Probabilistic numerical methods for stochastic control problem: FBSDE approach

Jean-François Chassagneux

Université Paris Cité & LPSM

Woudschoten conference,
27-29 September 2023, Zeist (The Netherlands).

Outline

Framework & notations

- Very classical setting: $(\Omega, \mathcal{A}, \mathbb{P})$ a complete probability space with a Brownian Motion W, $(\mathcal{F}_t)_{t\geqslant 0}$ is the natural filtration of the BM. We are given a time horizon $T > 0$.
  $\hookrightarrow$ All the randomness comes from W.
- $\mathbb{R}^d$-valued martingales with prescribed terminal value $\xi \in \mathcal{L}^2(\mathcal{F}_T)$

$$\mathcal{Y}_t = \mathbb{E}[\xi|\mathcal{F}_t] = \xi - \int_t^T \mathcal{Z}_s \,\mathrm{d}W_s \;,\; 0 \leqslant t \leqslant T \;,$$

  $\hookrightarrow (\mathcal{Y}, \mathcal{Z}) \in \mathscr{S}_2 \times \mathscr{H}_2$ i.e. $\mathcal{Y}$ is continous and adapted process, $\mathcal{Z}$ is progressively measurable & $\mathbb{E}\left[\sup_{t\in[0,T]} |\mathcal{Y}_t|^2 + \int_0^T |\mathcal{Z}_t|^2 \,\mathrm{d}t\right] < \infty$.
- Example in a Markovian setting: Let $u(\cdot)$ be the solution of

  $\partial_t u + \frac{1}{2}\partial^2_{xx} u = 0$ and $u(T,\cdot) = g(\cdot)$ (backward heat equation)

  Set $\xi = g(W_T)$, $\mathcal{Y}_t := u(t, W_t)$ and $\mathcal{Z}_t := \partial_x u(t, W_t)$, compute

  $\xi = u(T, W_T) = u(t, W_t) + \int_t^T (\partial_t u + \frac{1}{2}\partial^2_{xx} u)(s, W_s)\,\mathrm{d}s + \int_t^T \partial_x u(s, W_s)\,\mathrm{d}W_s$

  $\hookrightarrow \xi = \mathcal{Y}_t + \int_t^T \mathcal{Z}_s \,\mathrm{d}W_s$ ...

Backward Stochastic Differential Equation

- Backward SDEs: Non linear perturbation with a *driver* f. Solve for $(\mathcal{Y}, \mathcal{Z}) \in \mathscr{S}_2 \times \mathscr{H}_2$ s.t.

$$\mathcal{Y}_t = \xi + \int_t^T f(s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s\ ,\ 0 \leqslant t \leqslant T\ .$$

- If $(y,z) \mapsto f(s,y,z)$ is a Lispchitz function (possibly random) then there exists a unique solution to the above equation ! [PP90, PP92]
- One typically shows that the following Picard iteration scheme

$$\mathcal{Y}^n_t = \xi + \int_t^T f(s, \mathcal{Y}^{n-1}_s, \mathcal{Z}^{n-1}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}^n_s\,\mathrm{d}W_s$$

converges in $\mathscr{S}_2 \times \mathscr{H}_2$ to a unique fixed point...
- In dimension one (for $\mathcal{Y}$), they satisfy a *comparison theorem* namely consider, for $i \in \{1,2\}$, ξ^i terminal condition of BSDE solution $(\mathcal{Y}^i, \mathcal{Z}^i)$,
$\hookrightarrow$ If $\xi_1 \geqslant \xi_2$, then $\mathcal{Y}^1_t \geqslant \mathcal{Y}^2_t$, $t \in [0,T]$.

