



HAL
open science

Branching diffusion representation of semilinear PDEs and Monte Carlo approximation *

Pierre Henry-Labordère, Nadia Oudjane, Xiaolu Tan, Nizar Touzi, Xavier Warin

► To cite this version:

Pierre Henry-Labordère, Nadia Oudjane, Xiaolu Tan, Nizar Touzi, Xavier Warin. Branching diffusion representation of semilinear PDEs and Monte Carlo approximation *. 2017. <hal-01429549>

HAL Id: hal-01429549

<https://hal.science/hal-01429549v1>

Preprint submitted on 8 Jan 2017

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.



HAL Authorization

Branching diffusion representation of semilinear PDEs and Monte Carlo approximation *

Pierre Henry-Labordère[†] Nadia Oudjane[‡] Xiaolu Tan[§]
 Nizar Touzi[¶] Xavier Warin^{||}

March 8, 2016

Abstract

We provide a representation result of parabolic semi-linear PD-Es, with polynomial nonlinearity, by branching diffusion processes. We extend the classical representation for KPP equations, introduced by Skorokhod [23], Watanabe [27] and McKean [18], by allowing for polynomial nonlinearity in the pair (u, Du) , where u is the solution of the PDE with space gradient Du . Similar to the previous literature, our result requires a non-explosion condition which restrict to “small maturity” or “small nonlinearity” of the PDE. Our main ingredient is the automatic differentiation technique as in [15], based on the Malliavin integration by parts, which allows to account for the nonlinearities in the gradient. As a consequence, the particles of our branching diffusion are marked by the nature of the nonlinearity. This new representation has very important numerical implications as it is suitable for Monte Carlo simulation. Indeed, this provides the first numerical method for high dimensional nonlinear PDEs with error estimate induced by the dimension-free Central limit theorem. The complexity is also easily seen to be of the order of the squared dimension. The final section of this paper illustrates the efficiency of the algorithm by some high dimensional numerical experiments.

Key words. Semilinear PDEs, branching processes, Monte-Carlo methods.

*We are grateful to Vincent Bansaye, Julien Claisse, Emmanuel Gobet and Gaoyue Guo for valuable comments and suggestions. X. Tan and N. Touzi gratefully acknowledge the financial support of the ERC 321111 Rofirm, the ANR Isotace, and the Chairs Financial Risks (Risk Foundation, sponsored by Société Générale) and Finance and Sustainable Development (IEF sponsored by EDF and CA).

[†]Société Générale, Global Market Quantitative Research, pierre.henry-labordere@sgcib.com

[‡]EDF R&D & FiME, Laboratoire de Finance des Marchés de l’Energie, nadia.oudjane@edf.fr

[§]CEREMADE, University of Paris-Dauphine, PSL Research University, tan@ceremade.dauphine.fr

[¶]Ecole Polytechnique Paris, Centre de Mathématiques Appliquées, nizar.touzi@polytechnique.edu

^{||}EDF R&D & FiME, Laboratoire de Finance des Marchés de l’Energie, xavier.warin@edf.fr

1 Introduction

The objective of the present paper is to provide a probabilistic representation for the solution of a nonlinear parabolic second order partial differential equation (PDE) which is suitable for a high dimensional Monte Carlo approximating scheme. Our main results achieve this goal in the context of semilinear PDEs:

$$-\partial_t u - \mathcal{L}u = f(u, Du), \quad u_T = g, \quad t < T, \quad x \in \mathbb{R}^d,$$

with polynomial non-linearity $f_{t,x}(y, z)$ in the solution and its gradient, diffusion generator \mathcal{L} , and bounded terminal condition g .

Previous representation results were obtained in the literature by means of backward stochastic differential equations, as introduced by Pardoux and Peng [20]. The Monte Carlo numerical implications of this representation were introduced by Bally & Pagès [2], Bouchard & Touzi [5] and Zhang [28], and generated a large stream of the literature. However, these methods can be viewed as a Monte Carlo version of the finite elements methods, and as such, are subject to the problem of curse of dimensionality. Our primary goal is to avoid this numerical problem so as to be capable to handle genuinely high-dimensional problems. This however will be achieved at the cost of some limitations...

Our main representation result is obtained by using the branching diffusion trick to absorb the nonlinearity, as illustrated by Skorokhod [23], Watanabe [27] and McKean [18] in the context of the KPP equation, see also the extensions in Rasulov, Raimova & Mascagni [21] and our previous paper [14] where the representation is also shown to allow for path-dependency.

Since the gradient is also involved in the nonlinearity, our representation result is a significant improvement of the classically well-know representation of KPP equations. We observe that the polynomial nonlinearity naturally induces some restrictions needed to ensure the non-explosion of the corresponding solution. As a consequence, our representation holds under technical conditions of small maturity or small nonlinearity of the PDE.

The main idea for our representation is to use the Monte Carlo automatic differentiation technique in addition to the branching diffusion representation. The automatic differentiation in Monte Carlo approximation of diffusions was successfully used in the previous literature by Fournié et al. [11], Bouchard, Ekeland & Touzi [4], Henry-Labordère, Tan & Touzi [15], and Doumbia, Oudjane & Warin [7]. The resulting branching diffusion in the representation differs from that of the original founding papers [23, 27, 18] by introducing marks for the particles born at each branching. The mark of the particle determines the nature of the differentiation, and thus induces the corresponding automatic differentiation weight.

We next illustrate the main idea behind our representation in the context of the following extension of the one-dimensional Burgers equation:

$$\mathcal{L}u := \frac{1}{2}\Delta u, \quad \text{and} \quad f(y, z) = \frac{1}{2}(y^2 + yz).$$

Let W^1 be a Brownian motion, and τ^1 an independent random variable with density $\rho > 0$ on \mathbb{R}_+ , and denote $\bar{F}(t) := \int_t^\infty \rho(s) ds$. We also introduce another independent random variable I^1 which takes the values 0 and 1 with equal probability. Then, denoting by $\mathbb{E}_{t,x}$ the expectation operator conditional on the starting data $W_t = x$ at time t , we obtain from the Feynman-Kac formula the representation of the solution u as:

$$u(0, x) = \mathbb{E}_{0,x} \left[\bar{F}(T) \frac{g(W_T)}{\bar{F}(T)} + \int_0^T \frac{f(u, Du)(t, W_t)}{\rho(t)} \rho(t) dt \right] = \mathbb{E}_{0,x} [\phi(T_{(1)}, W_{T_{(1)}}^1)],$$

where $T_{(1)} := \tau^1 \wedge T$, and

$$\phi(t, y) := \frac{\mathbf{1}_{\{t \geq T\}}}{\bar{F}(T)} g(y) + \frac{\mathbf{1}_{\{t < T\}}}{\rho(t)} (uD^{I^1} u)(t, y). \quad (1.1)$$

We next consider the two alternative cases for the value of I^1 .

- On the event set $\{I^1 = 0\}$, it follows from the Markov property that:

$$(uD^{I^1} u)(t, y) = u(t, y)^2 = \mathbb{E}_{t,y} [\phi(t + \tau^1, W_{t+\tau^1}^1)]^2.$$

The tricky branching diffusion representation now pops up naturally by rewriting the last expression in terms of independent copies $(W^{1,1}, \tau^{1,1})$ and $(W^{1,2}, \tau^{1,2})$ as:

$$\begin{aligned} (uD^{I^1} u)(t, y) &= \mathbb{E}_{t,y} [\phi(t + \tau^{1,1}, W_{t+\tau^{1,1}}^{1,1})] \mathbb{E}_{t,y} [\phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,1})] \\ &= \mathbb{E}_{t,y} [\phi(t + \tau^{1,1}, W_{t+\tau^{1,1}}^{1,1}) \phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,2})]. \end{aligned}$$

where $\mathbb{E}_{t,y}$ denotes the expectation operator conditional on $W_t^{1,1} = W_t^{1,2} = y$. Substituting this expression in (1.1) and using the tower property, we see that the branching mechanism allows to absorb the nonlinearity.

- On the event set $\{I^1 = 1\}$, we arrive similarly to the expression

$$(uD^{I^1} u)(t, y) = \mathbb{E}_{t,y} [\phi(t + \tau^{1,1}, W_{t+\tau^{1,1}}^{1,1})] \partial_y \mathbb{E}_{t,y} [\phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,2})].$$

Our main representation is based on the following automatic differentiation:

$$\partial_y \mathbb{E}_{t,y} [\phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,2})] = \mathbb{E}_{t,y} \left[\frac{W_{t+\tau^{1,2}}^{1,2} - W_t^{1,2}}{\tau^{1,2}} \phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,1}) \right],$$

which is an immediate consequence of the differentiation with respect to the heat kernel, i.e. the marginal density of the Brownian motion. By the independence of $W^{1,1}$ and $W^{1,2}$, this provides:

$$(uD^{I^1} u)(t, y) = \mathbb{E}_{t,y} \left[\frac{W_{t+\tau^{1,2}}^{1,2} - W_t^{1,2}}{\tau^{1,2}} \phi(t + \tau^{1,1}, W_{t+\tau^{1,1}}^{1,1}) \phi(t + \tau^{1,2}, W_{t+\tau^{1,2}}^{1,2}) \right],$$

so that the branching mechanism allows again to absorb the nonlinearity by substituting in (1.1) and using the tower property.

The two previous cases are covered by denoting $T_{(1,i)} := T \wedge (\tau^1 + \tau^{1,i})$ for $i = 0, 1$, and introducing the random variable:

$$\mathcal{W}^1 := \mathbf{1}_{\{I^1=0\}} + \mathbf{1}_{\{I^1=1\}} \frac{\Delta W_{T_{(1,2)}}^{1,2}}{\Delta T_{(1,2)}}, \text{ with } \Delta W_{T_{(1,2)}}^{1,2} := W_{T_{(1,2)}}^{1,2} - W_{T_{(1)}}^{1,2}, \Delta T_{(1,2)} := T_{(1,2)} - T_{(1)},$$

so that

$$u(0, x) = \mathbb{E}_{0,x} \left[\mathbf{1}_{\{T_{(1)}=T\}} \frac{g(W_T)}{F(T)} + \mathbf{1}_{\{T_{(1)}<T\}} \frac{\mathcal{W}^1}{\rho(T_{(1)})} \prod_{i=1}^2 \left(\mathbf{1}_{\{T_{(1,i)}=T\}} \frac{g(W_T^{1,i})}{F(\Delta T_{(1,i)})} + \mathbf{1}_{\{T_{(1,i)}<T\}} \frac{(uD^{I^{1,i}}u)(T_{(1,i)}, W_{T_{(1,i)}}^{1,i})}{\rho(\Delta T_{(1,i)})} \right) \right].$$

Our main representation result is obtained by iterating the last procedure, and solving the integrability problems which arise because of the singularity introduced by the random variable \mathcal{W}^1 . The automatic differentiation, which is the main additional ingredient to the branching diffusion representation, is illustrated in the previous example when the operator \mathcal{L} corresponds to the Brownian motion. This extends to the case of a more general diffusion operator by the so-called Bismuth-Elworthy-Li formula based on the Malliavin integration by parts formula, see Fournié et al. [11] for its use in the context of Monte Carlo approximation and the extension to other sensitivities.

Our main result provides a probabilistic representation of the solution of the semi-linear PDE, with polynomial nonlinearity, in terms of a branching diffusion. This requires naturally a technical condition ensuring the existence of a non-exploding solution for the PDE which can be either interpreted as a small maturity or a small nonlinearity condition. This new representation provides a new ingredient for the analysis of the corresponding PDE as it can be used to argue about existence, uniqueness, and regularity. We shall indeed prove a C^1 -regularity result in order to prove the main Theorem 3.5.

Moreover, our new representation has an important numerical implication as it is suitable for high dimensional Monte Carlo approximation. This is in fact the first high dimensional general method for nonlinear PDEs ! The practical performance of the method is illustrated on a numerical example in dimension $d = 20$. The convergence of the numerical method is a direct consequence of the law of large numbers. The rate of convergence is also a direct consequence of the central limit theorem, and is therefore dimension-free. The complexity of the method is easily shown to be of the order of d^2 , which cannot be avoided by the very nature of the equation whose second order term involves $d \times d$ matrices calculations.

The paper is organized as follows. Section 2 introduces the marked branching diffusion. The main representation result is stated in Section 3. We next provide further discussions in Section 4 on the validity of our representation for systems of semilinear PDEs, and the possible combination with the unbiased simulation technique of [15, 7]. The Monte Carlo numerical implications of our representation in high dimension are reported in Section 5 with an illustration by a numerical experiment in dimension 20. Finally, we provide more numerical examples in Section 6.

2 The marked branching diffusion

2.1 Semilinear PDE with polynomial nonlinearity

Let $d \geq 1$, \mathbb{M}^d denotes the set of all $d \times d$ matrices, and $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$ the coefficient functions. For a function $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, we denote by Du and D^2u the gradient and the Hessian of the function $u(t, x)$ w.r.t. variable x . Let $m \geq 0$ be a positive integer, we consider a subset $L \subset \mathbb{N}^{m+1}$, and a sequence of functions $(c_\ell)_{\ell \in L}$ and $(b_i)_{i=1, \dots, m}$, where $c_\ell : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $b_i : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. For every $\ell = (\ell_0, \ell_1, \dots, \ell_m) \in L$, denote $|\ell| := \sum_{i=0}^m \ell_i$. A generator function $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$ is then defined by

$$f(t, x, y, z) := \sum_{\ell=(\ell_0, \ell_1, \dots, \ell_m) \in L} c_\ell(t, x) y^{\ell_0} \prod_{i=1}^m (b_i(t, x) \cdot z)^{\ell_i}. \quad (2.1)$$

Given two matrix $A, B \in \mathbb{M}^d$, denote $A : B := \text{Trace}(AB^\top)$. We will consider the following semilinear PDE:

$$\partial_t u + \mu \cdot Du + \frac{1}{2} \sigma \sigma^\top : D^2u + f(\cdot, u, Du) = 0, \text{ on } [0, T] \times \mathbb{R}^d, \text{ and } u(T, \cdot) = g, \quad (2.2)$$

for some bounded Lipschitz function $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

Remark 2.1. The nonlinearity (2.1) includes the simplest case of a source term. Indeed, for $\ell = (0, 0, \dots, 0)$, we have $c_\ell(t, x) y^{\ell_0} \prod_{i=1}^m (b_i(t, x) \cdot z)^{\ell_i} = c_\ell(t, x)$.

