Branching Diffusions Representation for Nonlinear PDEs

Nizar Touzi
Ecole Polytechnique, Paris

Joint work with Nadia Oudjane, Pierre Henry-Labordère, Xiaolu Tan and Xavier Warin

University of Southern California
February 24, 2016
Objective

Design numerical approximation for the equation:

$$\partial_t v + \mu \cdot Dv + \frac{1}{2} \sigma^2 : D^2 v + F(t, x, v, Dv, D^2 v) = 0, \quad v(T, .) = g$$

- Finite differences, finite elements: very efficient in $1 - 2$ dim, curse of dimensionality, path dependency increases dimension

- **Probabilistic representation** $\implies$ Monte Carlo / Probabilistic numerical methods

- **An important issue**: extension to the path-dependent case??

Nizar Touzi
Branching Diffusions and Nonlinear PDEs
The heat equation:

$$\partial_t v + \frac{1}{2} \Delta v = 0, \quad v(T,.) = g$$

has the following two possible probabilistic representations:

(i) $v(0,x) = \mathbb{E}[g(B_T)|B_0 = x]$; with $B$ a BM

(ii) $v(t,x) = e^{\beta T} \mathbb{E}[g(B_T) \mathbb{1}_{\{T < \tau\}}|B_0 = x]$; $\tau \sim \text{Exp}(\beta) \perp B$

Both representations are valid in the path-dependent case.
From linear representation (i) to nonlinear

Representation (i) extended by

- **BSDEs** (Pardoux & Peng, Bouchard & NT, Zhang, ...)

\[
dv_t = -F_t(v_t, \zeta_t)dt + \zeta_t dB_t, \quad v_T = g(B_T)
\]

- **2BSDEs** (Cheridito, Soner, NT & Victoire, Fahim, NT & Warin, Zhang & Zhuo, Possamaï & Tan)

\[
dv_t = -F_t(v_t, \zeta_t, \gamma_t)dt + \zeta_t dB_t, \quad d\zeta_t = \ldots dt + \gamma_t dB_t, \quad v_T = g(B_T)
\]

New formulation: Soner, NT & Zhang, and Possamaï, Tan & Zhou
$X^n$ : discrete-time approximation of diffusion with drift $\mu$ and diffusion $\sigma = 1$ (also $d = 1$ for simplicity)

$$
Y^n_{t_n} = g(X^n_{t_n}) ,
$$

$$
Y^n_{t_i-1} = \mathbb{E}^n_{i-1} [Y^n_{t_i}] + f\left(X^n_{t_i-1}, Y^n_{t_i-1}, Z^n_{t_i-1}, \Gamma^n_{t_i-1}\right) \Delta t_i , \hspace{1mm} 1 \leq i \leq n ,
$$

$$
Z^n_{t_i-1} = \mathbb{E}^n_{i-1} \left[Y^n_{t_i} \frac{\Delta W_{t_i}}{\Delta t_i}\right]
$$

$$
\Gamma^n_{t_i-1} = \mathbb{E}^n_{i-1} \left[Y^n_{t_i} \frac{\left|\Delta W_{t_i}\right|^2 - \Delta t_i}{|\Delta t_i|^2}\right]
$$

Then $Y^n_0 \longrightarrow \nu(0, x)$ as $n \to \infty$ + Error estimate...
Integration by parts

\[ \partial_x \mathbb{E}[\phi(X_t)] = \mathbb{E}\left[ \phi(X_t) \frac{W_h}{h} \right] \]

For simplicity, consider the one-dimensional case \( X_t = x + W_t \):

\[ \mathbb{E}[\phi_x(x + W_h)] = \int \phi_x(x + y) \frac{e^{-y^2/(2h)}}{\sqrt{2\pi}} dy \]

\[ = \int \phi_x(x + y) \frac{y e^{-y^2/(2h)}}{h \sqrt{2\pi}} dy \]

\[ = \mathbb{E}\left[ \phi(x + W_h) \frac{W_h}{h} \right] \]
From linear representation (ii) to nonlinear

- Consider KPP equations

\[ \partial_t v + \mu \cdot Dv + \frac{1}{2}\sigma^2 : D^2 v + \beta (\sum_{i=1}^{n} p_i v^i - v) = 0 \]

with \( p_i > 0 \) and \( \sum_{k=1}^{n} p_i = 1 \)