Markovian BSDEs - Link with PDEs

- Consider a forward SDE:

$$\mathcal{X}_t = x + \int_0^t b(\mathcal{X}_s)\,\mathrm{d}s + \int_0^t \sigma(\mathcal{X}_s)\,\mathrm{d}W_s$$

  for some (b, σ) Lipschitz continuous.
- Markovian BSDE: One sets $\xi := g(\mathcal{X}_T)$ and solve for

$$\mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s$$

  $(x, y, z) \mapsto f(x, y, z)$ is deterministic
- Then one has $\mathcal{Y}_t = u(t, \mathcal{X}_t)$ for some function u (and with a bit of smoothness $\mathcal{Z}_t = \sigma(\mathcal{X}_t)\partial_x u(t, \mathcal{X}_t)$)
- The function u satisfies

$$\partial_t u + \mathcal{L}_X u + f(x, u, \sigma(.)\partial_x u) = 0 \text{ and } u(T, \cdot) = g(\cdot), \qquad (1)$$

  where $\mathcal{L}_X u = b(x)\partial_x u + \frac{1}{2}\sigma^2(x)\partial^2_{xx} u$ (in dimension one.)

Elements of proof

- The representation $\mathcal{Y}_t = u(t, \mathcal{X}_t)$ comes from the Markov property of $(\mathcal{X}_t)_{0 \leqslant t \leqslant T}$ which is extended in this nonlinear setting.
- If one assumes smoothness of u then applying Ito's formula, we get

$$\mathrm{d}u(t, \mathcal{X}_t) = (\partial_t u + \mathcal{L}_X u)(t, \mathcal{X}_t)\,\mathrm{d}t + \sigma(\mathcal{X}_t)\partial_x u(t, \mathcal{X}_t)\,\mathrm{d}W_t$$

and we also have

$$\mathrm{d}u(t, \mathcal{X}_t) = \mathrm{d}\mathcal{Y}_t = -f(\mathcal{X}_t, \mathcal{Y}_t, \mathcal{Z}_t)\,\mathrm{d}t + \mathcal{Z}_t\,\mathrm{d}W_t$$

- Identifying the Brownian part leads to $\mathcal{Z}_t = \sigma(\mathcal{X}_t)\partial_x u(t, \mathcal{X}_t)$
- Then, identifying the '$\mathrm{d}t$'-part shows that u solves the semi-linear PDE, as claimed.
- Under various assumptions, one can show that u is smooth and solves the PDE classicaly. If not, viscosity solution theory can be used.
- The link between the FBSDE and the semi-linear PDE is often coined as "non-linear Feynman-Kac formula".

A Stochastic Target Problem

- Consider a state process X and a one dimensional controlled process:

$$Y_t^{y,Z} = y - \int_0^t f(\mathcal{X}_s, Y_s, Z_s)\,\mathrm{d}s + \int_t^T Z_s\,\mathrm{d}W_s \qquad (2)$$

  The controls are $y \in \mathbb{R}$ (initial value) and $Z \in \mathscr{H}^2$.

- Target problem: the target is $\xi \in \mathcal{L}^2(F_T)$, goal: find

$$p := \inf\{y \in \mathbb{R} \,|\, \exists Z \in \mathscr{H}^2,\ Y_T^{y,Z} \geqslant \xi\} \qquad (3)$$

- The solution is the initial value of the BSDE: $p := \mathcal{Y}_0$ where

$$\mathcal{Y}_t = \xi + \int_t^T f(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s$$

  The optimal control is then given by $\mathcal{Z}$.

- A typical example of stochastic target problem is the super-hedging problem in finance. Note that this link between BSDEs and stochastic target problem will be used later on in the numerical part.

Elements of proof

- The fact that $(\mathcal{Y}_0, \mathcal{Z})$ is a super-solution (namely $\mathcal{Y}_0 \geqslant p$) to the problem comes existence result for BSDEs. The more delicate question is the fact that $\mathcal{Y}_0$ gives exactly the infinimum.
- To answer this question, one uses the comparison theorem for one dimensional BSDEs, recall:
  Let $\xi_1 \geqslant \xi_2$. For $i = 1, 2$, denote $(\mathcal{Y}^i, \mathcal{Z}^i)$ the BSDEs associated to the terminal condition ξ^i. Then, $\mathcal{Y}^1_t \geqslant \mathcal{Y}^2_t$, for all $t \in [0, T]$.
  $\hookrightarrow$ This comparison theorem is proven using a linearisation argument, see e.g. [EKPQ97].
- Conclusion: assume that there exists y and Z such that $Y^{y,Z}_T \geqslant \xi = \mathcal{Y}_T$, one applies simply the comparison theorem to get that $y \geqslant \mathcal{Y}_0$. Then taking the infinimum on y, we get $p = \mathcal{Y}_0$ (recall (3))