2.2 Age-dependent marked branching process

In preparation of the representation result, let us first introduce a branching process, characterized by a distribution density function $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a probability mass function $(p_\ell)_{\ell \in L}$ (i.e. $p_\ell \geq 0$ and $\sum_{\ell \in L} p_\ell = 1$).

Instead of the usual exponential arrival time, we shall consider a branching particle process with arrival time of distribution density function ρ . At the arrival time, the particle branches into $|\ell|$ offsprings with probability p_ℓ , among which, ℓ_i particles carry the mark i , $i = 0, \dots, m$. Then regardless of its mark, each descendant particle performs the same but independent branching process as the initial particle.

To construct the above process, we will consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with

- a sequence of i.i.d. positive random variables $(\tau^{m,n})_{m,n \geq 1}$ of density function ρ ,
- a sequence of i.i.d. random elements $(I^{m,n})_{m,n \geq 1}$ with $\mathbb{P}(I^{m,n} = \ell) = p_\ell$, $\ell \in L$.

In addition, the sequences $(\tau^{m,n})_{m,n \geq 1}$ and $(I^{m,n})_{m,n \geq 1}$ are independent.

We now construct an age-dependent branching process, with $(\tau^{m,n})_{m,n \geq 1}$ and $(I^{m,n})_{m,n \geq 1}$, using the following procedure.

1. We start from a particle marked by 0, indexed by (1), of generation 1, whose arrival time is given by $T_{(1)} := \tau^{1,1} \wedge T$.

2. Let $k = (k_1, \dots, k_{n-1}, k_n) \in \mathbb{N}^n$ be a particle of generation n , with arrival time T_k . When $T_k < T$, we let $I_k = I^{n, \pi_n(k)}$, where

$$\pi_n \text{ is an injection from } \mathbb{N}^n \text{ to } \mathbb{N},$$

and at time T_k , it branches into $|I_k|$ offspring particles, which constitute $n+1$ -the generation, and are indexed by (k_1, \dots, k_n, i) for $i = 1, \dots, |I_k|$.

3. When $I_k = (\hat{\ell}_0, \hat{\ell}_1, \dots, \hat{\ell}_m)$, we have $|\hat{\ell}|$ offspring particles, among which we mark the first $\hat{\ell}_0$ particles by 0, the next $\hat{\ell}_1$ particles by 1, and so on, so that each particle has a mark i for $i = 0, \dots, m$.
4. For a particle $k = (k_1, \dots, k_n, k_{n+1})$ of generation $n+1$, we denote by $k^- := (k_1, \dots, k_n)$ the ‘‘parent’’ particle of k , and the arrival time of k is given by $T_k := (T_{k^-} + \tau^{n+1, \pi_{n+1}(k)}) \wedge T$.
5. In particular, for a particle $k = (k_1, \dots, k_n)$ of generation n , and T_{k^-} is its birth time and also the arrival time of k^- . Moreover, for the initial particle $k = (1)$, one has $k^- = \emptyset$, and $T_\emptyset = 0$.

The above procedure defines a marked age-dependent branching process. We denote further

$$\theta_k := \text{mark of } k, \quad \mathcal{K}_t^n := \begin{cases} \{k \text{ of generation } n \text{ s.t. } T_{k^-} \leq t < T_k\}, & \text{when } t \in [0, T), \\ \{k \text{ of generation } n \text{ s.t. } T_k = T\}, & \text{when } t = T, \end{cases}$$

and also

$$\bar{\mathcal{K}}_t^n := \cup_{s \leq t} \mathcal{K}_s^n, \quad \mathcal{K}_t := \cup_{n \geq 1} \mathcal{K}_t^n \quad \text{and} \quad \bar{\mathcal{K}}_t := \cup_{n \geq 1} \bar{\mathcal{K}}_t^n.$$

Clearly, \mathcal{K}_t (resp. \mathcal{K}_t^n) denotes the set of all living particles (resp. of generation n) in the system at time t , and $\bar{\mathcal{K}}_t$ (resp. $\bar{\mathcal{K}}_t^n$) denotes the set of all particles (resp. of generation n) which have been alive before time t .

Example 2.2. *Let us consider the case $d = 1$, with*

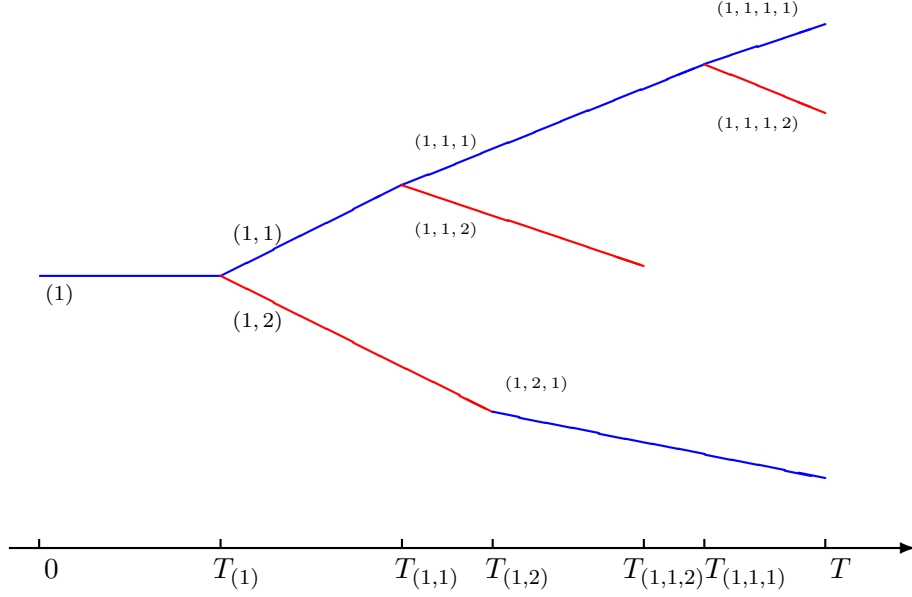
$$f(t, x, y, z) := c_{0,0}(t, x) + c_{1,0}(t, x)y + c_{1,1}(t, x)yz.$$

In this case, $m = 1$, $L = \{\bar{\ell}_1 = (1, 0), \bar{\ell}_2 = (1, 1)\}$. For the sake of clarity, we present an typical path of the associated age-dependent process, with graphical illustration below. The process starts from time 0 with one particle indexed by (1). At terminal time T , the number of particles alive is 3, with

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

- At time $T_{(1)}$, particle (1) branches into two particles (1, 1) and (1, 2).
- At time $T_{(1,1)}$, particle (1, 1) branches into (1, 1, 1) and (1, 1, 2).
- At time $T_{(1,2)}$, particle (1, 2) branches into (1, 2, 1).
- At time $T_{(1,1,2)}$, particle (1, 1, 2) dies out without any offspring particle.
- At time $T_{(1,1,1)}$, particle (1, 1, 1) branches into (1, 1, 1, 1) and (1, 1, 1, 2).

- The particles in blue are marked by 0, and the particles in red are marked by 1.



Proposition 2.3. *Assume that $\sum_{\ell \in L} |\ell| p_\ell < \infty$. Then the age-dependent branching process is well defined on $[0, T]$, i.e. the number of particles in $\bar{\mathcal{K}}_t$ is finite a.s. for all $t \in [0, T]$.*

Proof. See e.g. Theorem 1 of Athreya and Ney [1, Chapter IV.1], or Harris [12, pp. 138-139]. \square

2.3 The marked branching diffusion

We next equip each particle with a Brownian motion in order to define a branching Brownian motion.

We consider a sequence of independent d -dimensional Brownian motion $(W^{m,n})_{m,n \geq 1}$, which are also independent of $(\tau^{m,n}, I^{m,n})_{m,n \geq 1}$. Define $W_t^{(1)} = \Delta W_t^{(1)} := W_t^{1,1}$ for all $t \in [0, T_{(1)}]$ and then for each $k = (k_1, \dots, k_n) \in \bar{\mathcal{K}}_T \setminus \{(1)\}$, define

$$W_t^k := W_{T_{k-}}^{k-} + \Delta W_{t-T_{k-}}^k, \quad \text{with } \Delta W_{t-T_{k-}}^k := W_{t-T_{k-}}^{n, \pi_n(k)}, \quad \text{for all } t \in [T_{k-}, T_k]. \quad (2.3)$$

Then $(W^k)_{k \in \bar{\mathcal{K}}_T}$ is a branching Brownian motion. For each $k \in \bar{\mathcal{K}}_T$, we define an associated diffusion process $(X_t^k)_{t \in [T_{k-}, T_k]}$ by means of the following SDE

$$X_t^k = X_{T_{k-}}^{k-} + \int_{T_{k-}}^t \mu(s, X_s^k) ds + \int_{T_{k-}}^t \sigma(s, X_s^k) dW_s^k, \quad t \in [T_{k-}, T_k], \quad \mathbb{P}\text{-a.s.}, \quad (2.4)$$

where for particle (1), we fix the initial condition $X_0^{(1)} = x_0$ for some constant $x_0 \in \mathbb{R}^d$. The well-posedness of the last SDE is guaranteed by standard conditions on the coefficients μ, σ contained in Assumption 3.1.

The process $(X^k)_{k \in \bar{\mathcal{K}}_T}$ is our main marked branching diffusion process. We finally introduce the sub- σ -fields

$$\mathcal{F}_0 := \sigma\{\tau^{m,n}, I^{m,n} : m, n \geq 1\}, \quad \mathcal{F}_m := \sigma\{W^{i,n}, \tau^{i,n}, I^{i,n} : n \geq 1, i \leq m\}, \quad m \geq 1. \quad (2.5)$$

3 The main representation

We shall provide a representation result for a class the semilinear PDEs (2.2) under general abstract conditions. More explicit sufficient conditions are provided later.

3.1 Branching diffusion representation of semilinear PDEs

We first collect the conditions on the marked branching diffusion which are needed for our main results.

- Assumption 3.1.** (i) *The probability mass function $(p_\ell)_{\ell \in L}$ satisfies $p_\ell > 0$ for all $\ell \in L$, and $\sum_{\ell \in L} |\ell| p_\ell < \infty$. The density function $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is continuous and strictly positive on $[0, T]$, and such that $\bar{F}(T) := \int_T^\infty \rho(t) dt > 0$.*
(ii) $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$ are bounded continuous, and Lipschitz in x .
(iii) $c_\ell : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $b_i : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ are bounded continuous.

Our next assumption is the key automatic differentiation condition on the underlying diffusion $\bar{X}_s^{t,x}$ defined by

$$\bar{X}_s^{t,x} = x + \int_t^s \mu(r, \bar{X}_r^{t,x}) dr + \int_t^s \sigma(r, \bar{X}_r^{t,x}) dW_r, \quad s \in [t, T], \quad (3.1)$$

where W is a d -dimensional Brownian motion independent of the branching diffusion.

Assumption 3.2. *There is a measurable functional $\bar{W}(t, s, x, (W_r - W_t)_{r \in [t, s]})$ satisfying $(t, x) \mapsto \bar{W}(t, s, x, (W_r - W_t)_{r \in [t, s]})$ is continuous, and for any $s \in [t, T]$ and bounded measurable function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$, one has*

$$\partial_x \mathbb{E}[\phi(\bar{X}_s^{t,x})] = \mathbb{E}[\phi(\bar{X}_s^{t,x}) \bar{W}(t, s, x, (W_r - W_t)_{r \in [t, s]})].$$

Remark 3.3. *In case $(\mu, \sigma) \equiv (\mu_0, \sigma_0)$ for some constant $(\mu_0, \sigma_0) \in \mathbb{R}^d \times \mathbb{M}^d$, where σ_0 is not generate, then an example of such automatic differentiation function can be given by*

$$\bar{W}(t, s, x, (W_r - W_t)_{r \in [t, s]}) := (\sigma_0^\top)^{-1} \frac{W_s - W_t}{s - t}.$$

For general coefficient functions (μ, σ) satisfying some regularity and non-degeneracy conditions, one can find such functional \bar{W} using Malliavin calculus (see more discussions in Section 3.2).

Now, for each particle $k \in \bar{\mathcal{K}}_T$, we recall that it is born at time T_{k-} and dies out at time T_k , its mark is given by θ_k and its branching type is given by I_k . Let us denote

$$\mathcal{W}_k := \mathbf{1}_{\{\theta_k=0\}} + \mathbf{1}_{\{\theta_k \neq 0\}} b_{\theta_k}(T_{k-}, X_{T_{k-}}^k) \cdot \bar{W}(T_{k-}, T_k, X_{T_{k-}}^k, \Delta W^k). \quad (3.2)$$

We next introduce for a smooth function $u \in C^{1,2}([0, T] \times \mathbb{R}^d)$

$$\begin{aligned} \psi_n := & \left[\prod_{k \in \cup_{j=1}^n \mathcal{K}_T^j} \frac{g(X_T^k) - g(X_{T_{k-}}^k) \mathbf{1}_{\{\theta_k \neq 0\}} \mathcal{W}_k}{\bar{F}(\Delta T_k)} \right] \left[\prod_{k \in \cup_{j=1}^n (\bar{\mathcal{K}}_T^j \setminus \mathcal{K}_T^j)} \frac{c_{I_k}(T_k, X_{T_k}^k)}{p_{I_k}} \frac{\mathcal{W}_k}{\rho(\Delta T_k)} \right] \\ & \left[\prod_{k \in \bar{\mathcal{K}}_T^{n+1}} \left(\mathbf{1}_{\{\theta_k=0\}} u + \sum_{i=1}^m \mathbf{1}_{\{\theta_k=i\}} b_i \cdot Du \right) (T_{k-}, X_{T_{k-}}^k) \right], \end{aligned} \quad (3.3)$$

for all $n \geq 1$, and the corresponding limit

$$\psi := \left[\prod_{k \in \mathcal{K}_T} \frac{g(X_T^k) - g(X_{T_{k-}}^k) \mathbf{1}_{\{\theta_k \neq 0\}}}{\bar{F}(\Delta T_k)} \mathcal{W}_k \right] \left[\prod_{k \in \bar{\mathcal{K}}_T \setminus \mathcal{K}_T} \frac{c_{I_k}(T_k, X_{T_k}^k)}{p_{I_k}} \frac{\mathcal{W}_k}{\rho(\Delta T_k)} \right]. \quad (3.4)$$

Further, notice that the above branching diffusion process $(T_k, X^k)_{k \in \bar{\mathcal{K}}_T}$ and random variables ψ, ψ_n are defined with initial condition $(0, x_0)$ on interval $[0, T]$. By exactly the same way, we can define the system with initial condition (t, x) on interval $[t, T]$, let us denote them respectively by $(T_k^t, W^{t,k}, X^{t,x,k})_{k \in \bar{\mathcal{K}}_T^t}$, $\psi^{t,x}$, and $\psi_n^{t,x}$.

