# Nesting Monte Carlo for high-dimensional Non Linear PDEs

X. Warin

EDF

5th september, CAESARS conference

# Outline

1. Introduction
2. Non linearity in $u$
3. Non linearity in $Du$
4. Towards full non linear equations

Resolution in high dimension (10-100) of

$$\begin{aligned} (-\partial_t u - \mathcal{L}u)(t,x) &= f(t,x,u(t,x),Du(t,x)), \\ u_T &= g, \end{aligned} \tag{1}$$

where

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

Parameters $\mu$, $\sigma$ may be time and space dependent.

# Recent bibliography

- Deterministic, dimension 3 or 4 for $x$ : based on grids ,
- Classic BSDE with Monte Carlo [13], [9]: dimension till 6 or 7. Based on grids or basis functions for regressions;
- Recent Deep Learning techniques [6], [5] : numerical results in dimension over 100. No proof of convergence. Limitation not understood. Based on forward simulation of BSDE discretized with an Euler scheme.
- Branching techniques
  - for polynomial drivers [11] :dimension over 100, but restriction on maturity and non linearity size. Variance explodes rapidly,
  - General drivers [2], [1], no real restriction on maturity but grids needed : limited in dimension 4 or 5.
- In [8], [7], [12], algorithm based on Picard iterations, multi-level techniques and automatic differentiation

# Challenge

## How to solve such equation in dimension over than 10 , 100 ?

- No grids,
- No basis functions

Use pure Monte carlo.

# Non linearity in $u$

$$(-\partial_t u - \mathcal{L}u)(t,x) = f(t,x,u(t,x)),$$
$$u_T = g, \tag{3}$$

where

$$\mathcal{L}u(t,x) := \mu Du(t,x) + \frac{1}{2}\sigma\sigma^\top : D^2u(t,x), \tag{4}$$

so that $\mathcal{L}$ is the generator associated to

$$\hat{X}_t = x + \mu t + \sigma dW_t, \tag{5}$$

# Idea of the algorithm : **Time discretization of the SDE using Poisson process.**

Same as branching method: i.i.d. $(\tau_m)_{m\geq 1}$ density $\rho$ exponential law with parameter $\lambda$, CDF : $F = 1 - \bar{F}$.

$$\begin{cases} T_0 &= 0, \\ T_{k+1} &= (T_k + \tau_k) \wedge T. \end{cases} \tag{6}$$

$(W_t^m)_{m\geq 1}$, independent of $(\tau_m)_{m\geq 1}$.

- $W_t = W_t^1$ for $t \in [0, T_1]$ ,
- $W_t := W_{T_k} + W_{t-T_k}^{k+1}$, for all $t \in [T_k, T_{k+1}]$.
- $X_0 = x$
- $X_t = X_{T_k} + \mu(t - T_k) + \sigma W_{t-T_k}^{k+1}$ $t \in [T_k, T_{k+1}]$, ¶-a.s.,

# Idea of the algorithm : **Use Feynman Kac**

Same as branching method:

$$
\begin{aligned}
u(0,x) =& \mathbb{E}_{0,x}\Big[g(X_T) + \int_0^T f(t, X_t, u(t, X_t))dt\Big] \\
=& \mathbb{E}_{0,x}\Big[\overline{F}(T)\frac{g(X_T)}{\overline{F}(T)} + \int_0^T \frac{f(t, X_t, u(t, X_t))}{\rho(t)}\rho(t)dt\Big] \\
=& \mathbb{E}_{0,x}\big[\phi\big(0, T_1, X_{T_1}, u(T_1, X_{T_1})\big)\big],
\end{aligned}
$$

$$
\phi(s,t,y,z) := \frac{\mathbf{1}_{\{t \geq T\}}}{\overline{F}(T-s)} g(y) + \frac{\mathbf{1}_{\{t < T\}}}{\rho(t-s)} f(t,y,z).
$$

# Idea of the algorithm : **Use nesting**

Recursively noting $u_n = u(T_n, X_{T_n})$ :

$$u_n = \mathbb{E}_{T_n, X_{T_n}}\big[\phi(T_n, T_{n+1}, X_{T_{n+1}}, u_{n+1})\big], \tag{7}$$

Truncated operator after $p$ switches :

$$\begin{aligned} u_p^p &= g(X_{T_p}), \\ u_n^p &= \mathbb{E}_{T_n, X^n_{T_n}}\big[\phi(T_n, T_{n+1}, X_{T_{n+1}}, u^p_{n+1})\big], \quad n < p, \qquad \text{defined if } T_n < T \end{aligned} \tag{8}$$

# Assumption

- $f$ is uniformly Lipschitz in $u$ with constant $K$ :

$$|f(t,x,y) - f(t,x,w)| \leq K|y-w| \quad \forall t \in \mathbb{R}^+, x \in \mathbb{R}^d, (w,y) \in \mathbb{R}^2. \tag{9}$$

- $f$ is uniformly Lipschitz in $x$.
- Solution $u \in C^{1,2}([0,T] \times \mathbb{R}^d)$
  - $u$ is $\theta$-Hölder with $\theta \in (0,1]$ in time with constant $\hat{K}$ :

$$|u(t,x) - u(\tilde{t},x)| \leq \hat{K}|t-\tilde{t}|^{\theta} \qquad \forall (t,\tilde{t},x) \in [0,T] \times [0,T] \times \mathbb{R}^d,$$

  - $u(t,x)$ has a quadratic growth in $x$ uniformly in $t$,

# Notations

- $(N_0, .., N_{p-1}) \in \mathbb{N}^p$, **number of particles at each level**
- **Set of all particles till switch** $i$ : $Q_i = \{k = (k_1, ..., k_i)\}$ for $i \in \{1, .., p\}$ where $k_j \in [1, N_{j-1}]$.
- **Successors of** $k$ : $\tilde{Q}(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i\}\} \subset Q_{i+1}$

**Switching dates**

$$\begin{cases} T_{(j)} & = & \tau_{(j)} \wedge T, j \in \{1, ., N_0\} \\ T_{\tilde{k}} & = & (T_k + \tau_{\tilde{k}}) \wedge T, k = (k_1, ..k_i) \in Q_i, \tilde{k} \in \tilde{Q}(k) \end{cases} \tag{10}$$

**SDE**

$$\begin{aligned} X_t^{(i)} &= X_0^{\emptyset} + \mu t + \sigma \bar{W}_t^{(i)}, \quad t \in [0, T_{(i)}], i = 1, N_0 \\ X_t^{\tilde{k}} &= X_{T_k}^{k} + \mu(t - T_k) + \sigma \bar{W}_{t-T_k}^{\tilde{k}}, \text{ for } \tilde{k} \in \tilde{Q}(k), \quad t \in [T_k, T_{\tilde{k}}], \quad \P\text{-a.s.}, \end{aligned} \tag{11}$$

# Estimator

$$
\begin{cases}
\bar{u}^p_{\emptyset} = & \frac{1}{N_0}\sum_{j=1}^{N_0} \phi(0, T_{(j)}, X^{(j)}_{T_{(j)}}, \bar{u}^p_{(j)}), \\
\bar{u}^p_{k} = & \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} \phi(T_k, T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}), \\
 & \quad \text{for } k = (k_1, ..k_i) \in Q_i, 1 < i < p, T_k < T, \\
\bar{u}^p_{\tilde{k}} = & g(X^{\tilde{k}}_{T_{\tilde{k}}}) \qquad \text{for } \tilde{k} \in Q_p.
\end{cases}
$$