A classical stochastic control problem

We consider a control problem that leads to a toy model for prices in carbon markets, see last section. Here (W, B) are independent Brownian Motion.

- Let P be a multidimensional auxiliary process: $\mathrm{d}P_t = \sigma(P_t)\,\mathrm{d}W_t$
- Let (E_t^α) be the controlled emission process for $\alpha \in \mathscr{H}^2$ with dynamics

$$\mathrm{d}E_t^\alpha = (b(P_t) - \alpha_t)\,\mathrm{d}t + \eta\,\mathrm{d}B_t \quad (\eta \geqslant 0)$$

- Let the cost functional be given by

$$J(\alpha) = \mathbb{E}\left[\int_0^T \frac{1}{2}|\alpha_s|^2\,\mathrm{d}s + g(E_T^\alpha)\right]$$

  Here g is a C^1 convex function.
- One has to solve: $\min_{\alpha \in \mathcal{H}^2} J(\alpha)$ (a strictly convex minimisation problem which has thus a solution)
- The story is that g is a terminal cost that has to pay the company if it emits above a given level Λ, typically $g(e) = \lambda(\Lambda - e)_+$ (actually a smoother version), α is the abatement process to reduce emission and $\frac{1}{2}|\alpha|^2$ is the cost of abatement.

Stochastic Maximum Principle

In order to study the solvability of the problem, we compute the Gâteaux derivative of the functional $\alpha \mapsto J(\alpha)$.

- We first observe that

$$\partial_v J(\alpha) := \lim_{\epsilon \to 0} \frac{1}{\epsilon}(J(\alpha + \epsilon v) - J(\alpha)) = \mathbb{E}\left[\int_0^T \alpha_s v_s ds + g'(E_T^\alpha)\partial_v E_T^\alpha\right]$$

  where $\partial_v E_t^\alpha = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(E_t^{\alpha+\epsilon v} - E_t^\alpha) = -\int_0^t v_s \, \mathrm{d}s$.

- Define the "BSDE" $Y_t^\alpha = \mathbb{E}[g'(E_T^\alpha)|\mathcal{F}_t]$ (the adjoint process) to compute:

$$\partial_v J(\alpha) = \mathbb{E}\left[\int_0^T (\alpha_t - Y_t^\alpha) v_t \, \mathrm{d}t\right]$$

  at the optimum $\alpha^\star$, we have $\partial_v J(\alpha^\star) = 0$, $\forall v$ necessarily and thus $Y^\star = \alpha^\star$

- Then $(E^\star, Y^\star)$ solves a fully coupled BSDEs

  $\mathrm{d}E_t^\star = (b(P_t) - Y_t^\star)\,\mathrm{d}t + \eta\,\mathrm{d}B_t$ and $Y_t^\star = \mathbb{E}[g'(E_T^\star)|\mathcal{F}_t]$. ($f = 0$ here!)