- Branching diffusions representation:

\[ v(0, x) = \mathbb{E}\left[ \prod_{k \in \mathcal{K}_T} g(Z^k_T) \right], \text{ where } Z^k : k-th \text{ particle} \]

and

\[ \mathcal{K}_t := \{\text{All particles alive at time } t\} \]

[Skorokhod, Watanabe, McKean]
Unbiased simulation of SDEs
Age-dependent branching diffusions and semilinear PDEs

Branching diffusion \( (n = 2) \)

Nonlinearity:

\[ \nabla \beta (a_2 v^2 + a_1 v - v) \]
Let $a_i(t, x)$ be bounded functions, and consider the PDE

$$\partial_t v + \mu(t, x) \cdot Dv + \frac{1}{2} \sigma^2(t, x) : D^2 v + \beta \left( \sum_{i=1}^{n} p_i a_i(t, x) v^i - v \right) = 0$$

$v(T, .) = g$

Introduce the branching diffusion:

- $(\tau_k)_k \text{ iid Expo}(\beta)$ : branching times
- $(I_k)_k \text{ iid Multinomial}(p_1, \ldots, p_n)$ : number of descendents
- Particle $k$ dies out at the branching event $T_k$, and $I_k$ independent particles follow the diffusion with drift and diffusion $(\mu, \sigma)$
The branching diffusion representation

Recall

- $\mathcal{K}_T := \{\text{particles present at } T\}$
- $\overline{\mathcal{K}}_T := \bigcup_{t \leq T} \mathcal{K}_t : \text{all particles}$

**Theorem (Henry-Labordère, Tan & NT SPA ’14)**

$v(0, x) = \mathbb{E}[\psi_{0,x}]$ where

$$\psi_{0,x} := \prod_{k \in \mathcal{K}_T} g(Z^k_T) \prod_{k \in \overline{\mathcal{K}}_T \setminus \mathcal{K}_T} a_{l_k}(T_k, Z^k_{T_k})$$

Moreover, this representation extends to the path-dependent case

- Numerical implications

- In the rest of the talk: extension to more general nonlinearities
BSDE representation : (backward) regression-based methods $\implies$

- no explosion restrictions
- High complexity, curse of dimension is back!
- Markovian feature is crucial

Branching diffusions $\implies$

- Purely forward Monte Carlo
- Suitable for path-dependency
- Very easy to implement, complexity linear in $d^2$
- Need to control from explosion of solution
- and of the variance (in the subsequent extensions)...
Main objective

- **Branching diffusion representation** for a larger class of PDEs (beyond KPP)

  Including nonlinearity in the gradient

- Unbiased simulation / Monte Carlo approximation

  Treat both Gradient and Hessian as nonlinearities...
Outline

1. Unbiased simulation of SDEs
   - The constant diffusion case
   - Regime switching and automatic differentiation

2. Age-dependent branching diffusions and semilinear PDEs
   - Complexity of the Monte Carlo approximation
Weak approximation of SDEs

Objective is to approximate \textbf{without discretization error}:

\[ V_0 := \mathbb{E}[g(X_T)] \]

where \( X \) is solution of the SDE

\[ dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t \]

- \( W \) is a Brownian motion
- \( \mu \) and \( \sigma \) satisfy the Lipschitz bounded, \( \sigma^{-1} \) bounded
- more conditions on \( \mu \) and \( \sigma \) will pop up
Our algorithm in the case of constant diffusion $\sigma = I_d (I)$

- $(N_t)$: Poisson process with intensity $\beta$, arrival times $(\tau_i)_{i \geq 1}$

- Set $\tau_0 := 0$, $T_i := \tau_i \wedge T$, and
  
  $$\Delta T_i := T_i - T_{i-1}, \quad \Delta W_{T_i} := W_{T_i} - W_{T_{i-1}}$$

- Consider the "Euler discretization along the arrival times $\tau_i$"

  $$\hat{X}_{T_i} = \hat{X}_{T_{i-1}} + \mu(T_{i-1}, \hat{X}_{T_{i-1}}) \Delta T_i + \Delta W_{T_i},$$

  for $i = 1, \ldots, N_T + 1$

  $\implies$ branching diffusion with one descendent at each default
Define the exactly simulatable r.v.