# Convergence

## Proposition

*Under previous assumptions*

$$\mathbb{E}((\bar{u}^p_\emptyset - u(0,x))^2) \leq \prod_{i=1}^{p}(1+\frac{8}{N_{i-1}})\frac{K^{2p}e^{\lambda T}}{\lambda^p}T^{2\theta}\hat{K}^2\frac{T^p}{p\Gamma(p)} + \sum_{i=0}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\frac{T^i e^{\lambda T}}{\lambda^i}\kappa_i \tag{12}$$

*with*

$$\kappa_i = (\frac{4T}{\lambda(i+1)!}\sup_{t\in[0,T]}\mathbb{E}(f(t,X_t,u(t,X_t))^2) + 2\mathbb{E}(g(X_T)^2))(\frac{1_{i>0}}{i\Gamma(i)} + 1_{i=0}))$$

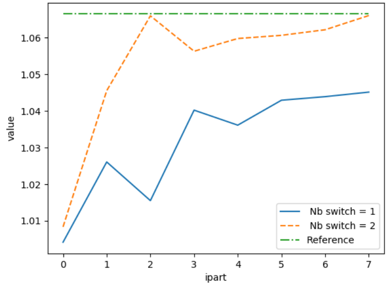
*and* $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ *is the gamma function.*

# Numerical results for first non linear case $y^2$ dim 100

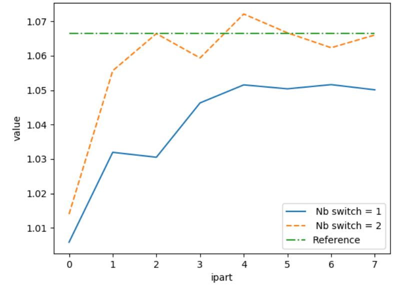
- $T = 1$, $T = 2$,
- $\mu = \frac{\mu_0}{d} \mathbb{1}_d$, $\sigma = \frac{\sigma_0}{\sqrt{d}} \mathbf{I}_d$ with $\mu_0 = 0.2$, $\sigma_0 = 1$.
- $g(x) = \cos(\sum_{i=1}^{d} x_i)$
- $$f(t,x,u) = \cos(\sum_{i=1}^{d} x_i)(a + \frac{\sigma_0^2}{2})e^{a(T-t)} + \sin(\sum_{i=1}^{d} x_i)\mu_0 e^{a(T-t)} - \\ r\cos(\sum_{i=1}^{d} x_i)^2 e^{2a(T-t)} + r(-e^{a(T-t)} \vee (u \wedge e^{a(T-t)}))^2$$

  with $a = 0.1$, $r = 0.1$.
- $u(t,x) = e^{a(T-t)} cos(\sum_{i=1}^{d} x_i)$.

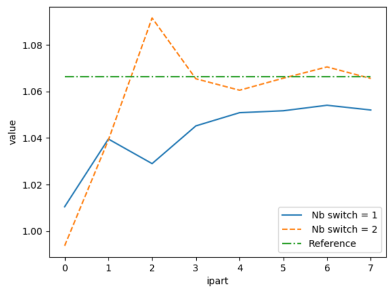
# $T = 1$




$\lambda = 0.2.$



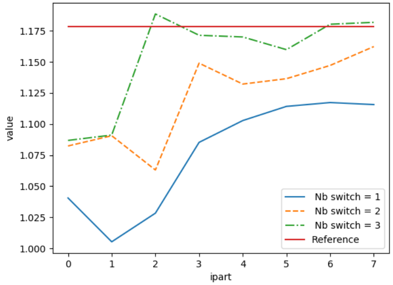

$\lambda = 0.4.$



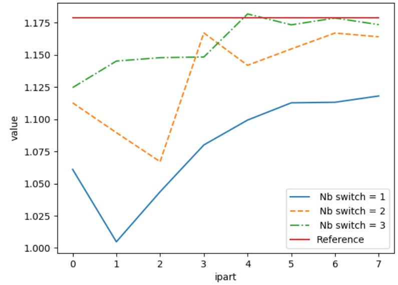

$\lambda = 0.8.$

Figure 1: Convergence for different number of switches for case 1, $T = 1$, $N_0 = 1000 \times 2^{\text{ipart}}$ and $N_1 = 50 \times 2^{\text{ipart}}$

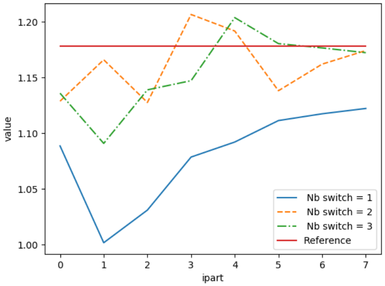
# $T = 2$




$\lambda = 0.2.$




$\lambda = 0.4.$




$\lambda = 0.8.$

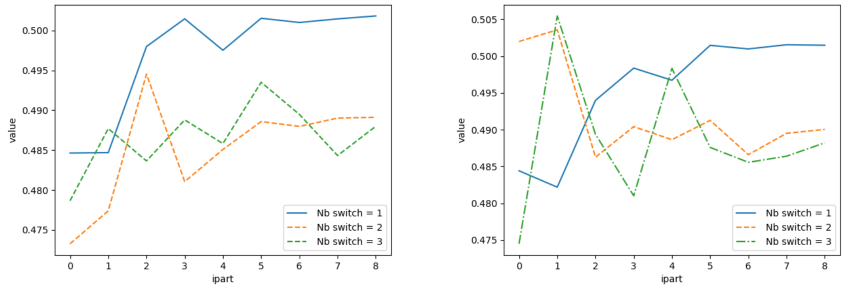
Figure 2: Convergence for different number of switches for case 1, $T = 2$, $N_0 = 1100 \times 2^{\text{ipart}}$, $N_1 = 110 \times 2^{\text{ipart}}$, $N_2 = 25 \times 2^{\text{ipart}}$

# CVA test case [10] Dim 6

- $\mu = -\frac{\sigma_0^2}{2}\mathbb{1}_d$, $\sigma = \sigma_0 \mathbf{I}_d$,
- $$f(t, x, u) = \beta(u^+ - u),$$

  $\beta = 0.03, \sigma_0 = 0.2.$
- $X_0 = \mathbb{1}_d$ .
- $g(x) = \sum_{i=1}^d (1 - 2 1_{e^{x_i} > 1})$
- $T = 1.$

Reference deep learning primal and dual : lower bound 48.80, upper bound 48.83, value BS 47.73.




$\lambda = 0.1.$ $\lambda = 0.2.$

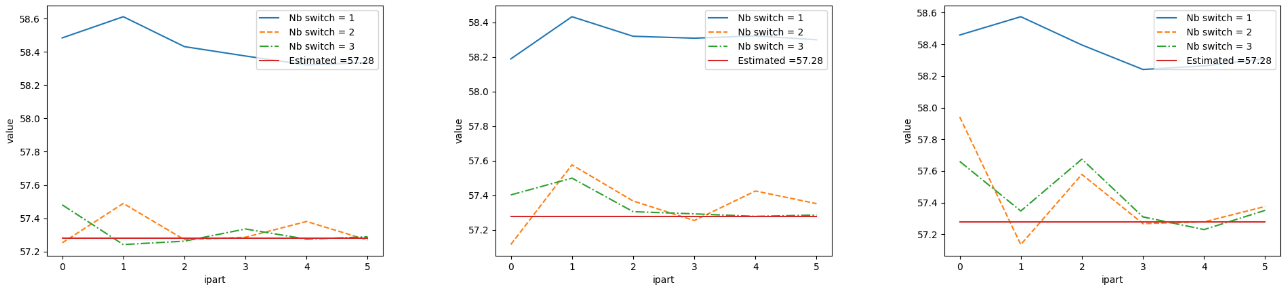
Figure 3: Convergence of the scheme on the CVA case.