We now provide a first result, under strong regularity conditions, which provides a better understanding of our representation. We emphasize that our main representation result in Theorem 3.5 below will be established under more transparent conditions.

Proposition 3.4. *Let Assumptions 3.1 and 3.2 hold true. Suppose that the PDE (2.2) has a solution $u \in C^{1,2}([0, T] \times \mathbb{R}^d)$ with $\mathbb{E} \left[\int_t^T |f(\cdot, u, Du)(s, \bar{X}_s^{t,x})| ds \right] < \infty$, for some $(t, x) \in [0, T] \times \mathbb{R}^d$. Assume further that $(\psi_n^{t,x})_{n \geq 1}$ is uniformly integrable. Then*

$$\psi^{t,x} \in \mathbb{L}^1 \quad \text{and} \quad u(t, x) = \mathbb{E}[\psi^{t,x}].$$

Proof. (i) It suffices to consider $(t, x) = (0, x_0)$. Since g is bounded, it follows from the integrability condition on the process $f(\cdot, u, Du)$ and the Feynma-Kac formula that

$$\begin{aligned} u(0, x_0) &= \mathbb{E} \left[\frac{1}{\bar{F}(T)} g(\bar{X}_T^{0,x_0}) \bar{F}(T) + \int_0^T \frac{1}{\rho(s)} f(\cdot, u, Du)(s, \bar{X}_s^{0,x_0}) \rho(s) ds \right] \\ &= \mathbb{E} \left[\frac{1}{\bar{F}(T_{(1)})} g(X_{T_{(1)}}^{(1)}) \mathbf{1}_{\{T_{(1)}=T\}} + \right. \\ &\quad \left. \frac{1}{\rho(T_{(1)})} \left(\frac{c_{I_{(1)}}}{p_{I_{(1)}}} u^{I_{(1)},0} \prod_{i=1}^m (b_i \cdot Du)^{I_{(1),i}} \right) (T_{(1)}, X_{T_{(1)}}^{(1)}) \mathbf{1}_{\{T_{(1)} < T\}} \right] \quad (3.5) \\ &= \mathbb{E}[\psi_1]. \end{aligned}$$

(ii) Next, let $b_0 \in \mathbb{R}^d$ be a constant vector, and assume in addition that the random variable $\psi_1(b_0 \cdot \bar{W}(0, T_{(1)}, x_0, \Delta W^{(1)}))$ is integrable. Then under Assumptions 3.2,

$$\begin{aligned} b_0 \cdot D_x u(0, x_0) &= \mathbb{E} \left[\psi_1 b_0 \cdot \bar{W}(0, T_{(1)}, x_0, \Delta W^{(1)}) \right] \\ &= \mathbb{E} \left[\left(\psi_1 - \frac{1}{\bar{F}(T_{(1)})} g(x_0) \mathbf{1}_{\{T_{(1)}=T\}} \right) b_0 \cdot \bar{W}(0, T_{(1)}, x_0, \Delta W^{(1)}) \right], \quad (3.6) \end{aligned}$$

where the first equality follows by Lemma A.3 of [15] and the second equality follows from the fact that $\mathbb{E}[\bar{W}(0, T, x, (W_s)_{s \in [0, T]})] = 0$.

(iii) For $k \in \bar{\mathcal{K}}_T^2$, change the initial condition from $(0, x_0)$ to $(T_{k-}, X_{T_{k-}}^k) = (T_{(1)}, X_{T_{(1)}}^{(1)})$ in formula (3.5) and (3.6). Then, with \mathcal{F}_1 defined in (2.5),

$$u(T_{k-}, X_{T_{k-}}^k) = \mathbb{E} \left[\mathbf{1}_{\{k \in \mathcal{K}_T^2\}} \frac{g(X_T^k)}{\bar{F}(\Delta T_k)} + \mathbf{1}_{\{k \in \bar{\mathcal{K}}_T^2 \setminus \mathcal{K}_T^2\}} \Psi_k \mid \mathcal{F}_1 \right],$$

by the Markov property, and by Assumption 3.2,

$$Du(T_{k-}, X_{T_{k-}}^k) = \mathbb{E} \left[\left(\mathbf{1}_{\{k \in \mathcal{K}_T^2\}} \frac{g(X_T^k) - g(X_{T_{k-}}^k)}{\bar{F}(\Delta T_k)} + \mathbf{1}_{\{k \in \bar{\mathcal{K}}_T^2 \setminus \mathcal{K}_T^2\}} \Psi_k \right) \bar{W}(T_{k-}, T_k, X_{T_{k-}}^k, \Delta W^k) \mid \mathcal{F}_1 \right],$$

where $\Psi_k := \frac{1}{\rho(\Delta T_k)} \frac{c_{I_k}(T_k, X_{T_k}^k)}{p_{I_k}} \left[\prod_{k'=-k} \left(\mathbf{1}_{\{\theta_{k'}=0\}} u + \sum_{i=1}^m \mathbf{1}_{\{\theta_{k'}=i\}} b_i \cdot Du \right) (T_{k'-}, X_{T_{k'-}}^{k'}) \right]$. Plugging these expressions in the definition of ψ_1 in (3.3), it follows from the integrability of ψ_2 and the tower property of conditional expectations that $u(0, x_0) = \mathbb{E}[\psi_2]$. (iv) Iterating this procedure, we see that

$$u(0, x_0) = \mathbb{E}[\psi_n], \quad \text{for all } n \geq 1, \quad \text{and therefore } u(0, x_0) = \lim_{n \rightarrow \infty} \mathbb{E}[\psi_n] = \mathbb{E}[\psi],$$

where the last equality follows by the uniform integrability condition of $(\psi_n)_{n \geq 1}$. \square

We now state our main representation result under abstract conditions on the automatic differentiation weight function $\bar{W}(\cdot)$ involving the slight modification of ψ :

$$\tilde{\psi} := \left[\prod_{k \in \mathcal{K}_T} \frac{g(X_T^k) - g(X_{T_{k-}}^k) \mathbf{1}_{\{\theta_k \neq 0 \text{ or } k=(1)\}}}{\bar{F}(\Delta T_k)} \mathcal{W}_k \right] \left[\prod_{k \in \bar{\mathcal{K}}_T \setminus \mathcal{K}_T} \frac{c_{I_k}(T_k, X_{T_k}^k)}{p_{I_k}} \frac{\mathcal{W}_k}{\rho(\Delta T_k)} \right], \quad (3.7)$$

with $\tilde{\psi}^{t,x}$ defined by an obvious change of origin. Explicit sufficient conditions for the validity of the next result will be reported in Section 3.2 below.

Theorem 3.5. *Let Assumptions 3.1 and 3.2 hold true, and suppose in addition that for all $(t, x) \in [0, T] \times \mathbb{R}^d$, there is some $\varepsilon > 0$ such that*

$$(\psi^{s,y})_{(s,y) \in B_\varepsilon(t,x)} \quad \text{and} \quad (\tilde{\psi}^{s,y} \bar{W}(s, T_{(1)}^s, y, \Delta W^{s,(1)}))_{(s,y) \in B_\varepsilon(t,x)}$$

are uniformly integrable, where $B_\varepsilon(t, x) := \{(s, y) \in [0, T] \times \mathbb{R}^d : |s - t| + |x - y| \leq \varepsilon\}$. Then, the function $u(t, x) := \mathbb{E}[\psi^{t,x}]$ is a continuous viscosity solution of the semilinear PDE (2.2). Moreover Du exists and is continuous.

Proof. (i) Notice that the solution of SDE (3.1) is continuous w.r.t. its initial condition (t, x) , and recall that $(t, x) \mapsto \bar{W}(t, s, x, (W_r - W_t)_{r \in [t,s]})$ is also continuous, then under the uniform integrability condition on $(\psi^{t,x})$, one obtains that $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is continuous. Similarly, let us define

$$v_i(t, x) := \mathbb{E}[\tilde{\psi}^{t,x} b_i(t, x) \cdot \bar{W}(t, T_{(1)}^t, x, \Delta W^{t,(1)})] = \mathbb{E}[\psi^{t,x} b_i(t, x) \cdot \bar{W}(t, T_{(1)}^t, x, \Delta W^{t,(1)})],$$

which is also continuous by the uniformly integrability condition.

(ii) Let us define $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$\begin{aligned} \phi(I_{(1)}, T_{(1)}, X_{T_{(1)}}^{(1)}) &:= \frac{1}{\bar{F}(T_{(1)})} g(X_T^{(1)}) \mathbf{1}_{\{T_{(1)}=T\}} \\ &+ \frac{1}{\rho(T_{(1)})} \left(\frac{c_{I_{(1)}}}{p_{I_{(1)}}} (\psi^{T_{(1)}, X_{T_{(1)}}^{(1)}})^{I_{(1)},0} \prod_{i=1}^m v_i(T_{(1)}, X_{T_{(1)}}^{(1)})^{I_{(1)},i} \right) \mathbf{1}_{\{T_{(1)} < T\}} \\ &= \mathbb{E}[\psi \mid \mathcal{F}_1], \end{aligned} \quad (3.8)$$

where \mathcal{F}_1 is defined in (2.5). Notice that $\phi(i, t, x)$ is continuous in x , then it follows by Assumption 3.2 and Lemma A.3 of [15] that

$$Du(0, x_0) = \mathbb{E}[\phi(I_{(1)}, T_{(1)}, X_{T_{(1)}}^{(1)}) \overline{\mathcal{W}}(0, T_{(1)}, x_0, \Delta W^{(1)})] = \mathbb{E}[\psi \overline{\mathcal{W}}(0, T_{(1)}, x_0, \Delta W^{(1)})].$$

By changing the initial condition from $(0, x_0)$ to (t, x) and notice that

$$\mathbb{E}[\overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W^{t, (1)}) \mathbf{1}_{\{T_{(1)}^t = T\}}] = 0,$$

it follows that

$$Du(t, x) = \mathbb{E}[\tilde{\psi}^{t, x} \overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W^{t, (1)})],$$

and one obtains that $Du : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuous from the uniform integrability of $(\tilde{\psi}^{t, x} \overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W^{t, (1)}))$. Moreover, one has $v_i(t, x) = b_i(t, x) \cdot Du(t, x)$.

(iii) Using the expression in (3.8) and the law of $I_{(1)}$ and $T_{(1)}$, and with similar arguments as in (3.5), it follows that

$$u(t, x) = \mathbb{E}[\psi^{t, x}] = \mathbb{E}\left[g(\overline{X}_T^{t, x}) + \int_t^T f(\cdot, u, Du)(s, \overline{X}_s^{t, x}) ds\right].$$

Let $h > 0$, denote $H_h := (t+h) \wedge \inf\{s > t : |\overline{X}_s^{t, x} - x| \geq 1\}$, then by the flow property of $\overline{X}^{t, x}$, one has

$$u(t, x) = \mathbb{E}\left[u(H_h, \overline{X}_{H_h}^{t, x}) + \int_t^{H_h} f(\cdot, u, Du)(s, \overline{X}_s^{t, x}) ds\right],$$

and we may verify by standard arguments that u is a viscosity solution of PDE (2.2). \square

3.2 More explicit sufficient conditions

We now provide some explicit sufficient conditions which guarantee the validity of the conditions of Theorem 3.5. Define $|\varphi|_\infty := \sup_{x \in \mathbb{R}^d} |\varphi(x)|$ for any bounded function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, and $|\phi|_\infty := \sup_{i=1}^d |\phi_i|_\infty$ for any bounded vector function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$.

We first recall the Bismut-Elworthy-Li formula from Malliavin calculus, which was used by Fournié, Lasry, Lebuchoux, Lions and Touzi [11] as an automatic differentiation tool, see also [4], [5] and [10] for subsequent usefulness of the automatic differentiation in the context of the Monte Carlo approximation of nonlinear PDEs. We emphasize that such automatic differentiation function is not unique.

Assumption 3.6. *The coefficients μ, σ are bounded continuous, with bounded continuous partial gradients $D\mu, D\sigma$, and σ is uniformly elliptic.*

Notice that $(\overline{X}_s^{t, x})_{s \in [t, T]}$, as defined by (3.1), is completely determined by $(t, x, (W_s - W_t)_{s \in [t, T]})$. We then introduce the corresponding first variation process Y :

$$Y_t := \mathbf{I}_d, \quad dY_s = D\mu(s, \overline{X}_s^{t, x}) Y_s ds + \sum_{i=1}^d D\sigma_i(s, \overline{X}_s^{t, x}) Y_s dW_s^i, \quad \text{for } s \in [t, T], \quad \mathbb{P}\text{-a.s.}, \quad (3.9)$$

where \mathbf{I}_d denotes the $d \times d$ identity matrix, and $\sigma_i(t, x) \in \mathbb{R}^d$ denotes the i -th column of matrix $\sigma(t, x)$. Then one has the following result (see e.g. Exercise 2.3.5 of Nualart [19, p.p. 125], or Proposition 3.2. of [11]).