Forward-Backward stochastic differential equation

- The solution (the optimal control process) to the previous optimisation problem is an example of *fully coupled* Forward-Backward SDE.
- Namely, a process $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ that solves on $[0, T]$

$$\mathcal{X}_t = x + \int_0^t b(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s + \int_0^t \sigma(\mathcal{X}_s, \mathcal{Y}_s)\,\mathrm{d}W_s \quad \text{(forward)}$$

$$\mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s \quad \text{(backward)}$$

- Wellposedness is more difficult to obtain. Lispchitz assumption of the coefficient is not enough. One needs on top
  - whether structural conditions (smallness condition, monotonicity),
  - or boundedness of terminal condition and ellipticity of σ.
- A key step is to prove that $\mathcal{Y}_t = u(t, \mathcal{X}_t)$ for some Lipschitz function $u(\cdot)$ called the *decoupling field*. With a bit of smoothness, one has that $u(\cdot)$ solves (denoting $v = \sigma(x, u)\partial_x u$) the parabolic quasilinear PDE

$$\partial_t u + b(x, u, v)\partial_x u + \frac{1}{2}\sigma^2(x, u)\partial^2_{xx} u + f(x, u, v) = 0 \text{ and } u(T, \cdot) = g(\cdot)$$

Further representation properties

FBSDE can be extended to obtain representation of the solution of more general control problem [ZZ17]

- When the volatility of the state process is not controlled, it is possible to represent directly the optimal *value* of the control problem by using quadratic BSDEs. The associated PDE are HJB equation.
- If the control problem is (related to) an optimal stopping problem then the definition of Reflected BSDE allow to represent the value function. The associated PDE are quasivariational inequalities.
- If the volatility of the state process is also controlled, then one introduces Second Order BSDE to obtain probabilistic representation of the value function. The associated PDE are HJB equations (fully non linear PDEs)
- The solution of large population stochastic control problem as Mean Field Games or Mean Field Control can be represented using McKean-Vlasov FBSDEs (FBSDEs depending on the law of the solution). The related PDE are then quasilinear PDE or HJB equation written on the space of probability measure.

Outline

Setting

- We want to approximate the solution $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})_{t\in[0,T]}$ to the system:

$$\mathcal{X}_t = x_0 + \int_0^t b(\mathcal{X}_s)\,\mathrm{d}s + \int_0^t \sigma(\mathcal{X}_s)\,\mathrm{d}W_s \tag{4}$$

$$\mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s \tag{5}$$

where b, σ, f and g are L-Lipschitz continous functions for some $L > 0$.

  - Let us be given a discrete-time grid $\pi = \{t_0 := 0 < \cdots < t_N = T\}$, with $t_{n+1} - t_n = T/N =: h$
  - We consider that we know how to similate the Brownian motion (W) on this grid and denote $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, h)$
  - The forward SDE (36) is approximated by a Euler-Maruyama scheme:

$$X_{n+1} = X_n + b(X_n)h + \sigma(X_n)\Delta W_n \text{ and } X_0 = x_0. \tag{6}$$

$\hookrightarrow$ Well studied and "easy" to simulate see e.g. [KPS12].

- Our main question is how to approximate (37)?

Heuristics for the scheme

- Write down the solution $(\mathcal{Y}, \mathcal{Z})$ on π:

$$\mathcal{Y}_{t_n} = \mathcal{Y}_{t_{n+1}} + \int_{t_n}^{t_{n+1}} f(\mathcal{Y}_t, \mathcal{Z}_t)\,\mathrm{d}t - \int_{t_n}^{t_{n+1}} \mathcal{Z}_t\,\mathrm{d}W_t$$

- Approximate the integrals to get:

$$\mathcal{Y}_{t_n} \simeq \mathcal{Y}_{t_{n+1}} + hf(\mathcal{Y}_{t_n}, \mathcal{Z}_{t_n}) - \mathcal{Z}_{t_n}\Delta W_n \qquad (7)$$

- Taking conditional expectation on both sides above, we obtain:

$$\mathcal{Y}_{t_n} \simeq \mathbb{E}_{t_n}\big[\mathcal{Y}_{t_{n+1}}\big] + hf(\mathcal{Y}_{t_n}, \mathcal{Z}_{t_n}) \to \text{scheme for Y}$$

$\hookrightarrow$ one needs an approximation for $\mathcal{Z}$!