- $\lambda = 0.1$ 3 switches and $ipart = 8$ we get 0.4880,
- $\lambda = 0.2$ we get 0.4882.

# BS default dim 100 risk [6]

- $\mu = (\mu_0 - \frac{\sigma_0^2}{2})\mathbb{I}_d$, $\sigma = \sigma_0 \mathbf{I}_d$.
- $g(x) = \min_{i=1}^{100}(e^{x_i})$,
-
$$f(t,x,u) = -\left((1-\delta)\min\{\gamma^h, \max\{\gamma^l, \frac{\gamma^h - \gamma^l}{v^h - v^l}(u - v^h) + \gamma^h\}\} + R\right) u$$
- $T = 1$, $\delta = \frac{2}{3}$, $\mu_0 = 0.02$, $\sigma_0 = 0.2$, $v^h = 50$, $v^l = 70$, $\gamma^h = 0.2$, $\gamma^l = 0.02$.

# Results Jentzen case




$\lambda = 0.2.$ $\lambda = 0.4.$ $\lambda = 0.8.$

Figure 4: Convergence of the scheme on the Black Scholes case with default $N_0 = 36000 \times 2^{\text{ipart}}$, $N_1 = 40 \times 2^{\text{ipart}}$, $N_2 = 2^{\text{ipart}}$

A good accuracy 57.28 is achieved with two switches taking $N_0 = 1152000$, $N_1 = 4480$ in 26 seconds (224 cores) with $\lambda = 0.2$.
Reference seems to be 57.30 according [6].

# Non linearity in $Du$

$$
\begin{aligned}
(-\partial_t u - \mathcal{L}u)(t,x) &= f(t,x,Du(t,x)), \\
u_T = g, &\quad t < T,\ x \in \mathbb{R}^d,
\end{aligned}
\tag{13}
$$

Assumptions:

- $\sigma$ non degenerated matrix,
- $f$ uniformly Lipschitz in $x$ and in $y$:

$$
|f(t,x,y) - f(t,x,w)| \leq K||y-w||_2, \forall t \in [0,T], x \in R^d, (w,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{14}
$$

- $u \in C^{1,2}([0,T] \times \mathbb{R}^d)$,
  - $Du$ is $\theta$-Hölder in time with $\theta \in (0,1]$ :

$$
||Du(t,x) - Du(\tilde{t},x)|| \leq \hat{K}|t-\tilde{t}|^{\theta} \quad \forall (t,\tilde{t},x) \in [0,T] \times [0,T] \times \mathbb{R}^d,
$$

  - $u(t,x)$ and $Du(t,x)$ have a quadratic growth in $x$ uniformly in $t$.
- $g$ Lipschitz

$$
|g(x) - g(y)| \leq \tilde{K}||x-y||_2 \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d.
$$

# General idea

$(\tau_m)_{m\geq 1}$ with gamma distribution $0 < u < 1$, $\rho(x) = \lambda^u x^{u-1}\frac{e^{-\lambda x}}{\Gamma(u)}$. Same as before

$$u(0,x) = \mathbb{E}_{0,x}\Big[\overline{F}(T)\frac{g(W_T)}{\overline{F}(T)} + \int_0^T \frac{f(t,X_t,Du(t,X_t))}{\rho(t)}\rho(t)dt\Big]$$
$$= \mathbb{E}_{0,x}\big[\hat{\phi}(0,T_1,X_{T_1},Du(T_1,X_{T_1}))\big], \tag{15}$$

with $\hat{\phi}(s,t,x,z) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)}g(x) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)}f(t,x,z)$.

Define $Du(T_1,X_{T_1})$ using the automatic differentiation rule :

$$Du(T_1,X_{T_1}) = \mathbb{E}_{T_1,X_{T_1}}\Big[\sigma^{-\top}\frac{W_{T_2}-W_{T_1}}{T_2-T_1}\phi(T_1,T_2,X_{T_1},X_{T_2},Du(T_2,X_{T_2}))\Big], \tag{16}$$

with $\phi(s,t,x,y,z) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)}(g(y)-g(x)) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)}f(t,y,z)$.

# A first estimator

$$
\begin{cases}
\bar{u}^p_{\emptyset} = & \frac{1}{N_0}\sum_{j=1}^{N_0} \hat{\phi}(0, T_{(j)}, X^{(j)}_{T_{(j)}}, D\bar{u}^p_{(j)}), \\
D\bar{u}^p_k = & \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} \phi(T_k, T_{\tilde{k}}, X^k_{T_k}, X^{\tilde{k}}_{T_{\tilde{k}}}, D\bar{u}^p_{\tilde{k}})\sigma^{-\top}\frac{\bar{W}^{\tilde{k}}_{T_{\tilde{k}}-T_k}}{T_{\tilde{k}}-T_k} \\
& \quad \text{for } k = (k_1, k_2, ..k_i) \in Q_i, i < p, \\
D\bar{u}^p_{\tilde{k}} = & Dg(X^{\tilde{k}}_{T_{\tilde{k}}}) \quad \text{for } \tilde{k} \in Q_p
\end{cases}
$$

# Convergence

## Proposition

*Under previous assumptions*

$$\mathbb{E}\Big((\bar{u}^p_\emptyset - u(0,x))^2\Big) \leq \prod_{i=1}^{p}(1+\frac{8}{N_{i-1}})\frac{\Gamma(u)^p e^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1+2\theta}}{(1-u)^{p-1}(2-u)}C(\sigma)^{p-1}\hat{K}^2 K^{2p}+$$

$$4\sum_{i=0}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^{i+1}}\frac{T^{(1-u)(i+1)+1}}{(1-u)^i(2-u)}\tilde{C}(\sigma)^i\hat{F}+$$

$$2\sum_{i=1}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^i}\frac{T^{(1-u)i+1}}{(1-u)^{i-1}(2-u)}\frac{\bar{C}(\mu,\sigma,T)C(\sigma)^{i-1}\tilde{K}^2}{\Gamma(u)-\gamma(u,\lambda T)}+$$

$$\frac{2}{N_0}\frac{\Gamma(u)}{\Gamma(u)-\gamma(u,\lambda T)}\mathbb{E}(g(X_T)^2)$$

*where*

$$\hat{F} = \sup_{t\in[0,T]}\mathbb{E}[f(t,X_t,Du(t,X_t))^4]^{\frac{1}{2}},$$

# Original particle and "antithetic" by example : $d = 1$, $\mu = 0$, $\sigma = 1$, the brownian case.

Let us consider the original particle $k = (k_1, k_2, k_3)$ such that $T_{(k_1,k_2,k_3)} = T$

$$
\begin{aligned}
X^{(k_1)}_{T_{(k_1)}} &= \bar{W}^{(k_1)}_{T_{(k_1)}}, \quad X^{(k_1^{(-)})}_{T_{(k_1)}} = -\bar{W}^{(k_1)}_{T_{(k_1)}} \\
X^{(k_1,k_2,k_3)}_{T} &= \bar{W}^{(k_1)}_{T_{(k_1)}} + \bar{W}^{(k_1,k_2)}_{T_{(k_1,k_2)}-T_{(k_1)}} + \bar{W}^{(k_1,k_2,k_3)}_{T-T_{(k_1,k_2)}} \\
X^{(k_1^-,k_2,k_3)}_{T} &= -\bar{W}^{(k_1)}_{T_{(k_1)}} + \bar{W}^{(k_1,k_2)}_{T_{(k_1,k_2)}-T_{(k_1)}} + \bar{W}^{(k_1,k_2,k_3)}_{T-T_{(k_1,k_2)}} \\
X^{(k_1,k_2^-,k_3)} &= \bar{W}^{(k_1)}_{T_{(k_1)}} - \bar{W}^{(k_1,k_2)}_{T_{(k_1,k_2)}-T_{(k_1)}} + \bar{W}^{(k_1,k_2,k_3)}_{T-T_{(k_1,k_2)}} \\
X^{(k_1,k_2^-,k_3^-)} &= \bar{W}^{(k_1)}_{T_{(k_1)}} - \bar{W}^{(k_1,k_2)}_{T_{(k_1,k_2)}-T_{(k_1)}} - \bar{W}^{(k_1,k_2,k_3)}_{T-T_{(k_1,k_2)}} \\
&\dots
\end{aligned}
$$

- for $f$ regular $f(X^{(k_1,k_2,k_3)}_T) - f(X^{(k_1^-,k_2,k_3)}_T) = O(\sqrt{T_{(k_1)}})$
- Notation antithetic $(k_1, k_2)^- = (k_1, k_2^-)$, $(k_1^-, k_2)^- = (k_1^-, k_2^-)$,
- Notation original particle $o((k_1, k_2^-, k_3^-)) = (k_1, k_2, k_3)$.