Proposition 3.7. *Let Assumption 3.6 hold true, then Assumption 3.2 holds true with the choice of automatic differentiation function \overline{W} defined by*

$$\overline{W}(t, s, x, (W_r - W_t)_{r \in [t, s]}) := \frac{1}{s-t} \int_t^s [\sigma^{-1}(r, \overline{X}_r^{t,x}) Y_r]^\top dW_r. \quad (3.10)$$

Remark 3.8. *When $\mu \equiv 0$ and $\sigma(t, x) \equiv \sigma_0$ for some non-degenerate constant matrix $\sigma_0 \in \mathbb{M}^d$, one then has $Y_t \equiv \mathbf{I}_d$ and so that*

$$\overline{W}(t, s, x, (W_r - W_t)_{r \in [t, s]}) = (\sigma_0^\top)^{-1} \frac{W_s - W_t}{s-t}.$$

With the above choice of automatic differentiation weight function (3.10), we can now derive some upper bounds for random variables $(\psi_n, n \geq 1)$. Recall that L_g is the Lipschitz constant of g , denote by $B_0^\infty(L_g) := \{(x_1, \dots, x_d) \in \mathbb{R}^d : |x_i| \leq L_g\}$ and $\overline{W}_{t,x,s} := \overline{W}(t, s, x, (W_r - W_t)_{r \in [t, s]})$. Then for $n \geq 1, q > 1$, we introduce two constants $C_{1,q}$ and $C_{2,q}$ by

$$C_{1,q} := |g|_\infty^q \vee \sup_{0 \leq t < s \leq T, x \in \mathbb{R}^d, i=1, \dots, m, b_0 \in B_0^\infty(L_g)} \mathbb{E} \left[\left| (b_0 \cdot (\overline{X}_s^{t,x} - x)) (b_i(t, x) \cdot \overline{W}_{t,x,s}) \right|^q \right]$$

and

$$C_{2,q} := \sup_{0 \leq t < s \leq T, x \in \mathbb{R}^d, i=1, \dots, m} \mathbb{E} \left[\left| \sqrt{s-t} b_i(t, x) \cdot \overline{W}_{t,x,s} \right|^q \right],$$

and then

$$\widehat{C}_{1,q} := \frac{C_{1,q}}{F(T)^{q-1}}, \quad \widehat{C}_{2,q} := C_{2,q} \sup_{\ell \in L, t \in (0, T]} \left(\frac{|c_\ell|_\infty t^{-\frac{q}{2(q-1)}}}{p_\ell \rho(t)} \right)^{q-1}.$$

Remark 3.9. (i) *Under Assumption 3.6, the tangent process Y is defined by a linear SDE, which has finite moment of any order $q \geq 1$. Then the two constant $C_{1,q}$ and $C_{2,q}$ are both finite. And for all $k \in \overline{\mathcal{K}}_T$, one has*

$$\max \left\{ |g|_\infty^q, \mathbb{E} \left[\left| (D_g \cdot \Delta X_k) \mathcal{W}_k \right|^q \middle| \mathcal{F}_0 \right] \right\} \leq C_{1,q}, \quad \mathbb{E} \left[\left(\sqrt{\Delta T_k} |\mathcal{W}_k| \right)^q \middle| \mathcal{F}_0 \right] \leq C_{2,q}, \quad (3.11)$$

where the sub- σ -field \mathcal{F}_0 is defined in (2.5).

(ii) *Notice that for a random variable $N \sim N(0, 1)$ and non-negative integer $q \geq 0$, one has $\mathbb{E}[|N|^q] = 2^{\frac{q}{2}} \Gamma(\frac{q+1}{2}) / \sqrt{\pi}$. Then if $(\mu, \sigma) \equiv (0, \sigma_0)$, for some constant $(\mu_0, \sigma_0) \in \mathbb{R}^d \times \mathbb{M}^d$, and \overline{W} as in Remark 3.8, it follows by direct computation that*

$$C_{1,q} \leq |g|_\infty^q \vee \left(\sup_{b_0 \in B_0^\infty(L_g)} (b_0^\top \sigma_0 \sigma_0^\top b_0) + \max_{i=1, \dots, m} \|b_i^\top (\sigma_0 \sigma_0^\top)^{-1} b_i\|_\infty \right)^q 2^{q-1} \Gamma\left(\frac{2q+1}{2}\right) / \sqrt{\pi},$$

and

$$C_{2,q} = \max_{i=1, \dots, m} \|b_i^\top (\sigma_0 \sigma_0^\top)^{-1} b_i\|_\infty^{\frac{q}{2}} 2^{\frac{q}{2}} \Gamma\left(\frac{q+1}{2}\right) / \sqrt{\pi}.$$

We are now ready for the main explicit sufficient conditions for the validity of the representation Theorem 3.5. Notice that the following conditions can be interpreted either as a small maturity or small nonlinearity restriction.

Assumption 3.10. For some $q > 1$, one of the following two items holds true.

- (i) Both $C_{1,q} \left(\frac{1}{\bar{F}(T)} \right)^q$ and $\sup_{\ell \in L, t \in (0, T]} C_{2,q} \left(\frac{|c_\ell|_\infty}{p_\ell} \frac{1}{\sqrt{t} \rho(t)} \right)^q$ are bounded by 1.
- (ii) $T < \int_{\widehat{\mathcal{C}}_{1,q}}^\infty (\widehat{\mathcal{C}}_{2,q} \sum_{\ell \in L} |c_\ell|_\infty x^{|\ell|})^{-1} dx$.

Remark 3.11. (i) To ensure that $\sup_{\ell \in L, t \in (0, T]} C_{2,q} \left(\frac{|c_\ell|_\infty}{p_\ell} \frac{1}{\sqrt{t} \rho(t)} \right)^q$ is bounded by 1, it is necessary to choose $(p_\ell)_{\ell \in L}$ such that $\frac{|c_\ell|_\infty}{p_\ell}$ is uniformly bounded, and to choose a density function such that $\rho(t) \geq Ct^{-1/2}$.

(ii) To ensure that $\widehat{\mathcal{C}}_{2,q}$ is finite, one needs to choose the density function ρ such that $\rho(t) \geq Ct^{-\frac{q}{2(q-1)}}$, and hence it is necessary that $q \in (2, \infty)$ so that $\frac{q}{2(q-1)} \in (\frac{1}{2}, 1)$.

Theorem 3.12. Consider the automatic differentiation function (3.10), and suppose that Assumptions 3.1, 3.6 and 3.10 hold true.

- (i) Then Assumption 3.2 holds, and $(\psi^{t,x}, \tilde{\psi}^{t,x} \overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W_{(1)}^t))_{(t,x) \in [0, T] \times \mathbb{R}^d}$ is uniformly integrable. Consequently, $u(t, x) := \mathbb{E}[\psi^{t,x}]$ is a viscosity solution of PDE (2.2).
- (ii) If Assumption 3.10 holds with some $q \geq 2$, then $\mathbb{E}[|\psi^{t,x}|^2] < \infty$.

Proof. (i) First, using Proposition 3.7, it is clear that Assumption 3.2 holds true with the choice of automatic differentiation function in (3.10).

(ii) Next, for $q \geq 1$, let us introduce

$$\chi_\infty^q := \left[\prod_{k \in \mathcal{K}_T} C_{1,q} \left(\frac{1}{\bar{F}(\Delta T_k)} \right)^q \right] \left[\prod_{k \in \overline{\mathcal{K}}_T \setminus \mathcal{K}_T} C_{2,q} \left(\frac{|c_{I_k}|_\infty}{p_{I_k}} \frac{1}{\sqrt{\Delta T_k} \rho(\Delta T_k)} \right)^q \right].$$

By conditioning on \mathcal{F}_0 , it follows from (3.11), together with direct computation, that

$$\mathbb{E}[|\psi|^q] \leq \mathbb{E}[\chi_\infty^q] \quad \text{and} \quad \mathbb{E}[|\tilde{\psi} \overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W_{(1)}^t)|^q] \leq C \mathbb{E}[\chi_\infty^q], \quad (3.12)$$

for some constant depending only on the Lipschitz constant L_g .

(iii) When Assumption 3.10 (i) holds true for some $q > 1$, then it is clear that $\mathbb{E}[|\psi|^q] \leq 1$. Notice that the above argument is independent of the initial condition $(0, x_0)$, it follows that $(\psi^{t,x}, \tilde{\psi}^{t,x} \overline{\mathcal{W}}(t, T_{(1)}^t, x, \Delta W_{(1)}^t))_{(t,x) \in [0, T] \times \mathbb{R}^d}$ is uniformly integrable.

(iv) When Assumption 3.10 (ii) holds true for some $q > 1$. Consider the ODE on $[0, T]$:

$$\eta(T) = \widehat{\mathcal{C}}_{1,q}, \quad \eta'(t) + \sum_{\ell \in L} \widehat{\mathcal{C}}_{2,q} \|c_\ell\|_\infty \eta(t)^{|\ell|} = 0.$$

Under Assumption 3.10 (ii), it is clear that the above ODE admits a unique finite solution on $[0, T]$. We next introduce a sequence of random variables

$$\widehat{\chi}_n^q := \left[\prod_{k \in \cup_{j=1}^n \mathcal{K}_T^j} \frac{\widehat{\mathcal{C}}_{1,q}}{\bar{F}(\Delta T_k)} \right] \left[\prod_{k \in \cup_{j=1}^n (\overline{\mathcal{K}}_T^j \setminus \mathcal{K}_T^j)} \widehat{\mathcal{C}}_{2,q} \frac{|c_{I_k}|_\infty}{p_{I_k}} \frac{1}{\rho(\Delta T_k)} \right] \left[\prod_{k \in \overline{\mathcal{K}}_T^{n+1}} \eta(T_{k-}) \right],$$

and

$$\widehat{\chi}_\infty^q := \lim_{n \rightarrow \infty} \widehat{\chi}_n^q = \left[\prod_{k \in \mathcal{K}_T} \frac{\widehat{\mathcal{C}}_{1,q}}{\bar{F}(\Delta T_k)} \right] \left[\prod_{k \in \overline{\mathcal{K}}_T \setminus \mathcal{K}_T} \widehat{\mathcal{C}}_{2,q} \frac{|c_{I_k}|_\infty}{p_{I_k}} \frac{1}{\rho(\Delta T_k)} \right].$$

Then by the same arguments as in the proof of Proposition 3.4, it is easy to check that

$$\eta(0) = \eta(T) + \int_0^T \sum_{\ell \in L} \widehat{C}_{2,q} |c_\ell|_\infty \eta(t)^{|\ell|} dt = \mathbb{E}[\widehat{\chi}_1^q] = \mathbb{E}[\widehat{\chi}_n^q], \quad \forall n \geq 1;$$

and hence by direct computation, it follows that

$$\mathbb{E}[|\psi|^q] \leq \mathbb{E}[\chi_\infty^q] \leq \mathbb{E}[\widehat{\chi}_\infty^q] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[\widehat{\chi}_n^q] = \eta(0) < \infty.$$

Changing the origin from $(0, x_0)$ to (t, x) , we see that

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \mathbb{E}[|\psi^{t,x}|^q] \leq \sup_{t \in [0,T]} \eta(t) < \infty,$$

and hence $(\psi^{t,x})_{(t,x) \in [0,T] \times \mathbb{R}^d}$ is uniformly integrable. The same arguments using (3.12) show that $(\tilde{\psi}^{t,x} \overline{W}(t, T_{(1)}^t, x, \Delta W_{(1)}^t))_{(t,x) \in [0,T] \times \mathbb{R}^d}$ is uniformly integrable. \square

4 Further discussions

Representation of the PDE system Let us consider a PDE system $(v_j)_{j=1, \dots, n}$, where for each j , $v_j : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies

$$\partial_t v_j + \mu_j \cdot Dv_j + \frac{1}{2} a_j : D^2 v_j + f(\cdot, v_1, \dots, v_n, Dv_1, \dots, Dv_n) = 0,$$

for some diffusion coefficient function $(\mu_j, a_j) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{S}^d$, and some polynomial function $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^n \times (\mathbb{R}^d)^n \rightarrow \mathbb{R}$. Our methodology immediately applies to this context, and provides a stochastic representation for the solution of the above PDE system, by means of a regime-changed branching diffusions: at every branching time, the independent offspring particles perform subsequently different branching diffusion regime.

Representation in view of unbiased simulation With the same idea of proof, we can also obtain an alternative representation result, with a frozen coefficient SDE in place of SDE (2.4). When the coefficient function $a \equiv a_0$ for some constant $a_0 \in \mathbb{S}^d$, this has significant application in terms of Monte Carlo approximation, as it leads to a representation random variable which can be simulated exactly, while the branching diffusion process X^k in (2.4) needs a time discretization technique and hence creates some discretization error in the simulation. Let us present this alternative representation formula in the case of constant diffusion coefficient case, i.e. $a \equiv a_0 = \sigma_0 \sigma_0^\top$ for some non-degenerate constant matrix $\sigma_0 \in \mathbb{M}^d$.

