev(PDE1)	Fair(PDE2)	Stdev(PDE2)	CPU (seconds)
12	7.40	0.25	5.63	0.26	0.3
14	7.28	0.12	5.60	0.13	1.1
16	7.20	0.06	5.47	0.07	4.3
18	7.24	0.03	5.48	0.03	17.0
20	7.24	0.02	5.50	0.02	68.3
22	7.24	0.01	5.51	0.01	272.9

Table 2: MC price quoted in percent as a function of the number of MC paths  $2^N$ . PDE pricer(PDE1) = **7.24**. PDE pricer(PDE2) = **5.51** (CPU PDE: 25 **seconds**). Maturity= 5 years. Non-linearities for PDE1 (resp. PDE2)  $F_1(u) = u^2$  (resp.  $F_2(u) = -u^2$ ). For completeness, the price with  $\beta = 0$  (which can be obtained using a classical Monte-Carlo pricer) is 10.24.

- [5] I. Ekren, C. Keller, N. Touzi and J. Zhang, *On Viscosity Solutions of Path Dependent PDEs*, Anna. Proba. to appear.
- [6] I. Ekren, N. Touzi and J. Zhang, *Viscosity Solutions of Fully Nonlinear Path Dependent PDEs: Part I*, preprint.
- [7] I. Ekren, N. Touzi and J. Zhang, *Viscosity Solutions of Fully Nonlinear Path Dependent PDEs: Part II*, preprint.
- [8] N. El Karoui, S. Peng, M.C. Quenez, *Backward stochastic differential equations in finance*, Mathematical finance 7(1), 1-71, 1997.
- [9] A. M. Etheridge, *An Introduction to Superprocesses*, American Mathematical Society, 2000.
- [10] C. Graham, D. Talay, *Simulation Stochastique et Méthodes de Monte-Carlo*, Éditions de l'École Polytechnique, 2011.
- [11] P. Henry-Labordère, *Cutting CVA's Complexity*, Risk magazine (Jul 2012). Extended version: *Counterparty risk valuation: A marked branching diffusion approach*, ssrn, <http://ssrn.com/abstract=1995503>.
- [12] G. Kersting and F.C. Klebaner, *Sharp conditions for nonexplosions and explosions in Markov jump processes*, Anna. Proba. 23(1), 268-272, 1995.
- [13] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, Vol. 23, Springer, 1992.
- [14] H. P. McKean, *Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure Appl. Math., Vol 28, 323-331, 1975.
- [15] E. Pardoux and S. Peng, *Adapted solutions of backward stochastic differential equations*, System and Control Letters, 14, 55-61, 1990.
- [16] P.E. Protter, *Stochastic integration and differential equations*, Second edition. version 2.1, volume 21 of *Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin, 2005.

- [17] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, American Mathematical Society, Graduate Studies in Mathematics, Volume 140, 2012.
- [18] D. W. Stroock, S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer, 1979.
- [19] S. Watanabe, *On the branching process for Brownian particles with an absorbing boundary*, J. Math. Kyoto Univ. 4(2), 385-398, 1964.
- [20] J. Zhang, *A numerical scheme for backward stochastic differential equations*, Annals of Applied Probability, 14(1), 459-488, 2004.
- [21] A. Rasulov, G. Raimova, M. Mascagni, *Monte Carlo solution of Cauchy problem for a nonlinear parabolic equation*, Mathematics and Computers in Simulation, 80(6), 1118-1123, 2010.