# A second framework : general notations

- $Q_1^o = \{(k_1), (k_1^-)\}$ where $k_1 \in \{1, .., N_1\}$,
- To a particle $(k_1) \in Q_1$ associate an antithetic particle noted $k_1^-$.
- The set $Q_i^o$ defined by recurrence :

$$Q_{i+1}^o = \{(k_1, .., k_i, k_{i+1})/(k_1, .., k_i) \in Q_i^o, k_{i+1} \in \{1, .., N_{i+1}, 1^-, .., N_{i+1}^-\}\}$$

- $k = (k_1, .., k_i) \in Q_i^o$ its original particle $o(k) = (\hat{k}_1, ..\hat{k}_i)$ where $\hat{k}_j = l$ if $k_j = l$ or $l^-$
- when $k = (k_1, \cdots, k_i)$ is such that $k_i \in \mathbb{N}$, $k^- := (k_1, \cdots, k_{i-1}, k_i^-)$.
- By convention $T_k = T_{o(k)}$, $\tau_k = \tau_{o(k)}$ and $\bar{W}_t^k = \bar{W}_t^{o(k)}$. For $k = (k_1, ..., k_i) \in Q_i^o$ we introduce the set
  - $\tilde{Q}^o(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i\}\} \subset Q_{i+1}^o$
  - and $\hat{Q}^o(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i, 1^-, .., N_i^-\}\} \subset Q_{i+1}^o$

# SDE scheme

$k = (k_1, .., k_i) \in Q_i^o$ and $\tilde{k} = (k_1, .., k_i, k_{i+1}) \in \hat{Q}^o(k)$ we define the following trajectories :

$$W_s^{\tilde{k}} := W_{T_k}^k + \mathbf{1}_{k_{i+1} \in \mathbb{N}} \bar{W}_{s-T_k}^{o(\tilde{k})} - \mathbf{1}_{k_{i+1} \notin \mathbb{N}} \bar{W}_{s-T_k}^{o(\tilde{k})}, \quad \text{and} \tag{17}$$

$$X_s^{\tilde{k}} := x + \mu s + \sigma W_s^{\tilde{k}}, \quad \forall s \in [T_k, T_{\tilde{k}}]. \tag{18}$$

# Add some assumptions

- $Du$ is uniformly Lipschitz in $x$ such that for $\bar{K} > 0$:

  $$||Du(t,x) - Du(t,y)||_2 \leq \bar{K}||x-y||_2 \quad \forall (t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d.$$

- $f$ is uniformly Lipschitz in $x$ :

  $$|f(t,x,z) - f(t,y,z)| \leq \underline{\mathrm{K}}||x-y||_2, \quad \forall (t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$$

# Estimator

$$\begin{cases} \bar{u}_0^p = & \frac{1}{N_0}\sum_{j=1}^{N_0} \frac{\left(\hat{\phi}\left(0, T_{(j)}, X_{T_{(j)}}^{(j)}, D\bar{u}_{(j)}^p\right) + \hat{\phi}\left(0, T_{(j)}, X_{T_{(j)}}^{(j)-}, D\bar{u}_{(j)-}^p\right)\right)}{2}\Big), \\ D\bar{u}_k^p = & \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}^o(k)} \tilde{W}^{\tilde{k}} \frac{1}{2}\left(\hat{\phi}\left(T_k, T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}}, D\bar{u}_{\tilde{k}}^p\right) - \hat{\phi}\left(T_k, T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}^-}, D\bar{u}_{\tilde{k}-}^p\right)\right), \\ & \quad \text{for } k = (k_1, ..k_i) \in Q_i^o, i < p, \\ D\bar{u}_{\tilde{k}}^p = & Dg(X_{T_{\tilde{k}}}^{\tilde{k}}) \quad \text{for } \tilde{k} \in Q_p^o \end{cases}$$

where

- $\hat{\phi}(s, t, x, z) := \frac{\mathbf{1}_{\{t \geq T\}}}{\overline{F}(T-s)} g(x) + \frac{\mathbf{1}_{\{t < T\}}}{\rho(t-s)} f(t, x, z).$
- $\tilde{W}^{\tilde{k}} = \sigma^{-\top} \frac{\bar{W}^{o(\tilde{k})}_{T_{\tilde{k}} - T_k}}{T_{\tilde{k}} - T_k}$ .

# Convergence

## Proposition

*Under previous assumptions:*

$$\mathbb{E}\Big((\bar{u}_{\emptyset}^{p} - u(0,x))^2\Big) \leq \prod_{i=1}^{p}(1 + \frac{8}{N_{i-1}})\frac{\Gamma(u)^p e^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1+2\theta}}{(1-u)^{p-1}(2-u)}C(\sigma)^{p-1}\hat{K}^2 K^{2p} +$$
$$\sum_{i=1}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\bar{C}(\sigma, K, \bar{K}, \underline{K})C(\sigma)^{i-1}\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^{i+1}}\frac{T^{(1-u)i+3-u}}{(2-u)^2(1-u)^{i-1}} +$$
$$\sum_{i=1}^{p-1}\frac{K^{2i}}{2N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})C(\sigma)^{i-1}\frac{\tilde{K}^2\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^i(\Gamma(u)-\gamma(u,\lambda T))}\bar{C}(\sigma)\frac{T^{(1-u)i+1}}{(2-u)(1-u)^{i-1}} +$$
$$\frac{4}{N_0}\frac{\Gamma(u)}{\lambda}e^{\lambda T}\frac{T^{2-u}}{2-u}\hat{F} + \frac{2}{N_0}\frac{\Gamma(u)}{\Gamma(u)-\gamma(u,\lambda T)}\mathbb{E}(g(X_T)^2) \qquad (19)$$

Using two switches, the use of an exponential law is possible.