Let $\widehat{L} := L \cup \{\partial\}$, where ∂ represents an artificial index; $\hat{p} = (\hat{p}_\ell)_{\ell \in \widehat{L}}$ be a probability mass function and $(\widehat{T}^{m,n})_{m,n \geq 1}$ be a sequence of i.i.d. random variables of distribution \hat{p} , and independent of the sequences of i.i.d Brownian motion $(\widehat{W}^{m,n})_{m,n \geq 1}$ and i.i.d positive random variable $(\widehat{T}^{m,n})_{m,n \geq 1}$ of density function ρ . Then following exactly the same procedure in Section 2.2, we can construct another age-dependent branching process, denoted by $(\widehat{T}_k)_{k \in \widehat{\mathcal{K}}_T}$ with branching type $\widehat{I}_k := \widehat{T}^{n, \pi_n(k)}$. Here, when $\widehat{I}_k =$

$(\hat{\ell}_0, \dots, \hat{\ell}_m) \in L$, it produces $|\hat{\ell}|$ offspring particles, marked by $i = 0, \dots, m$ exactly as in Step 3 in the construction of age-dependent process $\bar{\mathcal{K}}_T$ in Section 2.2; when $I_k = \partial$, it produces only one offspring particle, marked by $m + 1$. Then for every $k \in \widehat{\mathcal{K}}_T$, we equipped it with an independent Brownian motion \widehat{W}^k as in (3.2). Next, let us define $\widehat{X}_0^{(1)} = x_0$, and subsequently for every $k \in \widehat{\mathcal{K}}_T$,

$$\widehat{X}_{\widehat{T}_k}^k := \widehat{X}_{\widehat{T}_{k-}}^k + \mu(\widehat{T}_{k-}, \widehat{X}_{\widehat{T}_{k-}}^k) \Delta \widehat{T}_k + \sigma_0 \Delta \widehat{W}_{\Delta \widehat{T}_k}^k, \quad \text{with } \widehat{X}_{\widehat{T}_{k-}}^k := \widehat{X}_{\widehat{T}_{k-}}^{k-}. \quad (4.1)$$

For this case, the automatic differentiation functions take a particularly simple formula, which is compatible with the purpose of the unbiased simulation algorithm. Let us introduce

$$\begin{aligned} \widehat{W}_k &:= \mathbf{1}_{\{\theta_k=0\}} + b_{\theta_k}(\widehat{T}_{k-}, \widehat{X}_{\widehat{T}_{k-}}^k) \cdot (\sigma_0^\top)^{-1} \frac{\Delta \widehat{W}_{\Delta \widehat{T}_k}^k}{\Delta \widehat{T}_k} \mathbf{1}_{\{\theta_k \in \{1, \dots, m\}\}} \\ &+ \left(\mu(\widehat{T}_{k-}, \widehat{X}_{\widehat{T}_{k-}}^k) - \mu(\widehat{T}_{(k-)-}, \widehat{X}_{\widehat{T}_{(k-)-}}^{k-}) \right) \cdot (\sigma_0^\top)^{-1} \frac{\Delta \widehat{W}_{\Delta \widehat{T}_k}^k}{\Delta \widehat{T}_k} \mathbf{1}_{\{\theta_k=\partial\}}. \end{aligned} \quad (4.2)$$

Finally, setting $c_\partial \equiv 1$, and replacing (X^k, W_k, p_{I_k}) in the definition of ψ and $\tilde{\psi}$ (in and below (3.4)) by $(\widehat{X}^k, \widehat{W}_k, \widehat{p}_{\widehat{I}_k})$, we obtain

$$\widehat{\psi} := \left[\prod_{k \in \widehat{\mathcal{K}}_T} \frac{g(\widehat{X}_T^k) - g(\widehat{X}_{\widehat{T}_{k-}}^k) \mathbf{1}_{\theta_k \neq 0}}{\bar{F}(\Delta T_k)} \widehat{W}_k \right] \left[\prod_{k \in \widehat{\mathcal{K}}_T \setminus \widehat{\mathcal{K}}_T} \frac{c_{\widehat{I}_k}(\widehat{T}_k, \widehat{X}_{\widehat{T}_k}^k)}{\widehat{p}_{\widehat{I}_k}} \frac{\widehat{W}_k}{\rho(\Delta \widehat{T}_k)} \right], \quad (4.3)$$

and similarly $\tilde{\psi}$.

Next, given a constant vector $\mu_0 \in \mathbb{R}^d$, we keep the same branching Brownian motion $(\widehat{W}^k)_{k \in \widehat{\mathcal{K}}_T}$, and then introduce another diffusion process $\widehat{X}^{\mu_0, k}$ by

$$\widehat{X}_{\widehat{T}_{(1)}}^{\mu_0, k} := x_0 + \mu_0 \Delta \widehat{T}_{(1)} + \sigma_0 \Delta \widehat{W}_{\Delta T_{(1)}}^{(1)},$$

and the subsequent process $\widehat{X}_{\widehat{T}_k}^{\mu_0, k}$ for $k \in \widehat{\mathcal{K}}_T \setminus \{(1)\}$ by the same induction relation as in (4.1). We then introduce $\widehat{W}_k^{\mu_0}$ as in (4.2) by replacing \widehat{X}^k by $\widehat{X}^{\mu_0, k}$, and replacing $\mu(\widehat{T}_{(k-)-}, \widehat{X}_{\widehat{T}_{(k-)-}}^{k-})$ by μ_0 when $k = (1)$. Replacing $(\widehat{X}^k, \widehat{W}_k)$ by $(\widehat{X}^{\mu_0, k}, \widehat{W}_k^{\mu_0})$ in (4.3), it defines a new random variable $\widehat{\psi}^{\mu_0}$. Finally, by changing the initial condition $(0, x_0)$ and time interval $[0, T]$ to (t, x) and $[t, T]$, one obtains $\widehat{W}^{t, k}$, \widehat{T}_k^t , $\widehat{\psi}^{t, x}$, $\tilde{\psi}^{t, x}$, $\widehat{\psi}^{\mu_0, t, x}$ etc.

Proposition 4.1. *Suppose that Assumptions 3.1 holds true, and the semilinear PDE (2.2) has uniqueness for bounded viscosity solution. Suppose in addition that for every $(t, x) \in [0, T] \times \mathbb{R}^d$, and μ_0 lies in a neighborhood of $\mu(t, x)$, one has*

$$\widehat{\psi}^{\mu_0, t, x} \quad \text{and} \quad \widehat{\psi}^{\mu_0, t, x} \Delta \widehat{W}_{\Delta \widehat{T}_{(1)}^t}^{t, (1)} / \Delta \widehat{T}_{t, (1)} \quad \text{is integrable,}$$

and the family of random variables

$$(\widehat{\psi}^{t, x})_{(t, x) \in [0, T] \times \mathbb{R}^d} \quad \text{and} \quad (\tilde{\psi}^{t, x} \Delta \widehat{W}_{\Delta \widehat{T}_{(1)}^t}^{t, (1)} / \Delta \widehat{T}_{t, (1)})_{(t, x) \in [0, T] \times \mathbb{R}^d}$$

are uniformly integrable with uniformly bounded expectation, define $\hat{u}(t, x) := \mathbb{E}[\psi^{t,x}]$. Then the derivative $D\hat{u}$ exists, \hat{u} and $D\hat{u}$ are both continuous; and moreover, u is the unique bounded viscosity solution of semilinear PDE (2.2).

Sketch of proof. (i) First, by the uniform integrability condition, \hat{u} is bounded continuous. Let us introduce

$$\tilde{u}(\mu_0, t, x) := \mathbb{E}[\widehat{\psi}^{\mu_0, t, x}] \quad \text{and} \quad \hat{v}(t, x) := \mathbb{E}\left[\widehat{\psi}^{z^{t,x}} \Delta \widehat{W}_{\Delta \widehat{T}_{(1)}^t}^{t, (1)} / \Delta \widehat{T}_{(1)}^t\right].$$

Notice that \hat{v} is uniformly bounded and continuous. Recall that W is a standard d -dimensional Brownian motion independent of the branching diffusion process, we also introduce

$$\widehat{X}_s^{t,x} := x + \mu_0(s-t) + \sigma_0(W_s - W_t), \quad s \in [t, T],$$

where $\mu_0 \in \mathbb{R}^d$ is a constant vector in a neighborhood of $\mu(t, x)$. Then one obtains as in (3.5) that

$$\begin{aligned} \tilde{u}(\mu_0, t, x) &= \mathbb{E}\left[\frac{1}{F(\widehat{T}_{(1)}^t)} g(\widehat{X}_{\widehat{T}_{(1)}^t}^{t,x}) \mathbf{1}_{\{\widehat{T}_{(1)}^t = T\}} + \frac{\mathbf{1}_{\{\widehat{T}_{(1)}^t = \partial\}}}{\rho(\widehat{T}_{(1)}^t) \hat{p}_\partial} ((\mu - \mu_0) \cdot \hat{v})(\widehat{T}_{(1)}^t, \widehat{X}_{\widehat{T}_{(1)}^t}^{t,x}) \right. \\ &\quad \left. + \frac{\mathbf{1}_{\{\widehat{T}_{(1)}^t \neq \partial\}}}{\rho(\widehat{T}_{(1)}^t)} \left(\frac{c_{\widehat{T}_{(1)}^t}}{\hat{p}_{\widehat{T}_{(1)}^t}} \hat{u}_{\widehat{T}_{(1)}^t, 0} \prod_{i=1}^m (b \cdot \hat{v})^{\widehat{T}_{(1)}^t, i} \right) (\widehat{T}_{(1)}^t, \widehat{X}_{\widehat{T}_{(1)}^t}^{t,x}) \mathbf{1}_{\{\widehat{T}_{(1)}^t < T\}} \right] \\ &= \mathbb{E}\left[g(\widehat{X}_T^{t,x}) + \int_t^T \left((\mu - \mu_0) \cdot \hat{v} + f(\cdot, \hat{u}, \hat{v}) \right) (s, \widehat{X}_s^{t,x}) ds\right]. \end{aligned}$$

By standard argument, $(t, x) \mapsto \tilde{u}(\mu_0, t, x)$ is a viscosity solution of

$$-\partial_t u + \mu_0 \cdot Du + \frac{1}{2} a_0 : D^2 u + (\mu - \mu_0) \cdot \hat{v} + \sum_{\ell \in L} c_\ell \hat{u}^{\ell_0} \prod_{i=1}^m (b_i \cdot \hat{v})^{\ell_i} = 0,$$

with terminal condition g . Since \hat{u} and \hat{v} are bounded continuous, the above PDE has uniqueness for bounded viscosity solution, which induces that $\tilde{u}(\mu_0, t, x)$ is independent of μ_0 and $\hat{u}(t, x) = \tilde{u}(\mu_0, t, x)$ for μ_0 in a neighborhood of $\mu(t, x)$.

(ii) We can then compute the derivative $D_x \tilde{u}(\mu_0, t, x)$ and then set $\mu_0 := \mu(t, x)$, it follows that

$$D\hat{u}(t, x) = D_x \tilde{u}(\mu(t, x), t, x) = \hat{v}(t, x),$$

which is also bounded continuous. This implies that $\hat{u}(t, x)$ is a viscosity solution of (2.2), and we hence conclude the proof by uniqueness of the viscosity solution of (2.2). \square

The integrability and square integrability of $\widehat{\psi}$ can be analyzed in exactly the same way as in Theorem 3.12. We just notice that the above defined random variable $\widehat{\psi}$ can be simulated exactly from a sequence of Gaussian random variable, discrete distributed random variables $\widehat{T}^{m,n}$ and r.v. $\widehat{T}^{m,n}$ of distribution density function ρ . It is then in particular interesting to serve as a Monte-Carlo estimator for $u(0, x_0)$.

On the representation of fully nonlinear PDEs Formally, one can also obtain a representation result for fully nonlinear PDE, using the same automatic differentiation functions of order 2. However, this raises a serious integrability problem which can not be solved by conditions as in Assumption 3.10. To illustrate the main difficulty, let us consider the following PDE in the one-dimensional case $d = 1$:

$$u(T, x) = g(x), \quad \partial_t u + \frac{1}{2} D^2 u + f_0(D^2 u) = 0, \quad \text{on } [0, T] \times \mathbb{R}^d, \quad (4.4)$$

where $f_0(\gamma) = c_0 \gamma$ for some constant $c_0 > \frac{1}{2}$. Notice that there is only one term in function f_0 , then a natural guess for the representation is to consider a branching Brownian motion with exactly one offspring particle at every arrival time. This can be seen as a Brownian motion W equipped with a sequence of random time mark $(T_i)_{i=1, \dots, N_T}$, where

$$T_i := T \wedge \sum_{j=1}^i \tau^{j,1}, \quad N_T := \inf \left\{ i : T_i \geq T \right\}.$$

Notice that for any $t > 0$ and bounded measurable function $\phi : \mathbb{R} \rightarrow \mathbb{R}$, one has

$$\partial_{xx}^2 \mathbb{E}[\phi(x + W_t)] = \mathbb{E}\left[\phi(x + W_t) \frac{W_t^2 - t}{t^2}\right].$$

Then arguing as in Theorem 3.5, we may expect that $u(0, x_0) = \mathbb{E}[\widehat{\psi}]$, with

$$\widehat{\psi} := g(x + W_T) \frac{1}{\bar{F}(T - T_{N_T-1})} \prod_{i=1}^{N_T-1} c_0 \frac{(W_{T_{i+1}} - W_{T_i})^2 - (T_{i+1} - T_i)}{(T_{i+1} - T_i)^2 \rho(T_i - T_{i-1})},$$

provided that $\widehat{\psi}$ is integrable. However, the integrability of $\widehat{\psi}$ could fail in general. For simplicity, let $g \equiv 1$, and notice that $\bar{F} \leq 1$. Then by taking conditional expectation, one has, for some constant $C > 0$ and $c_1 := \mathbb{E}[|c_0(W_1^2 - 1)|]$, that

$$\begin{aligned} \mathbb{E}[|\widehat{\psi}|] &\geq \mathbb{E}\left[\prod_{i=1}^{N_T-1} \frac{c_1}{(T_{i+1} - T_i) \rho(T_i - T_{i-1})}\right] \\ &\geq \mathbb{E}\left[\frac{c_1}{\rho(T_1)} \frac{c_1}{(T_2 - T_1) \rho(T_2 - T_1)} \frac{c_1}{T_3 - T_2} \mathbf{1}_{\{T_1 \leq T/2, T_2 - T_1 < T/2, T_3 - T_2 \geq T\}}\right] \\ &\geq C \mathbb{E}\left[\frac{c_1}{(T_2 - T_1) \rho(T_2 - T_1)} \mathbf{1}_{\{T_2 - T_1 < T/2\}}\right] = C \int_0^{T/2} \frac{1}{t} dt = \infty. \end{aligned}$$

Of course, for linear PDEs as in (4.4), one can simulate a Brownian motion with volatility coefficient $1 + 2c_0$ whenever $1 + 2c_0 > 0$ to obtain the solution. But it is not the case for general fully nonlinear PDEs.