- Multiply (7) by ΔW_n and take conditional expectation on both sides:

$$0 \simeq \mathbb{E}_{t_n}\big[\mathcal{Y}_{t_{n+1}}\Delta W_n\big] - h\mathcal{Z}_{t_n} \to \text{scheme for Z}$$

Scheme Definition for BSDE

- Euler scheme for BSDEs [BT04, Zha04]:
  - at $t_N = T$: set

$$(Y_N, Z_N) = (g(X_N), 0) \tag{8}$$

  - for $n < N$, compute

$$\begin{cases} Z_n = \mathbb{E}_{t_n}\big[Y_{n+1}\frac{\Delta W_n}{h}\big] \\ Y_n = \mathbb{E}_{t_n}[Y_{n+1}] + hf(Y_n, Z_n) \end{cases} \tag{9}$$

  ($\mathbb{E}_{t_n}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t_n}]$ conditional expectation operator)

- Some remarks:
  - This is an implicit scheme in y, so one must impose $hL < 1$ to obtain well-posedness. In practice, Picard iteration will allow to compute the fixed point rapidly.
  - One could also use explicit scheme: $Y_n = \mathbb{E}_{t_n}[Y_{n+1} + hf(Y_{n+1}, Z_n)]$
  - In practice, a difficult question is the estimation of the conditional expectation.

Functional definition of the scheme

- From the Markov property of the Euler scheme (X_n), one deduces that $Y_n = \bar{u}_n(X_n)$ and $Z_n = \bar{v}_n(X_n)$, for some measurable functions $(\bar{u}_n, \bar{v}_n)$.
- More precisely, define one step of Euler scheme for $\mathcal{X}$:

$$X_{n+1}^{t_n,x} := x + b(x)h + \sigma(x)\Delta W_n$$

and, setting $\bar{u}_N = g$, compute from $n+1$ to n:

$$\bar{v}_n(x) = \mathbb{E}\left[\bar{u}_{n+1}(X_{n+1}^{t_n,x})\frac{\Delta W_n}{h}\right]$$

$$\bar{u}_n(x) = \mathbb{E}\left[\bar{u}_{n+1}(X_{n+1}^{t_n,x})\right] + hf(\bar{u}_n(x), \bar{v}_n(x))$$

- If one introduces a grid for x-values, then one can obtain approximation of $\bar{u}_n$, $\bar{v}_n$ at this grid point. This has to be combined with some interpolation procedure to evaluate e.g. $\bar{u}_{n+1}(X_n^{t_{n-1},x})$...
- Later on we discuss instead regression method that also yields approximation of $(\bar{u}_n, \bar{v}_n)$.

L^2-stability of the Euler scheme

- Perturbation approach: Consider, for $\zeta_n \in L^2(\mathcal{F}_{t_n})$,

$$\tilde{Y}_n = \mathbb{E}_{t_n}\left[\tilde{Y}_{n+1}\right] + hf(\tilde{Y}_n, \tilde{Z}_n) + \zeta_n \quad (10)$$

$$\tilde{Z}_n = \mathbb{E}_{t_n}\left[\tilde{Y}_{n+1}\frac{\Delta W_n}{h}\right] \quad (11)$$

Definition (L^2-stability)

The scheme given in (38) is L^2-stable if there exists a constant $C > 0$ s.t.

$$\max_n \mathbb{E}\left[|Y_n - \tilde{Y}_n|^2\right] + \sum_{n=0}^{N-1} h\mathbb{E}\left[|Z_n - \tilde{Z}_n|^2\right] \leqslant C\mathbb{E}\left[|Y_N - \tilde{Y}_N|^2 + N\sum_{n=0}^{N-1}\zeta_n^2\right]$$

for h small enough, for all perturbation ζ.

Theorem

If f is Lipschitz continous, the scheme (38) is L^2-stable.