# Numerical results on a Bürgers case [5], [3]

- $\mu = 0$, $\sigma = d\mathbf{I}_d$, $T = 1$,
- $$f(t, x, y, z) = (y - \frac{2+d}{2d})(d\sum_{i=1}^{d} z_i),$$
- $$g(x) = \frac{e^{T + \frac{1}{d}\sum_{i=1}^{d} x_i}}{1 + e^{T + \frac{1}{d}\sum_{i=1}^{d} x_i}}.$$

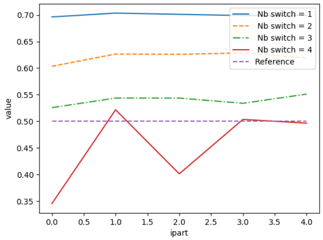
  The explicit solution given by [5] is

  $$u(t, x) = \frac{e^{t + \frac{1}{d}\sum_{i=1}^{d} x_i}}{1 + e^{t + \frac{1}{d}\sum_{i=1}^{d} x_i}}.$$

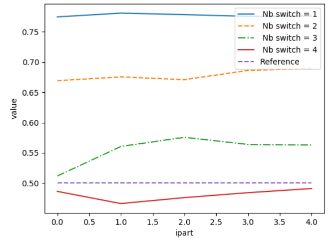
  Solution estimated by the two estimators by

  $$N_i^{ipart} = N_i^0 \times 2^{ipart}.$$

# Bürgers, estimator 1 , reference 0.5, $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 30)$

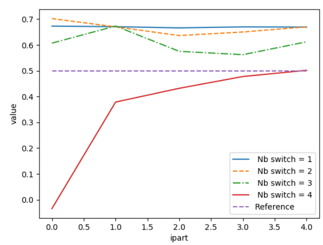



$\lambda = 0.1.$

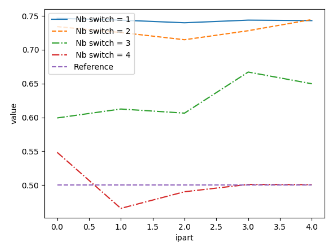



$\lambda = 0.2.$

Bürgers case: convergence in dimension 10 using a gamma law with $u = 0.8$




$\lambda = 0.1$




$\lambda = 0.2$

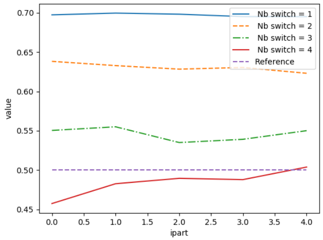
Bürgers case: convergence with estimator 1 in dimension 20 using a gamma law with $u = 0.8$.

# Bürgers estimator 1, time for 224 cores

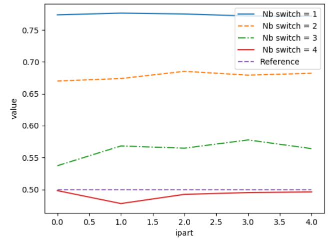
In dimension 10 and 20, we obtain very good results with 4 switches and $ipart = 4$,

- getting in dimension 10 a value 0.496 for $\lambda = 0.1$ in 80 seconds and 0.4910 for $\lambda = 0.2$ in 1000 seconds,
- getting in dimension 20 a value 0.501 for $\lambda = 0.1$ in 350 seconds and 0.5006 for $\lambda = 0.2$ in 1400 seconds.

# Bürgers : estimator 2, reference 0.5, gamma law, $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 30)$

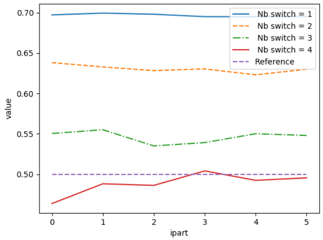



$\lambda = 0.1.$

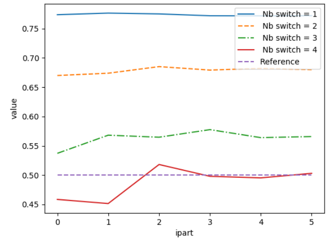



$\lambda = 0.2.$

Bürgers case: convergence with estimator 2 in dimension 10 using a gamma law with $u = 0.9$.



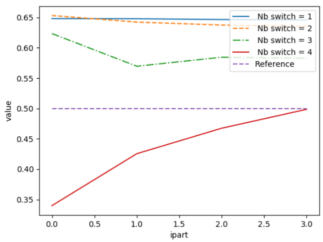

$\lambda = 0.1.$



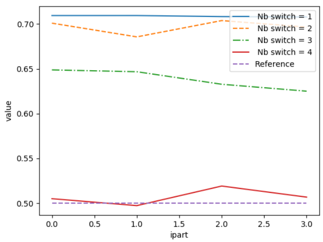

$\lambda = 0.2.$

Bürgers case: convergence with estimator 2 in dimension 20 with a gamma law with $u = 0.9$.

# Estimator 2 and exponential law, $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 4)$, time for 224 cores




$\lambda = 0.1.$




$\lambda = 0.2.$

Bürgers case: convergence with estimator 2 in dimension 20 with an exponential law, time for 224 cores

With 4 switches, $ipart = 3$ we get very good results :

- Solution 0.499, computational time 14 seconds with $\lambda = 0.1$,
- Solution 0.506 , computational time 55 seconds with $\lambda = 0.2$.

The variance using exponential laws seems to be lower and generation of an exponential law takes far less time than with general gamma laws.

# A HJB case in dimension 100 [5],[6], [4]

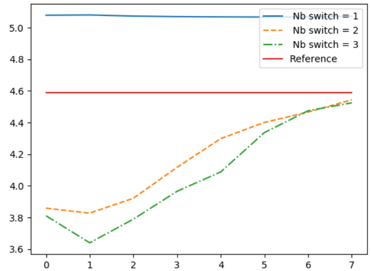
- $$\begin{aligned}\mu =&0, \quad \sigma = \sqrt{2}\mathbf{I}_d, \quad T = 1,\\ f(t,x,z) =& -\theta ||z||_2^2,\end{aligned}$$

- $$u(t,x) = -\frac{1}{\theta} \log \left( \mathbb{E}[e^{-\theta g(x+\sqrt{2}W_{T-t})}] \right). \tag{20}$$

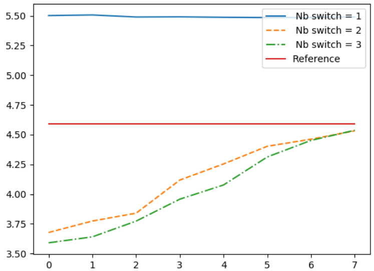
- $$g(x) = \log(\frac{1 + ||x||_2^2}{2}),$$

- solution searched for $t = 0$, $x = 0 \mathbb{1}_d$.
- References 4.59 with $\theta = 1$, 4.49 with $\theta = 10$, and 4.36 with $\theta = 20$

# Results with estimator 1 $(N_0^0, N_1^0, N_2^0) = (1000, 20, 20)$




$\lambda = 0.1$

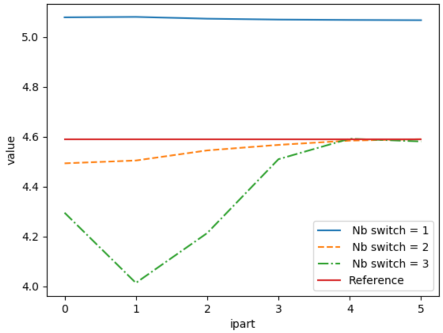



$\lambda = 0.2$

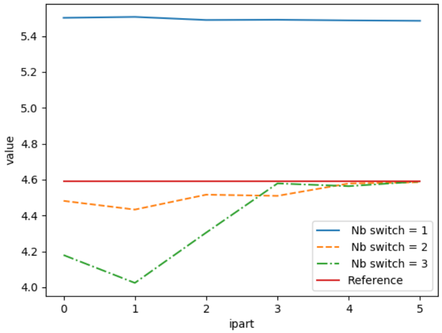
HJB convergence case for $\theta = 1$ and estimator 1 using a gamma law with $u = 0.8$.

Very slow convergence with this non linearity. Estimator 1 too costly for $\theta = 10$, $\theta = 20$.