On the representation results by BSDE Another probabilistic representation of semilinear parabolic PDE is the Backward Stochastic Differential Equation (BSDE) proposed by Pardoux and Peng [20]. Namely, given a classical solution u of semilinear PDE (2.2), we define

$$(Y_t, Z_t) := (u(t, \bar{X}_t^{0, x_0}), \sigma Du(t, \bar{X}^{0, x_0})).$$

Then (Y, Z) provides a solution to BSDE

$$Y_t = g(\bar{X}_T^{0,x_0}) + \int_t^T f(s, \bar{X}_s^{0,x_0}, Y_s, \sigma^{-1}(s, \bar{X}_s^{0,x_0})Z_s)ds - Z_s dW_s, \quad t \in [0, T], \quad \mathbb{P}\text{-a.s.}$$

Based on the discretization technique on the BSDE, one can then obtain a probabilistic numerical solution for semilinear parabolic PDEs, see e.g. Bouchard and Touzi [5], and Zhang [28], etc. Generally speaking, these numerical schemes for BSDE need a (time-consuming) simulation-regression technique to compute the conditional expectation appearing in the schemes.

Our representation result induces a pure Monte Carlo simulation algorithm, which avoids the regression procedure in the numerical schemes of BSDEs. Nevertheless, our numerical method provides only the solution of PDE at time 0, and it needs some restrictive conditions on the coefficient functions f such as Assumption 3.10 to obtain a finite variance estimator. We will provide more numerical examples as well as some variance reduction techniques in Section 5 below.

5 A Monte Carlo algorithm

5.1 The implementation of the numerical algorithm

The above representation result in Theorem 3.5 induces a Monte Carlo algorithm to compute the solution of PDE (2.2), by simulating the random variable ψ or $\hat{\psi}$. We provide here some discussion on the implementation of the numerical algorithm.

The choice of density function ρ As discussed in Remark 3.11, to ensure Assumption 3.10, a necessary condition is to choose $\rho(t) \geq Ct^{-1/2}$. A natural candidate as distribution, which is also easy to be simulated, is the gamma distribution $\Gamma(\kappa, \theta)$, with $\kappa \leq \frac{1}{2}$, whose density function is given by

$$\rho_0(t) = \frac{1}{\Gamma(\kappa)\theta^\kappa} t^{\kappa-1} \exp(-t/\theta) \mathbf{1}_{\{t>0\}}, \quad (5.1)$$

where $\Gamma(\kappa) := \int_0^\infty s^{\kappa-1} e^{-s} ds$. In particular, one has

$$\bar{F}_k := \int_{\Delta T_k}^\infty \rho_0(t) dt = 1 - \frac{\gamma(\kappa, \Delta T_k / \theta)}{\Gamma(\kappa)}, \quad \text{where } \gamma(\kappa, t) := \int_0^t s^{\kappa-1} e^{-s} ds.$$

Complexity The dimension d of the problem, the choice of $(p_\ell)_{\ell \in L}$ and ρ will of course influence the complexity of algorithm. First, the complexity is proportional to the number of particles in the branching process, i.e. $\#\bar{\mathcal{K}}_T$, and for each particle, the complexity of simulation and calculation is of order Cd^2 . Let us denote $n_0 := \sum_{\ell \in L} p_\ell |\ell|$ and $m(t) := \mathbb{E}[\#\bar{\mathcal{K}}_t]$.

Proposition 5.1. (i) *The function $m(t)$ is given by*

$$m(t) = \sum_{k=0}^{\infty} n_0^k F^{*,k}(t), \quad \text{where } F^{*,k}(t) := \mathbb{P}[\tau^{1,1} + \dots + \tau^{1,k} < t]. \quad (5.2)$$

(ii) Let ρ be given by (5.1), then $F^{*,k}(t) = \frac{1}{\Gamma(k\kappa)}\gamma(k\kappa, x/\theta)$ and hence

$$m(t) = \sum_{k=0}^{\infty} \frac{\gamma(k\kappa, t/\theta)n_0^k}{\Gamma(k\kappa)}.$$

Proof. (i) Using Lemma 4.4.3 of Athreya and Ney [1], one has that satisfies the equation $m(t) = 1 + n_0 \int_0^t m(t-s)\rho(s)ds$, whose solution is given explicitly by (5.2). Further, when ρ is the density function of Gamma distribution, the function $F^{*,k}(t)$ can be computed explicitly. \square

5.2 A high dimensional numerical example

We first focus on a simple numerical example in high dimension. Let $(\mu, \sigma) \equiv (0, \sigma_0)$ for some constant matrix $\sigma_0 = \frac{1}{\sqrt{d}}\mathbf{I}_d$, and $f(t, x, y, z) = k(t, x) + cy(b \cdot z)$, where $b := \frac{1}{d}(1 + \frac{1}{d}, 1 + \frac{2}{d}, \dots, 2)$ and

$$k(t, x) := \cos(x_1 + \dots + x_d) \left(\alpha + \frac{\sigma^2}{2} + c \sin(x_1 + \dots + x_d) \frac{3d+1}{2d} e^{\alpha(T-t)} \right) e^{\alpha(T-t)}.$$

With terminal condition $g(x) = \cos(x_1 + \dots + x_d)$, the explicit solution of semilinear PDE (2.2) is given by

$$u(t, x) = \cos(x_1 + \dots + x_d) e^{\alpha(T-t)}.$$

In our numerical experiment, we set $\alpha = 0.2$, $c = 0.15$, $T = 1$, and $x_0 = 0.5\mathbb{1}_d$, where $\mathbb{1}_d$ stands for the unit vector in \mathbb{R}^d for $d = 5, 10$ and 20 . We would like to emphasize that, to the best of our knowledge, no alternative methods are available for solving such a high-dimensional semilinear PDE. In Table 1, we report the analytic solution of the semilinear PDE and that of the corresponding linear PDE by setting $c = 0$. The different results indicate that the nonlinearity term has an impact.

Dimension	5	10	20
Linear Solution	-1.0436	0.3106	-0.9661
Non linear solution	-0.97851	0.34646	-1.0248

Table 1: Analytical solution for the linear PDE (i.e., $c = 0$) versus analytical solution for the semilinear PDE in $d = 5, 10$ and 20 .

For numerical implementations, we use gamma distribution (5.1), with $\kappa = 0.5$ and $\theta = 2.5$. On each test performed, a computation is achieved with n particles. An estimation E with n particles is then calculated. The standard deviation of E is estimated with 1000 runs of n particles and its log-plot is reported below on the different figures for different values of n . We also show on some figures the convergence of the solution obtained on the average of the 1000 runs.

On Figures 1, 2, 3, we illustrate that the Monte Carlo method converges easily to our analytic solution. Computational costs are estimated on one core of a Laptop core I7 processor 2.2 GHz and are reported in Table 2 for a number of simulations equal to 96000 permitting to get a solution with an error less than 0.1%.

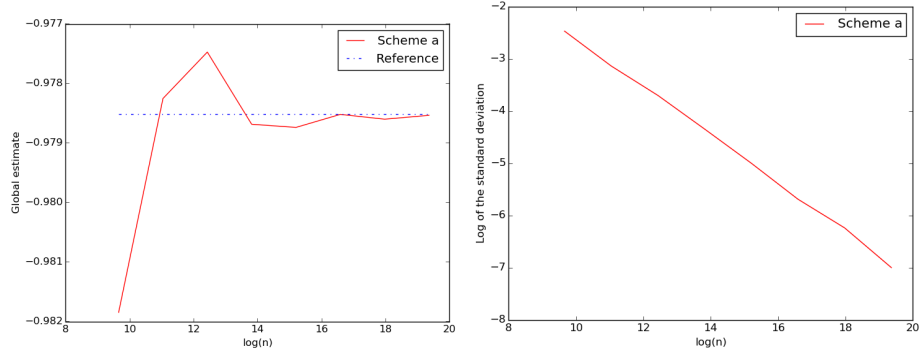


Figure 1: Estimation and standard deviation observed in $d = 5$ depending on the log of the number of particles used.

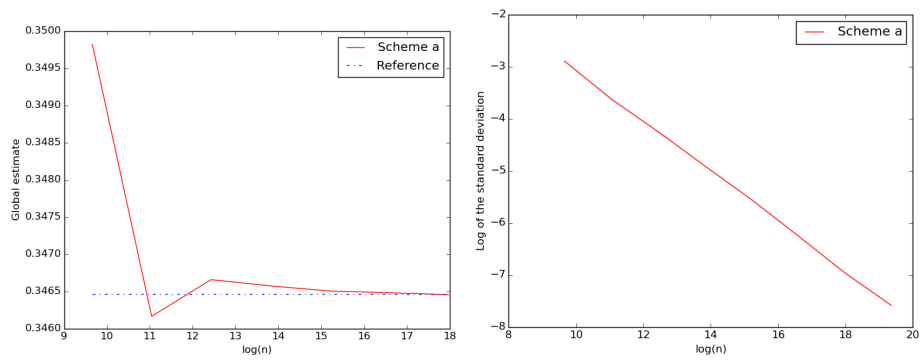


Figure 2: Estimation and standard deviation observed in $d = 10$ depending on the log of the number of particles used.

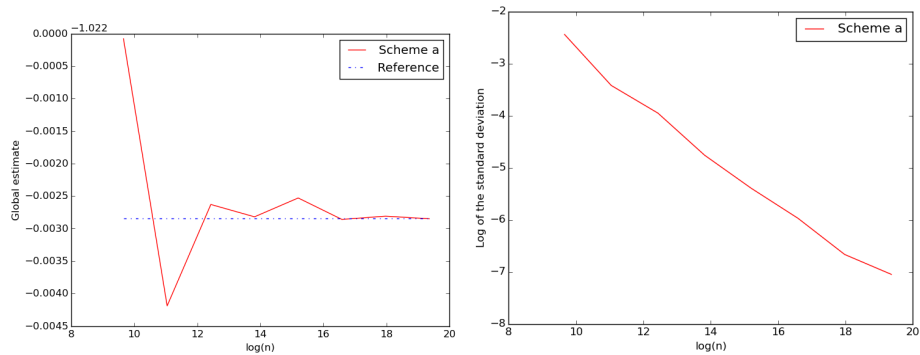


Figure 3: Estimation and standard deviation in $d = 20$ depending on the log of the number of particles used.

6 Some extensive tests

This section is devoted to additional tests. Having illustrated previously that our algorithm is efficient for solving high-dimensional semilinear PDEs, we focus on some examples from dimension 1 to 3. Note that our results have been benchmarked against

Dimension	5	10	20
Time	550	717	956

Table 2: Computational time in seconds for 96000 trajectories computed 1000 times on one core for $\kappa = 0.5$, $\theta = 2.5$.

a Finite Difference method in $d = 1$ and $d = 2$. Unfortunately, the finite difference method is no more available in $d = 3$. All our numerical examples share the following characteristics: $\mu(t, x) = 1 - x$, $\sigma \equiv 0.5\mathbf{I}_d$ and $x_0 = \mathbf{1}$. T is chosen equal to 1, $g(x) = (\frac{1}{d} \sum_i^d x(i) - 1)^+$ for $x \in \mathbb{R}^d$. Notice that with the above coefficients, SDE (3.1) is a linear SDE, whose solution can exactly simulated:

$$\bar{X}_t^{0,x} = (1 - e^{-t})\mathbf{1} + e^{-t}x + \sigma \sqrt{\frac{1 - e^{-2t}}{2}}Z, \quad Z \sim N(0, \mathbf{I}_d). \quad (6.1)$$

The Malliavin weight used in the algorithm can be computed explicitly and is given by $Z/(\sigma \sqrt{\frac{e^{2t}-1}{2}})$. We will compare numerical results from four different schemes.

- **(scheme a)** using the representation (3.4) with the explicit solution (6.1) of the SDE (3.1).
- **(scheme b)** using the representation (4.3) with freezing coefficient techniques.
- **(scheme c)** using the representation (4.3), enhanced by the resampling scheme (see Appendix for more details).
- **(scheme d)** using the representation (3.4), enhanced by the resampling scheme.

The density function ρ is that of the gamma law with parameters κ and θ . If not indicated, the parameters of the law are set to $\kappa = 0.5$ and $\theta = 2.5$ and the probability p_l are chosen equal. On each test, a calculation is achieved with n particles (starting with $n = 1562$ for scheme **a** and with $n = 100000$ for schemes **b** and **c**). The n particles are shared on 96 processors and each processor i calculates an estimation E_i of the solution with $\frac{n}{96}$ particles. Then an estimation E with n particles is achieved with $E = \frac{1}{96} \sum_i E_i$. When importance sampling is used, in order to avoid communications that breaks parallelism, it is used on each processor so with $\frac{n}{96}$ particles on each processor. The standard deviation of E is estimated with 1000 runs of n particles and its log is reported on the different figures below for different values of n . We expect that by quadrupling the values of n , the standard deviation std divides by a factor 2 and the plot $(\log(n), \log(std))$ should be linear with a slope equal to $-\frac{1}{2}$. The theoretical rate of convergence is also plotted on each figure (as in our previous example, the solutions are obtained on the average of the 1000 runs).

6.1 Some examples in one space dimension

- For $d = 1$, we take $f(t, x, y, z) := 0.2y^2 + 0.3y^3$. Results on Figure 4 show that the method converges. Scheme **a** is far more effective than scheme **b** and that the importance sampling of scheme **c** is effective. The log of the standard

deviation decreases for all schemes linearly with the log of the particle number as predicted by the theory. Note that the computational cost for 1000 runs with 25000 particles on one core is equal to 490 seconds for scheme **a**, 200 seconds with scheme **b** and 260 seconds with scheme **c**.