Proof of the stability result

In the proof below, C is a constant that will change from line to line but that does not depend on the discretisation grid.
A key point is to observe that the scheme can be rewritten "almost as a BSDE". Indeed, we observe that

$$Y_n = Y_{n+1} + hf(Y_n, Z_n) - hZ_nH_n - \Delta M_n \tag{12}$$

where $H_n = \frac{\Delta W_n}{h}$. Note that (12) defines ΔM_n, moreover it satisfies

$$\mathbb{E}_{t_n}[\Delta M_n] = \mathbb{E}_{t_n}[\Delta M_n H_n] = 0 \text{ and } \mathbb{E}\left[|\Delta M_n|^2\right] < \infty \tag{13}$$

These properties are obtained by using the definition of the scheme given in (38).
For the perturbed scheme, we have similarly:

$$\tilde{Y}_n = \tilde{Y}_{n+1} + hf(\tilde{Y}_n, \tilde{Z}_n) + \zeta_n - h\tilde{Z}_nH_n - \Delta\tilde{M}_n \tag{14}$$

Denoting $\delta f_n = f(Y_n, Z_n) - f(\tilde{Y}_n, \tilde{Z}_n)$ and $\delta\Delta M_n = \Delta M_n - \Delta\tilde{M}_n$, we observe that

$$\delta Y_n + h\delta Z_nH_n + \delta\Delta M_n = \delta Y_{n+1} + h\delta f_n + \zeta_n\,. \tag{15}$$

Squaring both sides and taking conditional expectation, we compute, using Young's inequality,

$$|\delta Y_n|^2 + h|\delta Z_n|^2 \leqslant (1 + Ch)\mathbb{E}_{t_n}\left[(\delta Y_{n+1} + h\delta f_n)^2\right] + \frac{C}{h}|\zeta_n|^2\,. \tag{16}$$

Note that

$$(\delta Y_{n+1} + h\delta f_n)^2 \leqslant (|\delta Y_{n+1}| + Ch|\delta Y_n| + Ch|\delta Z_n|)^2 \tag{17}$$

$$\leqslant (1 + \frac{h}{\epsilon})(|\delta Y_{n+1}| + Ch|\delta Y_n|)^2 + C(1 + \frac{\epsilon}{h})h^2|\delta Z_n|^2 \tag{18}$$

Choosing h and ϵ such that $C(h + \epsilon) \leqslant \frac{1}{2}$, we obtain

$$(\delta Y_{n+1} + h\delta f_n)^2 \leqslant (1 + Ch)|\delta Y_{n+1}|^2 + Ch|\delta Y_n|^2 + \frac{1}{2}h^2|\delta Z_n|^2 \tag{19}$$

Inserting the previous inequality in (16), we get

$$|\delta Y_n|^2 + \frac{1}{2}h|\delta Z_n|^2 \leqslant (1 + Ch)\mathbb{E}_{t_n}\big[|\delta Y_{n+1}|^2\big] + \frac{C}{h}|\zeta_n|^2 \tag{20}$$

$$|\delta Y_n|^2 \leqslant e^{Ch}\mathbb{E}_{t_n}\big[|\delta Y_{n+1}|^2\big] + \frac{C}{h}|\zeta_n|^2 \tag{21}$$

Taking expectation on both sides and iterating over n, we obtain the stability result for the Y part. For the Z part we sum over n (20) and used the stability obtained for the Y part to conclude. □

Truncation error

- Let us introduce

$$\hat{Z}_{t_n} = \mathbb{E}_{t_n}\left[\mathcal{Y}_{t_{n+1}} \frac{\Delta W_n}{h}\right]. \tag{22}$$

- We define the *local truncation error* as

$$\hat{\zeta}_n := \mathbb{E}_{t_n}\left[\int_{t_n}^{t_{n+1}} [f(\mathcal{Y}_t, \mathcal{Z}_t) - f(\mathcal{Y}_{t_n}, \hat{Z}_{t_n})]\,\mathrm{d}t\right] . \tag{23}$$

- It measures how well the true solution satisfies the scheme.
- The *global truncation error* is then defined as

$$\mathcal{T}(\pi) := \sum_{n=0}^{N-1} \mathbb{E}\left[N|\hat{\zeta}_n|^2\right] . \tag{24}$$