# Estimator 2 , gamma law , $\theta = 1$, $(N_0^0, N_1^0, N_2^0) = (1000, 10, 1)$



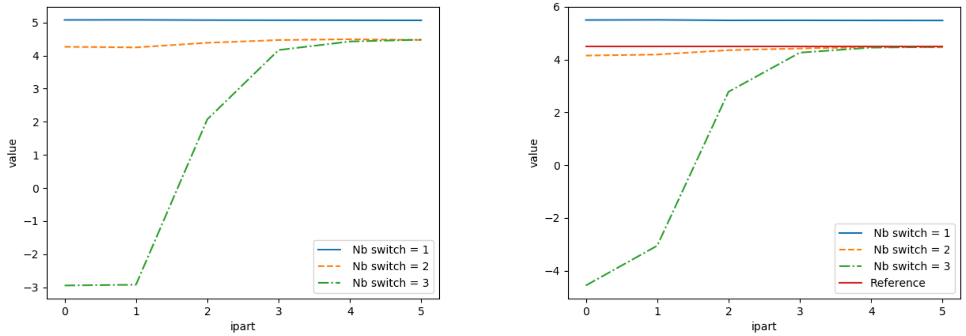

$\lambda = 0.1$




$\lambda = 0.2$

Figure 5: HJB convergence case for $\theta = 1$ and estimator 2 using a gamma law with $u = 0.9$.

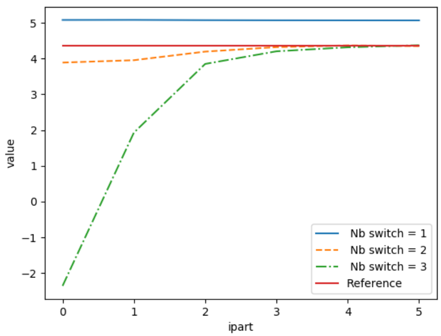
# Estimator 2 , gamma law , $\theta = 10$, $(N_0^0, N_1^0, N_2^0) = (1000, 10, 5)$



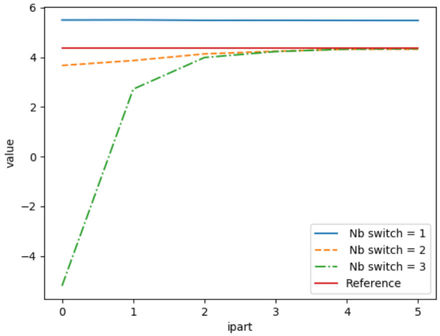

$\lambda = 0.1$ $\lambda = 0.2$

Figure 6: HJB convergence case for $\theta = 10$ and estimator 2 using a gamma law with $u = 0.9$.

# Estimator 2 , gamma law , $\theta = 20$, $(N_0^0, N_1^0, N_2^0) = (1000, 40, 20)$



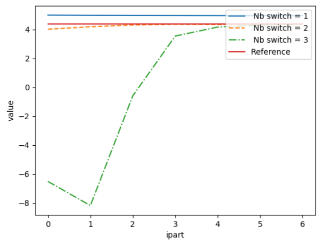

$\lambda = 0.1$



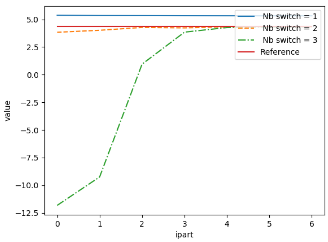

$\lambda = 0.2$

Figure 7: HJB convergence case for $\theta = 20$ and estimator 2 using a gamma law with $u = 0.9$.

# Estimator 2 , exponential law, $\theta = 20$, $(N_0^0, N_1^0, N_2^0) = (1000, 40, 10)$




$\lambda = 0.1$




$\lambda = 0.2$

HJB convergence case for $\theta = 20$ and estimator 2 using an exponential law.

Solution in 10 seconds, $\lambda = 0.1$, 2 switches, precision less than 0.5% using $(N_0, N_1) = (8000, 320)$.

Very accurate solution, precision 0.1% with both $\lambda$ necessary to use 3 switches with $(N_0, N_1, N_2) = (64000, 2560, 640)$ and computational time explodes to 30000 seconds with $\lambda = 0.1$ and 80000 seconds with $\lambda = 0.2$.

# The general problem Full Non Linear problem

$$(-\partial_t u - \mathcal{L}u)(t,x) = f(t,x,u(t,x),Du(t,x),D^2u(t,x)),$$
$$u(T,x) = g(x), \quad t < T,\ x \in \mathbb{R}^d, \tag{21}$$

- $\mu \in \mathbb{R}^d$, and $\sigma \in \mathbb{M}^d$ is some constant matrix.
-

$$\rho(x) = \lambda^\alpha x^{\alpha-1} \frac{e^{-\lambda x}}{\Gamma(\alpha)}, 1 \geq \alpha > 0. \tag{22}$$

# Notations

- $$Q_1^o = Q_1$$

  $$Q_{i+1}^o = \{(k_1, .., k_i, k_{i+1})/(k_1, .., k_i) \in Q_i^o, k_{i+1} \in \{1, .., N_{i+1}, 1_1, .., (N_{i+1}$$

- For $k = (k_1, .., k_i) \in Q_i^o$ its original particle $o(k) \in Q_i$ such that $o(k) = (\hat{k}_1, ..\hat{k}_i)$ where $\hat{k}_j = l$ if $k_j = l$, $l_1$ or $l_2$.
- For $k = (k_1, ..., k_i) \in Q_i^o$ set of its non fictitious sons

  $$\tilde{Q}(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i\}\} \subset Q_{i+1}^o,$$

- $$\begin{aligned} \kappa(k) =& 0 \text{ for } k_i \in \mathbb{N}, \\ \kappa(k) =& 1 \text{ for } k_i = l_1, l \in \mathbb{N} \\ \kappa(k) =& 2 \text{ for } k_i = l_2, l \in \mathbb{N} \end{aligned}$$

# Switching dates and trajectories

$$\begin{cases} T_{(j)} & = & \tau_{(j)} \wedge T, j \in \{1, ., N_0\} \\ T_{\tilde{k}} & = & (T_k + \tau_{\tilde{k}}) \wedge T, k = (k_1, ..k_i) \in Q_i, \tilde{k} \in \tilde{Q}(k) \end{cases} \tag{23}$$

$$W_s^{\tilde{k}} \ := W_{T_k}^k \ + \ \mathbf{1}_{\kappa(\tilde{k})=0} \bar{W}_{s-T_k}^{o(\tilde{k})} \ - \ \mathbf{1}_{\kappa(\tilde{k})=1} \bar{W}_{s-T_k}^{o(\tilde{k})}, \quad \text{and} \tag{24}$$

$$X_s^{\tilde{k}} := x + \mu s + \sigma W_s^{\tilde{k}}, \quad \forall s \in [T_k, T_{\tilde{k}}], \tag{25}$$

# Example: $d = 1$, $\mu = 0$, $\sigma = 1$

Let us consider the original particle $k = (1, 1, 1)$ such that $T_{(1,1,1)} = T$

$$
\begin{aligned}
X_T^{(1,1,1)} &= \bar{W}^{(1)}_{T_{(1)}} + \bar{W}^{(1,1)}_{T_{(1,1)}-T_{(1)}} + \bar{W}^{(1,1,1)}_{T-T_{(1,1)}} \\
X_T^{(1_1,1,1)} &= -\bar{W}^{(1)}_{T_{(1)}} + \bar{W}^{(1,1)}_{T_{(1,1)}-T_{(1)}} + \bar{W}^{(1,1,1)}_{T-T_{(1,1)}} \\
X^{(1,1_1,1)} &= \bar{W}^{(1)}_{T_{(1)}} - \bar{W}^{(1,1)}_{T_{(1,1)}-T_{(1)}} + \bar{W}^{(1,1,1)}_{T-T_{(1,1)}} \\
X_T^{(1_2,1_1,1)} &= -\bar{W}^{(1,1)}_{T_{(1,1)}-T_{(1)}} + \bar{W}^{(1,1,1)}_{T-T_{(1,1)}} \\
&\dots
\end{aligned}
$$

- For $f$ regular, $f(X_T^{(1,1,1)}) + f(X_T^{(1_1,1,1)}) - 2f(X_T^{(1_2,1,1)}) = O(T_{(1)})$.
- Notation $(1, 1, 1)^1 = (1, 1, 1_1)$, $(1, 1, 1)^2 = (1, 1, 1_2)$.