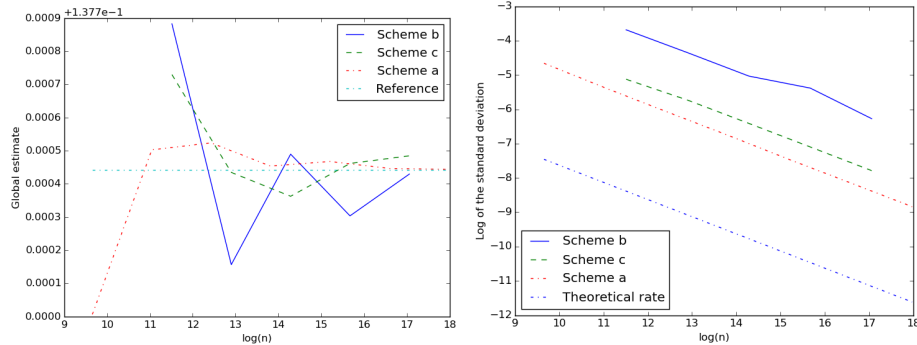


Figure 4: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) := 0.2y^2 + 0.3y^3$

- As a second example in $d = 1$, we take a Burgers type nonlinearity $f(t, x, y, z) = 0.15yz$. Results on Figure 5 show that all the schemes converge to our numerical finite difference solution. Note that the computational cost for 1000 runs with 25000 particles on one core is roughly equal to 200 seconds for scheme **a**, 100 seconds for scheme **b**, 300 seconds for scheme **c**.

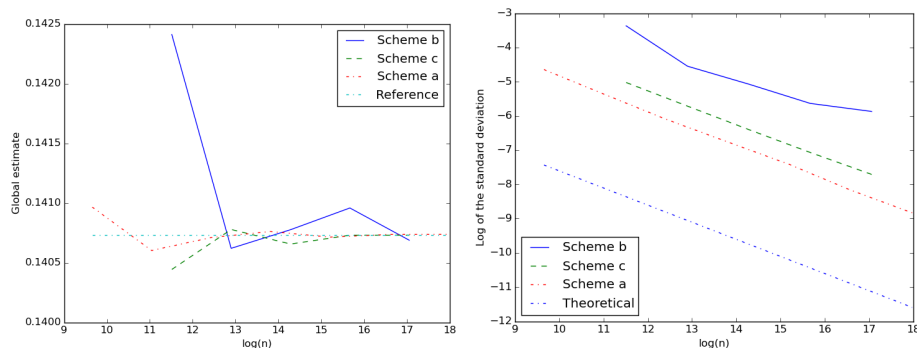


Figure 5: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) = 0.15yz$.

- As a third example in $d = 1$, we keep the same nonlinearity with $f(t, x, y, z) = 0.3yz$. We expect that the variance of the results will be higher than in the previous case. This is observed in Figure 6. Scheme **a** still converges. Scheme **b** converges slowly and Importance Sampling of scheme **c** permits to get faster convergence and to recover the good rate in the log of the standard deviation decay. The computational times are the same as in our previous test.
- As a fourth example in $d = 1$, we take a nonlinearity with $f(t, x, y, z) = 0.08z^2$. Results are shown in Figure 7. The importance sampling of scheme **c** is required

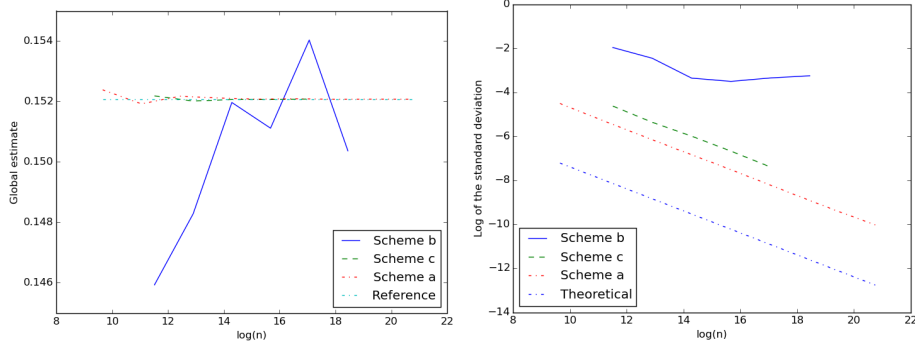


Figure 6: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) = 0.3yz$.

to achieve proper convergence. Scheme **a** converges quickly. Computational times are the same as before (same type of branching).

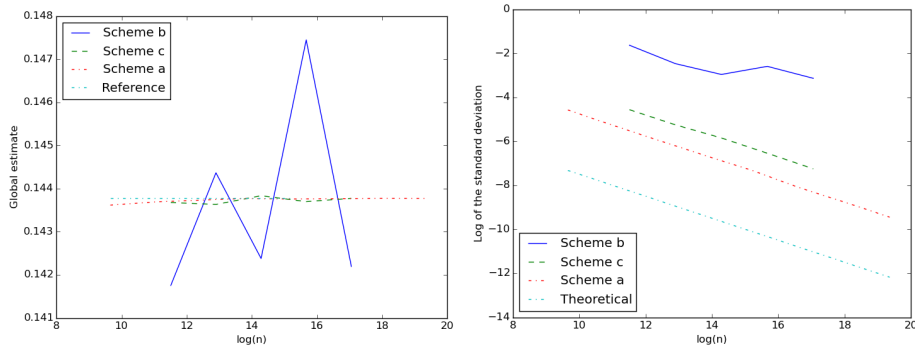


Figure 7: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) = 0.08z^2$.

- As a last example in $d = 1$, we keep the same type of nonlinearity $f(t, x, y, z) = 0.2z^2$. Schemes **b** and **c** don't converge anymore. We only test scheme **a** using different values for the parameters κ and θ (see Figure 8). The change in θ does not seem to change convergence properties. The change in κ (from 0.5 to 0.4) does not seem to modify our results. However, some tests, not reported here, show that the variance can increase a lot using κ around 0.25. Then, as the average jump size is proportional to θ , it is more efficient to take some quite high values for θ in order to reduce the computational time. For the same reason, it is optimal to choose a κ equal to 0.5. In Table 3, we report the computational time, associated to different choices of (κ, θ) , as a multiplicative factor of the computational effort with benchmark parameters $\kappa = 0.5, \theta = 2.5$.

We notice that for all the parameters, the decay in the variance is far from the expected theoretical one (see Figure 8). We then use our benchmark parameters and compare the results obtained using scheme **a** and scheme **d** (importance sampling is used here). Results are reported on Figure 9. They illustrate that the importance sampling method allows to improve the convergence rate.

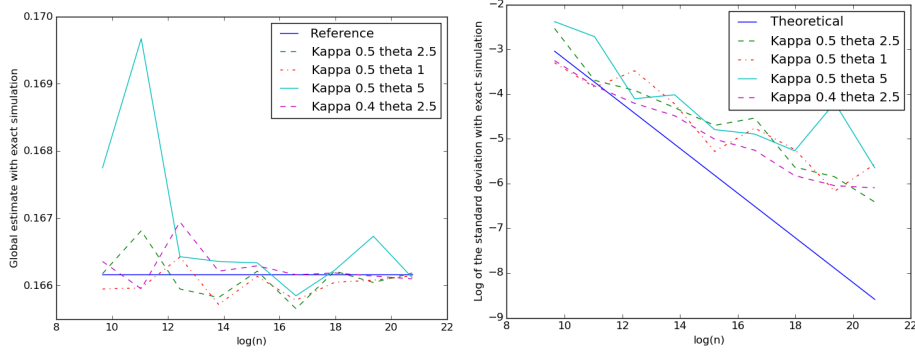


Figure 8: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) = 0.2z^2$. Different values for κ and θ are used.

Table 3: Computational time, associated to different choices of (κ, θ) , as a multiplicative factor of the computational effort with benchmark parameters $\kappa = 0.5, \theta = 2.5$

κ	0.5	0.5	0.5	0.4	0.4
θ	1	2.5	5	2.5	5.
Time	6.63	1	0.49	2.85	1.02

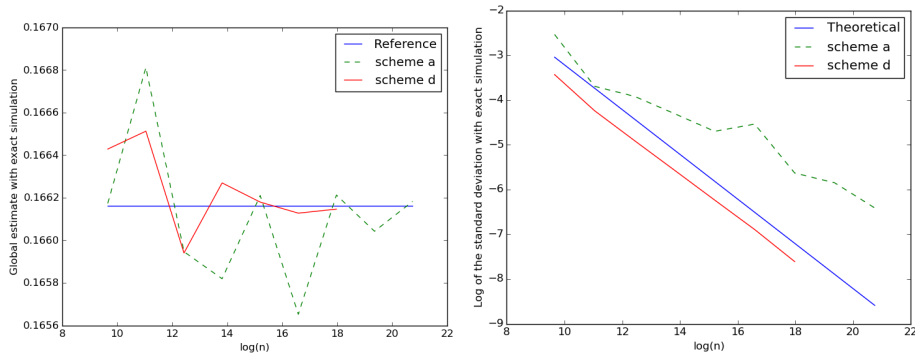


Figure 9: Estimation and standard deviation obtained in $d = 1$ for $f(t, x, y, z) = 0.2z^2$ with and without importance sampling.

6.2 Some examples in two space dimensions

Although the efficiency of our algorithm was illustrated on our previous experiments, this Monte-Carlo method cannot compete a PDE deterministic methods in $d = 1$. In this section, we focus on $d = 2$, where advantages of PDE implementation remain but are not so obvious.

- For the first example in $d = 2$, we take $f(t, x, y, z) := 0.15y\mathbb{I}.z$. On Figure 10, we give the results obtained using our three schemes showing that Importance Sampling is needed. Note that the computation cost for 1000 runs with 25000 particles on one core is roughly equal to 230 seconds for scheme **a**, 90 seconds for

scheme **b**, 580 seconds for scheme **c**.

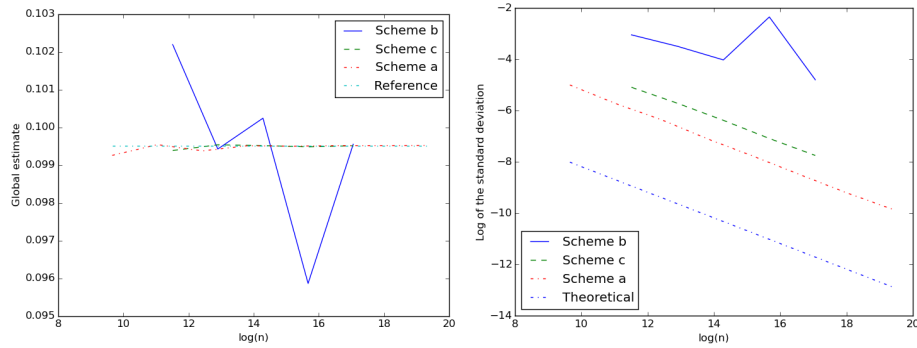


Figure 10: Estimation and standard deviation obtained in $d = 2$ for $f(t, x, y, z) := 0.15y\mathbb{I}.z$.

- For the second example in $d = 2$, we take $f(t, x, y, z) := 0.04(z.\mathbb{I})^2$. The convergence of Scheme **a** is easily achieved while Scheme **b** converge poorly as shown in Figure 11. Importance sampling method improve the convergence. Computational costs are the same as in our first $d = 2$ tests.

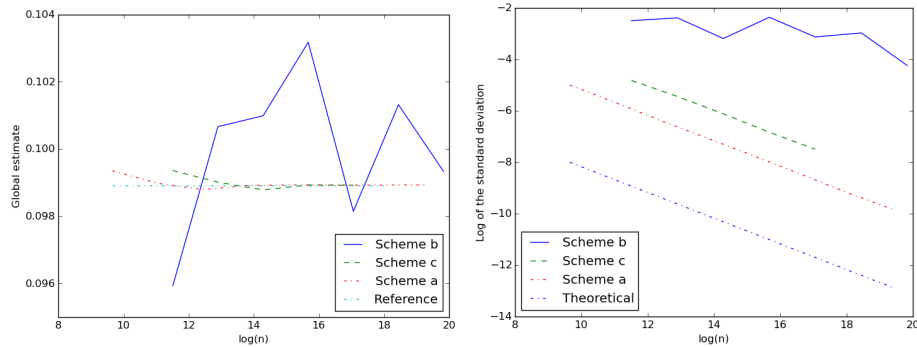


Figure 11: Estimation and standard deviation obtained in $d = 2$ for $f(t, x, y, z) := 0.04(z.\mathbb{I})^2$.

- For the third example, we test the influence of the coefficients on Scheme **a** for a non linearity $f(t, x, y, z) := K(z.\mathbb{I})^2$ with $K = 0.05, K = 0.1, K = 0.2$. Using Scheme **b** and **c**, we cannot get proper convergence due to high variances observed. On Figure 12, we give the convergence obtained with the different K values and on Figure 13 the standard deviation associated. As the coefficients grow, the variance of the results gets higher preventing the method from converge when $K = 0.2$.
- At last we test the influence of the function g . The representation of the solution involves the product of g functions so we expect that the variance of the result is highly sensitive to the scaling of this function. Here we choose to keep $f(t, x, r, p) := 0.05(Du.\mathbb{I})^2$ and take different values for the g function. On figure 14 we take $g(x) = 2(\frac{1}{d}\sum_i^d x(i) - 1)^+$ and give the convergence of schemes **a** and

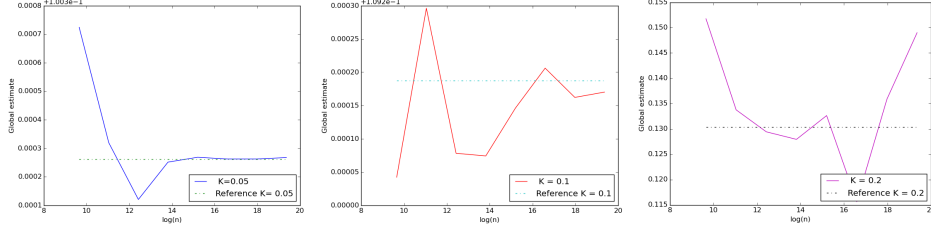


Figure 12: Convergence of scheme **a** for different K values.