- Assume at this point that there is **no error made on the forward process**, thus $Y_N - Y_T = 0$ and we have

$$\max_n \mathbb{E}\left[|\mathcal{Y}_{t_n} - Y_n|^2\right] \leqslant C\mathcal{T}(\pi) . \tag{25}$$

Order of convergence

Theorem

In the regular case, we have that

$$\mathcal{T}(\pi) \leqslant C\mathbf{h^2}$$

and thus the scheme is of order 1*.*

Theorem

In the Lispchitz coefficient case, we have that

$$\mathcal{T}(\pi) \leqslant C\left(h + \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \mathbb{E}\big[|\mathcal{Y}_t - \mathcal{Y}_{t_i}|^2\,\mathrm{d}t\big] + \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \mathbb{E}\big[|\mathcal{Z}_t - \bar{\mathcal{Z}}_{t_i}|^2\big]\,\mathrm{d}t\right) \leqslant C\mathbf{h}$$

where $\bar{\mathcal{Z}}_{t_i} = \frac{1}{h_i}\mathbb{E}_{t_i}\Big[\int_{t_i}^{t_{i+1}} Z_t\,\mathrm{d}t\Big]$. *The scheme is of order one half.*

Proof in the regular case

(proof in the one dimensional case to simplify notations) We denote for a smooth function ϕ

$$\phi^{(0)}(t,x) := (\partial_t \phi + \mathcal{L}_X \phi)(t,x), \phi^{(1)} = \sigma()\partial_x \phi$$

We observe that, since $\hat{\zeta}_n = \mathbb{E}_{t_n}\Big[\int_{t_n}^{t_{n+1}} [f(\mathcal{Y}_t, \mathcal{Z}_t) - f(\mathcal{Y}_{t_n}, \hat{Z}_{t_n})]\,\mathrm{d}t\Big]$

$$|\hat{\zeta}_n| \leqslant |\mathbb{E}_{t_n}\Big[\int_{t_n}^{t_{n+1}} [f(\mathcal{Y}_t, \mathcal{Z}_t) - f(\mathcal{Y}_{t_n}, \mathcal{Z}_{t_n})]\,\mathrm{d}t\Big]| + Ch|\mathcal{Z}_{t_n} - \hat{Z}_{t_n}|. \quad (26)$$

Using the PDE satisfied by u i.e. $u^{(0)} + f(u, u^{(1)}) = 0$, we get

$$|\mathbb{E}_{t_n}[f(\mathcal{Y}_t, \mathcal{Z}_t) - f(\mathcal{Y}_{t_n}, \mathcal{Z}_{t_n})]| = |\mathbb{E}_{t_n}\Big[u^{(0)}(t, \mathcal{X}_t) - u^{(0)}(t_n, \mathcal{X}_{t_n})\Big]| \quad (27)$$

$$= |\mathbb{E}_{t_n}\Big[\int_{t_n}^{t_{n+1}} u^{(0,0)}(t, \mathcal{X}_t)\,\mathrm{d}t\Big]| \quad (28)$$

$$\leqslant Ch \quad (29)$$

where we used Ito's formula for the last equality.

2. Now we compute, setting $H_n := \frac{\Delta W_n}{h_n}$

$$\hat{Z}_{t_n} = \mathbb{E}_{t_n}[u(t_{n+1}, \mathcal{X}_{t_{n+1}}) H_n] = \mathbb{E}_{t_n}\left[H_n \int_{t_n}^{t_{n+1}} u^{(0)}(t, \mathcal{X}_t)\,\mathrm{d}t + \frac{1}{h}\int_{t_n}^{t_{n+1}} u^{(1)}(t, \mathcal{X}_t)\,\mathrm{d}t\right] \quad (30)$$