# Weights and $\phi$ function

$$\phi(s,t,x,y,z,\theta) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)} g(x) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)} f(t,x,y,z,\theta), \tag{26}$$

and gradiant weight:

$$\mathbb{V}^k = \sigma^{-\top} \frac{\bar{W}^k_{T_k - T_{k^-}}}{T_k - T_{k^-}},$$

second order derivative weight:

$$\mathbb{W}^k = (\sigma^\top)^{-1} \frac{\bar{W}^k_{T_k - T_{k^-}} (\bar{W}^k_{T_k - T_{k^-}})^\top - (T_k - T_{k^-}) I_d}{(T_k - T_{k^-})^2} \sigma^{-1}. \tag{27}$$

# The scheme

$$
\begin{cases}
\bar{u}^p_{\emptyset} = \dfrac{1}{N_0} \displaystyle\sum_{j=1}^{N_0} \phi\Big(0, T_{(j)}, X^{(j)}_{T_{(j)}}, \bar{u}^p_{(j)}, D\bar{u}^p_{(j)}, D^2\bar{u}^p_{(j)}\Big), \\
\bar{u}^p_k = \dfrac{1}{N_i} \displaystyle\sum_{\tilde{k} \in \tilde{Q}(k)} \frac{1}{2}\Big(\phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}, D\bar{u}^p_{\tilde{k}}, D^2\bar{u}^p_{\tilde{k}}\Big) + \\
\qquad \phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}1}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}1} D\bar{u}^p_{\tilde{k}1}, D^2\bar{u}^p_{\tilde{k}1}\Big)\Big), \quad \text{for } k = (k_1, \ldots, k_i) \in Q^o_i, 0 < i < p, \\
D\bar{u}^p_k = \dfrac{1}{N_i} \displaystyle\sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{V}^{\tilde{k}} \frac{1}{2}\Big(\phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}, D\bar{u}^p_{\tilde{k}}, D^2\bar{u}^p_{\tilde{k}}\Big) - \\
\qquad \phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}1}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}1} D\bar{u}^p_{\tilde{k}1}, D^2\bar{u}^p_{\tilde{k}1}\Big)\Big), \quad \text{for } k = (k_1, \ldots, k_i) \in Q^o_i, 0 < i < p, \\
D^2\bar{u}^p_k = \dfrac{1}{N_i} \displaystyle\sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{W}^{\tilde{k}} \frac{1}{2}\Big(\phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}, D\bar{u}^p_{\tilde{k}}, D^2\bar{u}^p_{\tilde{k}}\Big) + \\
\qquad \phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}1}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}1}, D\bar{u}^p_{\tilde{k}1}, D^2\bar{u}^p_{\tilde{k}1}\Big) - \\
\qquad 2\phi\Big(T_k, T_{\tilde{k}}, X^{\tilde{k}2}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}2}, D\bar{u}^p_{\tilde{k}2}, D^2\bar{u}^p_{\tilde{k}2}\Big)\Big), \quad \text{for } k = (k_1, \ldots, k_i) \in Q^o_i, 0 < i < p, \\
\bar{u}^p_k = g(X^k_{T_k}), D\bar{u}^p_k = Dg(X^k_{T_k}), D^2\bar{u}^p_k = D^2 g(X^k_{T_k}), \quad \text{for } k \in Q^o_p,
\end{cases}
\tag{28}
$$

# $f(\theta) = A : \theta$

## Assumption

*Equation* (21) *has a solution* $u$ *such that*

- $u \in C^{1,2p}([0,T] \times \mathbb{R}^d)$ *with uniformly bounded derivatives in* $x$ *and* $t$.
- $D^{2i}u$ *is* $\theta$*-Hölder with* $\theta \in (0,1]$ *in time with constant* $\hat{K}$ *for* $i = 1$ *to* $p$*:*

$$|D^{2p}u(t,.) - D^{2p}u(\tilde{t},.)|_\infty \leq \hat{K}|t - \tilde{t}|^\theta \qquad \forall (t,\tilde{t}) \in [0,T] \times [0,T]. \tag{29}$$

# Convergence

## Proposition

*There exists some functions of u:* $C_1(u)$, $C_2(u)$, $C_3(u)$, *and two functions* $\hat{C}(T)$ *and* $C(\sigma)$ *such that we have the following error given by the estimator* (28)*:*

$$\mathbb{E}((\bar{u}_\emptyset^p - u(0,x))^2) \leq C_1(u)\hat{C}(T)^{2p}C(\sigma)^{p-1}||A||_2^{2p}T^{2\theta}\frac{\gamma(\alpha,\lambda Tp)}{\Gamma(\alpha)}+$$

$$\sum_{i=0}^{p-1}\frac{C_2(u)}{N_i}\hat{C}(T)^{2i+2}C(\sigma)^i||A||_2^{2i+2}\frac{\gamma(\alpha,\lambda T(i+1))}{\Gamma(\alpha)}+$$

$$\sum_{i=0}^{p-1}\frac{C_3(u)}{N_i}\frac{\hat{C}(T)^{2i}||A||_2^{2i}C(\sigma)^i}{\bar{F}(T)^2}\frac{\gamma(\alpha,\lambda Ti)}{\Gamma(\alpha)}$$

# Semi Linear degenerated, $\sigma$ not invertible

$$(-\partial_t u - \mathcal{L}u)(t,x) = f(t,x,u(t,x),Du(t,x)),$$

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

with $\hat{\sigma}$ invertible.

$$\begin{aligned}
(-\partial_t u - \hat{\mathcal{L}}u)(t,x) =& \tilde{f}(t,x,u(t,x),Du(t,x),D^2u(t,x)) \\
\tilde{f}(t,x,u(t,x),Du(t,x),D^2u(t,x)) :=& f(t,x,u(t,x),Du(t,x)) - \\
& \frac{1}{2}(\hat{\sigma}\hat{\sigma}^\top - \sigma\sigma^\top) : D^2u(t,x)
\end{aligned}$$

# Semi linear degenerated case

- $$\mathcal{L}u(t,x) := k(m-x)Du(t,x) + \frac{1}{2}\bar{\sigma}(x)^2 : D^2u(t,x),$$
  and $k = \hat{k}I_d$, $\hat{k} \in \mathbb{R}^+$, $m = \hat{m}\mathbf{1}_d$, $\hat{m} \in \mathbb{R}^+$, $\bar{\sigma}(x)$ diagonal $\bar{\sigma}_{i,i}(x) = \hat{\sigma}\sqrt{x_i}$, $\hat{\sigma} \in \mathbb{R}^+$.
- Generator associated to multidimensional CIR process :
  $$dS_t^i = \hat{k}(\hat{m} - S^i)dt + \hat{\sigma}\sqrt{S_t^i}dW_t^i$$
- Non linearity and solution:
  $$f(x,y,z) = ay\sum_{i=1}^{d} z_i + c(t,x)$$
  $$u(t,x) = \cos(\sum_{i=1}^{d} x_i)e^{-\alpha(T-t)}.$$