\[ \hat{\xi} := \beta^{-N_T} e^{\beta T} \left[ g(\hat{X}_T) - g(\hat{X}_{T_{N_T}}) \mathbb{I}_{\{N_T > 0\}} \right] \prod_{k=1}^{N_T} \hat{\mathcal{W}}^1_k \]

where

\[ \hat{\mathcal{W}}^1_k := \left( \mu(T_k, \hat{X}_{T_k}) - \mu(T_{k-1}, \hat{X}_{T_{k-1}}) \right) \cdot \frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}} \]

**Theorem (Henry-Labordère, Tan & NT '15)**

Assume \( \mu \frac{1}{2} - \)Hölder in \( t \), Lip in \( x \), and \( g \) Lipschitz. Then

\[ \hat{\xi} \in L^2 \quad \text{and} \quad \mathbb{E}[g(X_T)] = \mathbb{E}[\hat{\xi}] \]
Define

\[ \hat{X}_0 := X_0, \quad d\hat{X}_t = \mu(\Theta_t)dt + \sigma dW_t \]

with \( \Theta_t := (T_{N_t}, \hat{X}_{T_{N_t}}) \). In other words,

\[ \hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \int_{T_k}^{T_{k+1}} \mu(T_k, \hat{X}_{T_k}) ds + \int_{T_k}^{T_{k+1}} \sigma dW_s \]

i.e. the drift coefficient changes at each arrival time \( T_k \).
First main idea

Define \( u(t, x) := \mathbb{E}_{t,x}[g(X_T)], \ t \leq T, \ x \in \mathbb{R} \)

**Proposition**

Let \( \beta > 0, \ \theta \in [0, T) \times \mathbb{R}^d, \ (t, x) \in [0, T) \times \mathbb{R}^d \). Then

\[
u(t, x) = e^{\beta(T-t)}\mathbb{E}_{t,x,\theta} \left[ 1_{\{N_T = 0\}} g(\hat{X}_T) 
+ 1_{\{N_T > 0\}} \frac{1}{\beta} \Delta \mu \cdot Du(T_1, \hat{X}_{T_1}) \right]
\]

where \( \Delta \mu := \mu - \mu(\theta) \)
Sketch of proof of the lemma

The function $\tilde{u} := e^{-\beta(T-t)}E_{t,x}[g(X_T)]$ solves

$$-\partial_t \tilde{u} - \mu \cdot D\tilde{u} - \frac{1}{2} \sigma^2 : D^2 \tilde{u} + \beta \tilde{u} = 0 \quad \text{and} \quad \tilde{u}(T, .) = g$$

Equivalently, with $\phi := (\mu - \mu(\theta)) \cdot D\tilde{u}$,

$$-\partial_t \tilde{u} - \mu(\theta) \cdot D\tilde{u} - \frac{1}{2} \sigma^2 : D^2 \tilde{u} + \beta \tilde{u} = \phi \quad \text{and} \quad \tilde{u}(T, .) = g$$

By the Feynman-Kac representation:

$$u(0, X_0) = e^{\beta T}E\left[e^{-\beta T}g(\hat{X}_T) + \int_0^T e^{-\beta t} \phi(t, \hat{X}_t) dt\right]$$

$$= e^{\beta T}E\left[g(\hat{X}_T)I\{\tau \geq T\} + \frac{1}{\beta} \phi(\tau, \hat{X}_\tau)I\{\tau < T\}\right]$$

where $\tau$ is an independent Expo($\beta$)
By the last proposition,

\[ u(t, x) = \mathbb{E}_{t,x,\theta} \left[ e^{\beta(T_1-t)} \left( \mathbb{I}_{\{N_T=0\}} g(\hat{X}_T) 
\quad + \mathbb{I}_{\{N_T>0\}} \frac{\Delta \mu T_1}{\beta} \cdot Du(T_1, \hat{X}_{T_1}) \right) \right] \]

\[ = \mathbb{E}_{t,x,\theta} \left[ e^{\beta(T_1-t)} \left( \mathbb{I}_{\{N_T=0\}} g(\hat{X}_T) 
\quad + \mathbb{I}_{\{N_T=1\}} \frac{\Delta \mu T_1}{\beta} \cdot \frac{\Delta W_{T_2}}{\Delta T_2} g(\hat{X}_T) 
\quad + \mathbb{I}_{\{N_T>1\}} \frac{\Delta \mu T_2}{\beta^2} \cdot Du(T_2, \hat{X}_{T_2}) \right) \right] \]

by the assumption. And so on...
Back to unit diffusion: square integrability lost... in general