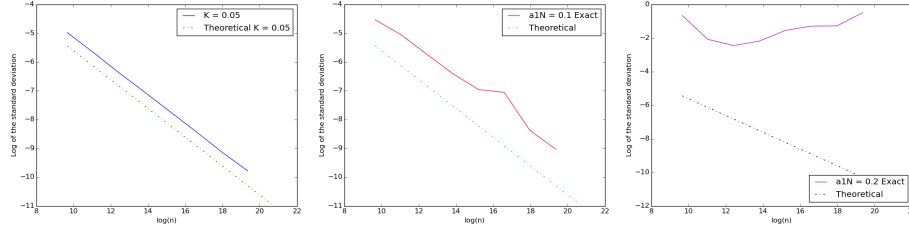


Figure 13: Standard deviation of the scheme **a** for different K values.

b and the standard deviation associated. Comparing to figure 13 ($K = 0.05$), we see a net increase in the variance of the result for scheme **a**. When importance sampling is used (scheme **d**) the decay in term of variance is more regular. Increasing the function g such that $g(x) = 3(\frac{1}{d} \sum_i^d x(i) - 1)^+$, we give the results obtained on figure 15. Here importance sampling is really necessary to recover a good rate of convergence.

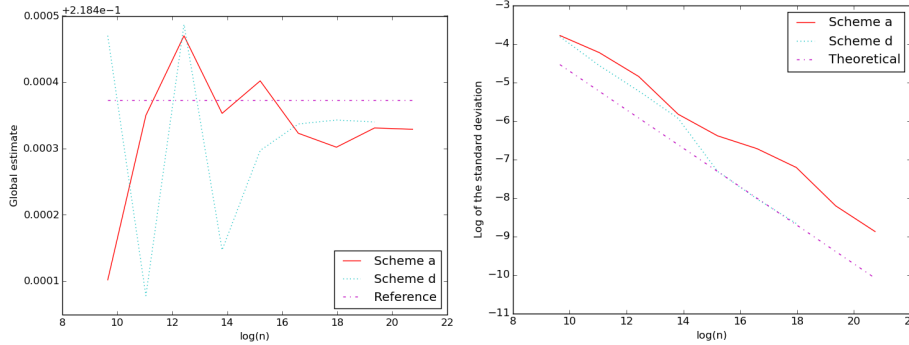


Figure 14: Estimation and standard deviation observed in dimension 2 for case 4, $g(x) = 2(\frac{1}{d} \sum_i^d x(i) - 1)^+$.

6.3 An example in three space dimensions

We take $f(t, x, y, z) := 0.15(z \cdot \mathbf{1})^2$. Results are given on Figure 16, still showing that importance sampling is necessary while using discretization of the scheme and that the exact scheme has a lower variance.

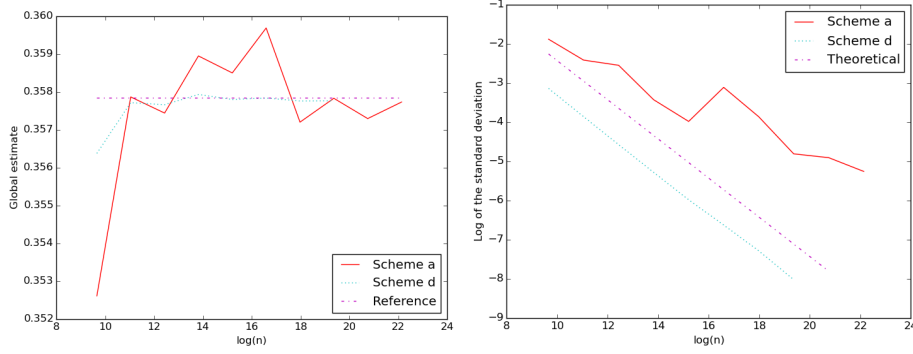


Figure 15: Estimation and standard deviation observed in dimension 2 for case 4, $g(x) = 3(\frac{1}{d} \sum_i^d x(i) - 1)^+$.

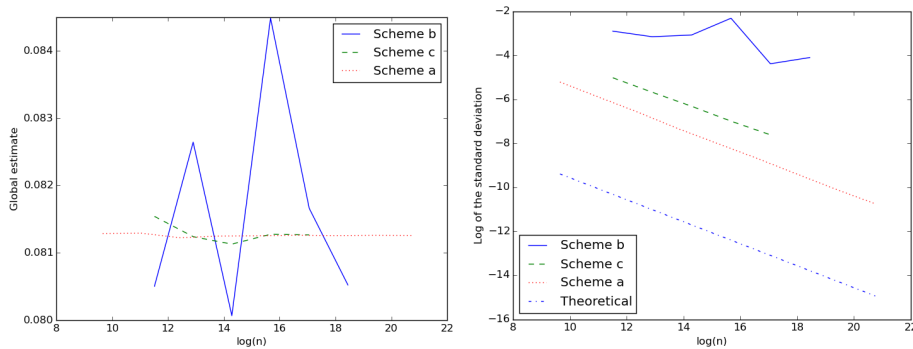


Figure 16: Estimation and standard deviation observed in dimension 3 for case 1.

A Resampling scheme for branching processes

Notice that our estimator (3.4) and (4.3) are provided as a product of some random variables. Then similar to Doumbia, Oudjane and Warin [7], one can use the resampling scheme (or interacting particle systems), see Del Moral [6]. Intuitively, this scheme replaces the expectation of a product by a product of expectations, which potentially stabilizes the Monte-Carlo estimator.

Let us first introduce the Markov chain $(\mathcal{X}_n)_{n \geq 1}$, taking values in $\cup_{p \geq 1}([0, T]^2 \times \mathbb{R}^{2d} \times \{0, \dots, m\} \times L)^p$ such that $\mathcal{X}_1 = (0, 0, x_0, x_0, 0, I_0)$ with $I_0 = (1, 0, \dots, 0) \in \mathbb{N}^{m+1}$ and for any $n \geq 1$, one defines

$$\mathcal{X}_{n+1} := \left(\mathcal{X}_n, (T_{k^-}, T_k, X_{T_{k^-}}^k, X_{T_k}^k, \theta_k, I_k)_{k \in \cup_{p=1}^n \bar{\mathcal{K}}_T^p} \right).$$

Notice that this Markov chain has an absorbing state since for any $\omega \in \Omega$ there is a generation $n(\omega)$ for which all branches have died (either having no offspring before reaching T or having reached T) implying $\bar{\mathcal{K}}_T^{n+1}(\omega) = \emptyset$ and consequently $\mathcal{X}_{n+1}(\omega) = \mathcal{X}_n(\omega)$. Then $(\mathcal{X}_n)_{n \geq 0}$ is a Markov chain. We next introduce

$$G_n(\mathcal{X}_n) := \left[\prod_{k \in \bar{\mathcal{K}}_T^n} \frac{g(X_T^k) - g(X_{T_{k^-}}^k) \mathbf{1}_{\{\theta_k \neq 0\}}}{\bar{F}(\Delta T_k)} \mathcal{W}_k \right] \left[\prod_{k \in (\bar{\mathcal{K}}_T^n \setminus \mathcal{K}_T^n)} \frac{c_{I_k}(T_k, X_{T_k}^k)}{p_{I_k}} \frac{\mathcal{W}_k}{\rho(\Delta T_k)} \right], \quad (\text{A.1})$$

so that

$$\psi = \prod_{n=1}^{\infty} G_n(\mathcal{X}_n).$$

Notice that the above representation consists of a product from contributions from each generation $n \geq 1$. Since the number of generation prior to the maturity T is finite a.s., the last product only involves finite number of terms, a.s. We also observe that except for the trivial case of constant function g , $\mathbb{E}[|G_n(\mathcal{X}_n)|] \neq 0$. By iteration, it is easy to see that

$$\mathbb{E}^{\mathbb{P}_0}[\psi] = \left(\prod_{n=1}^{\infty} \mathbb{E}^{\mathbb{P}_{n-1}}[|G_n(\mathcal{X}_n)|] \right) \mathbb{E}^{\mathbb{P}_\infty} \left[\prod_{n=1}^{\infty} \text{sgn}(G_n(\mathcal{X}_n)) \right],$$

where given \mathbb{P}_0 , one defines \mathbb{P}_n by $\frac{d\mathbb{P}_n}{d\mathbb{P}_{n-1}} := \frac{|G_n(\mathcal{X}_n)|}{\mathbb{E}^{\mathbb{P}_{n-1}}[|G_n(\mathcal{X}_n)|]}$, for $n \geq 1$.

The particle algorithm consists in simulating the dynamics of an interacting particle system of size N , $(\xi_p^{1,N}, \dots, \xi_p^{N,N})$, on $\cup_{p \geq 1} ([0, T]^2 \times \mathbb{R}^{2d} \times \{0, \dots, m\} \times L)^p$, from step $n = 1$ to $n = \infty$ and then to approximate each expectation $\mathbb{E}^{\mathbb{P}_{n-1}}[|G_n(\mathcal{X}_n)|]$ by the empirical mean value of the simulation. The algorithm can be given as an iteration of the following two steps, initiated by $n = 1$,

Selection step Given N copies of simulation $(\xi_n^{i,N})_{i=1, \dots, N}$ of \mathcal{X}_n , one draws randomly and independently N particles among the current particle system with a probability $\frac{|G_p(\xi_p^{i,N})|}{\sum_{j=1}^N |G_p(\xi_p^{j,N})|}$;

Evolution step Each new selected particle evolves randomly and independently according to the transition of the Markov chain (\mathcal{X}_n) between n and $n + 1$.

Finally $u(0, x_0)$ is approximated as a product of empirical averages:

$$\prod_{n=1}^{\infty} \left(\frac{1}{N} \sum_{i=1}^N |G_n(\xi_n^{i,N})| \right) \left(\frac{1}{N} \sum_{i=1}^N \prod_{n=1}^{\infty} \text{sgn}(G_n(\xi_n^{i,N})) \right). \quad (\text{A.2})$$

Notice again that, for every simulation $(\xi^{i,N})$, the maturity T is attained for some finite generation, then the above product $\prod_{n=1}^{\infty}$ can be restricted to the a finite product $\prod_{n=1}^{n_N}$, where $n_N := \inf\{n \geq 1 \mid \xi_n^{i,N} \text{ has reached } T \text{ for all } i = 1, \dots, N\}$.

References

- [1] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [2] V. Bally and P. Pages, *Error analysis of the quantization algorithm for obstacle problems*, Stochastic Processes & Their Applications, 106(1), 1-40, 2003.
- [3] H. BAUKE AND S. MERTENS, *Random numbers for large-scale distributed Monte Carlo simulations*, Physical Review E, 75(6):066701, 2007.
- [4] B. Bouchard, I. Ekeland and N. Touzi, *On the Malliavin approach to Monte Carlo approximation of conditional expectations*, Finance and Stochastics, 8, 45-71, 2004.

- [5] 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.
- [6] P. Del Moral, Feynman-Kac formulae, Genealogical and interacting particle systems with applications, Probability and its Applications, Springer-Verlag, New York, 2004.
- [7] M. Doumbia, N. Oudjane and X. Warin, *Computing expectations for general SDE with pure Monte Carlo methods*, preprint, 2016.
- [8] 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.
- [9] A. M. Etheridge, *An Introduction to Superprocesses*, American Mathematical Society, 2000.
- [10] A. Fahim, N. Touzi, and X. Warin. *A probabilistic numerical method for fully nonlinear parabolic PDEs*. The Annals of Applied Probability, 1322-1364, 2011.
- [11] Fournié, E., Lasry, J.M., Lebuchoux, J., Lions, P.L, Touzi, N. : *Applications of Malliavin Calculus to Monte Carlo Methods in Finance.*, Finance and Stochastics, 3, 391-412, 1999.
- [12] T. E. Harris, *The Theory of Branching Processes*, Die Grundlehren der Mathematischen Wissenschaften 119, Springer, Berlin, 1963.
- [13] 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>.
- [14] P. Henry-Labordère, X. Tan, N. Touzi. *A numerical algorithm for a class of BSDEs via the branching process*, Stochastic Processes and their Applications, 124(2):1112-1140, 2014.
- [15] P. Henry-Labordère, X. Tan, N. Touzi. *Unbiased simulation of stochastic differential equations*, preprint, 2015.
- [16] G. Kersting and F.C. Klebaner, *Sharp conditions for nonexplosions and explosions in Markov jump processes*, Anna. Proba. 23(1), 268-272, 1995.
- [17] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, Vol. 23, Springer, 1992.
- [18] H. P. McKean, *Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure Appl. Math., Vol 28, 323-331, 1975.
- [19] D. Nualart. *The Malliavin calculus and related topics*. Vol. 1995. Berlin: Springer, 2006.
- [20] E. Pardoux and S. Peng, *Adapted solutions of backward stochastic differential equations*, System and Control Letters, 14, 55-61, 1990.

- [21] A. Rasulov, G. Raimova, and M. Mascagni, *Monte Carlo solution of Cauchy problem for a nonlinear parabolic equation*. Mathematics and Computers in Simulation, 80(6):1118-1123, 2010.
- [22] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, American Mathematical Society, Graduate Studies in Mathematics, Volume 140, 2012.
- [23] A.V. Skorokhod *Branching diffusion processes*. Theory of Probability & Its Applications, 9(3):445-449, 1964.
- [24] D. W. Stroock, S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer, 1979.
- [25] J. Zhang, *A numerical scheme for backward stochastic differential equations*, Annals of Applied Probability, 14(1), 459-488, 2004.
- [26] 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.
- [27] S. Watanabe, *On the branching process for Brownian particles with an absorbing boundary*. Journal of Mathematics of Kyoto University, 4(2):385-398, 1965.
- [28] J. Zhang, *A numerical scheme for backward stochastic differential equations*, Annals of Applied Probability, 14(1), 459-488, 2004.