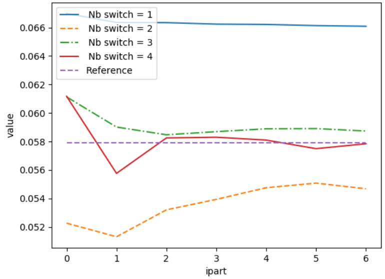
# PDE rewritten as

$$
\begin{aligned}
(-\partial_t u - \tilde{\mathcal{L}}u)(t,x) =& \tilde{f}(x, u(t,x), Du(t,x), D^2u(t,x)), \\
\tilde{f}(x, u(t,x), Du(t,x), D^2u(t,x)) =& \frac{1}{2}(\bar{\sigma}(x)^2 - \tilde{\sigma}^2)D^2u(t,x) + f(x, u(t,x), Du(t,x))), \\
\tilde{\mathcal{L}}u(t,x) :=& k(m-x)Du(t,x) + \frac{1}{2}\tilde{\sigma}^2 : D^2u(t,x), \\
\tilde{\sigma} =& \bar{\sigma} I_d, \quad \bar{\sigma} \in \mathbb{R}^+
\end{aligned}
$$

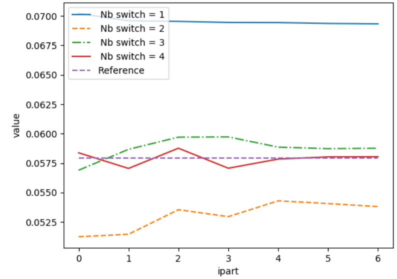
Using Ornstein Uhlenbeck forward process:

$$dS_t^i = \hat{k}(\hat{m} - S^i)dt + \bar{\sigma} dW_t^i.$$

# Results : $N_i^{ipart} = N_i^0 \times 2^{ipart}$

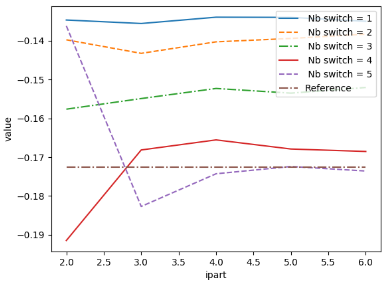



$\lambda = 0.1.$

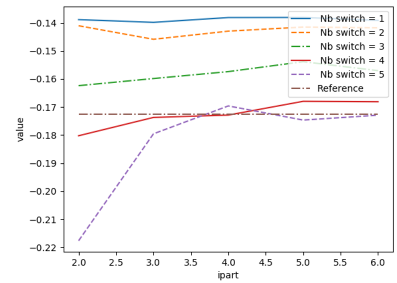



$\lambda = 0.15.$

CIR case dimension 5, $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 50, 25, 12)$




$\lambda = 0.05.$




$\lambda = 0.075.$

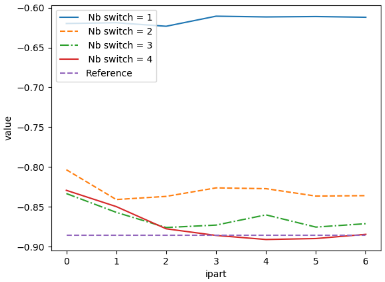
CIR case dimension 15, $(N_0^0, N_1^0, N_2^0, N_3^0, N_4^0) = (1000, 40, 20, 10, 5)$

# Full non linear case (no theoretical result)

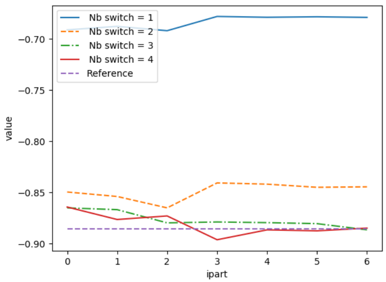
$$\mu = \frac{\mu_0}{d} \mathbb{1}_d,$$
$$\sigma = \frac{\sigma_0}{\sqrt{d}} \mathbf{I}_d,$$
$$f(t,x,y,z,\theta) = e(t,x) + \frac{a}{\sqrt{d}} (-e^{2\alpha(T-t)}) \vee (e^{2\alpha(T-t)} \wedge (y \sum_{i=1}^{d} \theta_{i,i})),$$

$$u(t,x) = e^{\alpha(T-t)} \cos(\sum_{i=1}^{d} x_i).$$

Take: $\mu_0 = 0.2$, $\sigma_0 = 1$, $\alpha = 0.1$, $x_0 = 0.5\mathbb{1}_d$, $T = 1$.

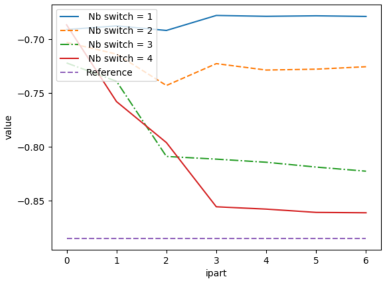



$\lambda = 0.1.$

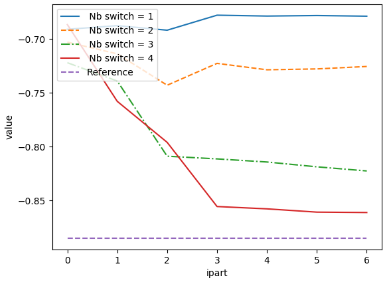



$\lambda = 0.2.$

Full non linear toy example $a = 0.1$, $d = 5$.




$\lambda = 0.1.$




$\lambda = 0.2.$

Full non linear toy example $a = 0.4$, $d = 5$.

Bruno Bouchard, Xiaolu Tan, and Xavier Warin.
Numerical approximation of general lipschitz bsdes with branching processes.
*arXiv preprint arXiv:1710.10933*, 2017.

Bruno Bouchard, Xiaolu Tan, Xavier Warin, and Yiyi Zou.
Numerical approximation of bsdes using local polynomial drivers and branching processes.
*Monte Carlo Methods and Applications*, 23(4):241–263, 2017.

Jean-François Chassagneux.
Linear multistep schemes for bsdes.
*SIAM Journal on Numerical Analysis*, 52(6):2815–2836, 2014.

Jean-François Chassagneux, Adrien Richou, et al.
Numerical simulation of quadratic bsdes.
*The Annals of Applied Probability*, 26(1):262–304, 2016.

Weinan E, Jiequn Han, and Arnulf Jentzen.

Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations.
*Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

Weinan E, Jiequn Han, and Arnulf Jentzen.
Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning.
*arXiv preprint arXiv:1707.02568*, 2017.

Weinan E, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse.
On multilevel picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations.
*arXiv preprint arXiv:1607.03295*, 46, 2016.

Weinan E, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse.

Linear scaling algorithms for solving high-dimensional nonlinear parabolic differential equations.
*SAM Research Report*, 2017, 2017.

Emmanuel Gobet, Jean-Philippe Lemor, Xavier Warin, et al.
A regression-based monte carlo method to solve backward stochastic differential equations.
*The Annals of Applied Probability*, 15(3):2172–2202, 2005.

Pierre Henry-Labordere.
Deep primal-dual algorithm for bsdes: Applications of machine learning to cva and im.
2017.

Pierre Henry-Labordere, Nadia Oudjane, Xiaolu Tan, Nizar Touzi, and Xavier Warin.
Branching diffusion representation of semilinear pdes and monte carlo approximation.
*arXiv preprint arXiv:1603.01727*, 2016.

Martin Hutzenthaler and Thomas Kruse.

Multi-level picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities.
*arXiv preprint arXiv:1711.01080*, 2017.

Jean-Philippe Lemor, Emmanuel Gobet, Xavier Warin, et al.
Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations.
*Bernoulli*, 12(5):889–916, 2006.