Iterating as above, and passing to limits, we would arrive at

$$\mathbb{E}[\xi] \quad \text{where} \quad \xi := \beta^{-NT} e^{\beta T} g(\hat{X}_T) \prod_{k=1}^{NT} \hat{W}_k^1$$

where, in the case of unit diffusion:

$$\hat{W}_k^1 := \left[ \mu(T_k, \hat{X}_{T_k}) - \mu(T_{k-1}, \hat{X}_{T_{k-1}}) \right] \cdot \frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}}$$

Notice that $\frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}} \sim (\Delta T_1)^{-1/2}$, then:

$$\mu \text{ Lip in } x, \quad \frac{1}{2} - \text{Hölder in } t \quad \Rightarrow \quad \hat{\xi} \in L^2$$
Gaining more integrability

Oudjane & Warin ’15 give up on the Expo(β) distribution \(\Rightarrow\) Age-dependent branching...

\[
-\partial_t u - \mu(\theta) \cdot Du - \frac{1}{2} \sigma^2 : D^2 u = \phi \quad \text{and} \quad u(T, .) = g
\]

with \(\phi := (\mu - \mu(\theta)) \cdot Du\). By the Feynman-Kac representation:

\[
u(0, X_0) = \mathbb{E}\left[ g(\hat{X}_T) + \int_0^T \phi(t, \hat{X}_t) dt \right]
\]

\[
= \mathbb{E}\left[ \bar{\rho}_T^{-1} g(\hat{X}_T) \mathbb{1}_{\{\tau \geq T\}} + \rho(\tau)^{-1} \phi(\tau, \hat{X}_T) \mathbb{1}_{\{\tau < T\}} \right]
\]

where \(\tau\) is an independent r.v. with density \(\rho\), and \(\bar{\rho}_T := \int_T^\infty \rho(t) dt\)

Choose \(\rho\) so as to guarantee square integrability...

**Gamma distribution does the job**

\[
\rho(t) = \Gamma(\kappa)^{-1} \beta^\kappa t^{\kappa-1} e^{-\beta t}, \quad \kappa \leq \frac{1}{2}
\]
Outline

1. Unbiased simulation of SDEs
   - The constant diffusion case
   - Regime switching and automatic differentiation

2. Age-dependent branching diffusions and semilinear PDEs
   - Complexity of the Monte Carlo approximation
Consider the PDE (unit diffusion for simplicity)

$$
\partial_t u + \frac{1}{2} \Delta u + f(t, x, u, Du) = 0, \quad u_T = g
$$

with nonlinearity

$$
f(t, x, y, z) = \sum_{(\ell_i)_{0 \leq i \leq n} \in L} p_\ell c_\ell(t, x) y^{\ell_0} \prod_{i=1}^{n} (b_i(t, x) \cdot z)^{\ell_i}
$$

- $L$ finite subset of $\mathbb{N}^{n+1}$
- $p_\ell > 0$ with $\sum_{\ell \in L} p_\ell = 1$
- $b_i(t, x)$ bounded functions

**Example**: Burgers equation $d = 1$ and $f(t, x, u, u_x) = u u_x$
Branching diffusion for the Burger equation

Burger's equation, nonlinearity:
\[ \beta(v v_x - v) \]
Marked branching diffusion representation

- \((\tau_k)_k\) iid arrival times, \(T_k := \tau_k \land T\)

- If \(T_1 < T\) : particle dies out, and is replaced with probability \(p_\ell\) by \(\ell_i\) particles of type \(i\), \(i = 0, \ldots, n\)

- For a particle \(k \in \overline{\mathcal{K}}_T\), denote by
  - \(D(k)\) its type
  - \(k^-\) its parent particle
  \[\implies\] Particle \(k\) lives between \(T_{k^-}\) and \(T_k\)
Using automatic differentiation

- Automatic differentiation:

\[
\mathcal{W}_k := \mathbb{I}_{\{D(k)=0\}} + \mathbb{I}_{\{D(k)\neq 0\}} b_{D(k)}(T_k, X_{T_k}^k) \cdot \frac{\Delta W_{T_k}}{\Delta T_k}
\]

The limiting random variable is:

\[
\psi := \prod_{k \in \mathcal{K}_T} \bar{F}_{\rho}(\Delta T_k)^{-1} \left[ g(X_{T_k}^k) - \mathbb{I}_{\{D_k \neq 0\}} g(X_{T_k}^k) \right] \mathcal{W}_k \\
\times \prod_{k' \in \overline{\mathcal{K}_T} \setminus \mathcal{K}_T} [\rho(\Delta T_{k'})]^{-1} b_{k'}(T_{k'}, X_{T_{k'}}^{k'}) \mathcal{W}_{k'}
\]
Sufficient condition for square integrability

For independent BM $W$, $\tau \sim \rho$, and $T_1 := \tau \wedge T$, define:

$$A_p := \max_{\ell} \frac{|g|^p_\infty \lor \|W_{T_1}\|^p \|b_\ell \cdot \frac{W_{T_1}}{T_1}\|^p}{\bar{F}_\rho(T)^{p-1}}$$

$$B_p := \max_{\ell} \|\Delta \tau\|^p/2 \|b_\ell \cdot \frac{W_{T_1}}{T_1}\|^p \left[|b_\ell|_\infty \sup_{t \leq T} \frac{t^{-\frac{p}{2(p-1)}}}{\rho(t)}\right]^{p-1}$$

Theorem (Henry-Labordère, Oudjane, Tan, NT, Warin '16)

Assume that $g$ Lipschitz and, for some $p > 1$,

$$\int_{A_p} \left[ B_p \sum |b_\ell|_\infty |x| |\ell|\right]^{-1} dx > T$$

Then $v(0, x) = \mathbf{E}_{0, x}[\psi]$, and $\psi \in \mathbb{L}^2$
The complexity of the algorithm

- **Average number of particles** in one simulation \( m(t) := \mathbb{E}[\#\mathcal{K}_t] \) satisfies, for \( n_0 := \sum_\ell \ell |p_\ell| \),

\[
m(t) = 1 + n_0 \int_0^t m(t-s)\rho(s)ds.
\]

When \( \rho \sim \Gamma(\kappa, \theta) \), one has

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

- **Number of v.a. simulated** for one particle

\[
d + 1 + 1
\]

- **Number of computation** for one particle

\[
Cd
\]
Numerical example

• Define

\[ u(t, x) = \cos(x_1 + \cdots + x_d) \exp(\alpha(T - t)) \]

is solution of semilinear PDE

\[ \partial_t u + \frac{1}{2} \Delta u + c u(b_1 \cdot Du) + b_0 = 0. \]

• For numerical implementation, we choose

\[ \alpha = 0.2, \quad c = 0.15, \quad b_1 = (1 + \frac{1}{d}, 1 + \frac{2}{d}, \cdots, 2). \]
A numerical example of dimension $d = 20$

**Figure:** Estimation and standard deviation observed in dimension $d = 20$ depending on the log of the number of simulation used.
Comments on the Monte-Carlo method

- Choice of $\rho, (p_\ell)_{\ell \in L}$, the expression of nonlinearity ...
- Possible to use importance sampling, particles method, ...
- Open to parallel computing
Fully nonlinear PDEs... (e.g. HJB equations)

If $T_1 < T$: particle dies out, and is replaced with probability $p_\ell$ by

- $i_\ell$ particles of type 0
- $j_\ell$ particles of type 1 $\implies$ first order differentiation weight
- $h_\ell$ particles of type 2 $\implies$ second order differentiation weight

Automatic differentiation for particles $k$ of type $D(k) = 2$:

$$\frac{\Delta W_T^2 - \Delta T}{(\Delta T)^2}$$
THANK YOU FOR YOUR ATTENTION