Observe that

$$|\mathbb{E}_{t_n}\left[H_n \int_{t_n}^{t_{n+1}} u^{(0)}(t, \mathcal{X}_t)\,\mathrm{d}t\right]| = |\mathbb{E}_{t_n}\left[H_n \int_{t_n}^{t_{n+1}} \{u^{(0)}(t, \mathcal{X}_t) - u^{(0)}(t_n, \mathcal{X}_{t_n})\}\,\mathrm{d}t\right]| \quad (31)$$

$$\leqslant Ch \quad (32)$$

and that

$$\mathbb{E}_{t_n}\left[u^{(1)}(t, \mathcal{X}_t)\right] = \mathbb{E}_{t_n}\left[u^{(1)}(t_n, \mathcal{X}_{t_n}) + \int_{t_n}^{t} u^{(0,1)}(s, \mathcal{X}_s)\,\mathrm{d}s\right] \quad (33)$$

We thus get

$$|\mathbb{E}_{t_n}\left[\frac{1}{h}\int_{t_n}^{t_{n+1}} u^{(1)}(t, \mathcal{X}_t)\,\mathrm{d}t\right] - u^{(1)}(t_n, \mathcal{X}_{t_n})| \leqslant Ch \quad (34)$$

leading to $|\mathcal{Z}_{t_n} - \hat{Z}_{t_n}|^2 \leqslant Ch$. We then obtain that

$$|\hat{\zeta}_n| \leqslant Ch^2. \quad (35)$$

And, by summing this local error estimate, we conclude that $\mathcal{T}(\pi) \leqslant Ch^2$. □

Remarks

- When $\mathcal{X}$ cannot be perfectly simulated, the discrete time error due to its approximation has to be taken into account: the convergence results do not change.
- One key question is how to estimate the conditional expectation (see example below). This is also a new source of error, that propagates along the backward recursion. This point is also well understood for most of the methods.
- The theoretical analysis has also been conducted for coefficients that are not globally Lipschitz or with irregular terminal conditions. The scheme might need to be modified slightly and the convergence rate is lower in some cases.
- Higher order schemes: The discrete time error can be reduced bu using e.g. Crank-Nicolson scheme:

$$Y_n = \mathbb{E}\left[Y_{n+1} + \frac{h}{2}\left(f(Y_n, Z_n) + f(Y_{n+1}, Z_{n+1})\right)|\mathcal{F}_{t_n}\right]$$

  $\hookrightarrow$ (the scheme for Z has to be modified as well)
  RK scheme, linear multi-step scheme have also been introduced, see numerical example below.

Small recap

- We want to approximate the solution $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})_{t\in[0,T]}$ to the system:

$$\mathcal{X}_t = x_0 + \int_0^t b(\mathcal{X}_s)\,\mathrm{d}s + \int_0^t \sigma(\mathcal{X}_s)\,\mathrm{d}W_s \tag{36}$$

$$\mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s \tag{37}$$

where b, σ, f and g are L-Lipschitz continous functions for some $L > 0$.

- Euler scheme for BSDEs:
  - at $t_N = T$: set $(Y_N, Z_N) = (g(X_N), 0)$
  - for $n < N$, compute

$$\begin{cases} Z_n = \mathbb{E}_{t_n}\big[Y_{n+1}\frac{\Delta W_n}{h}\big] \\ Y_n = \mathbb{E}_{t_n}\big[Y_{n+1}\big] + hf(Y_n, Z_n) \end{cases} \tag{38}$$

- The scheme is wellposed, stable and convergent with a rate. We need to understand how to compute the conditional expectations involved.

Implementation via (non-linear) regression

- In practice, one needs to estimate conditional expectation. Many methods can be used, e.g. Malliavin method [CMT10, BT04, BET04], quantization method [BP03, PS18], cubature and tree based method [CM12, Cha14, CT17], Fourier method [RO15].
- Regression methods, as the one used for US options, can be easily adapted to the BSDEs setting, introduced in [GLW05, LGW06] and extensively studied in [GT16b, GT16a, GLSTV16, GT17]
- With regression methods, one obtains an approximation the functions $u(t, \cdot)$ for $t \in \pi$ (and also of it gradient through the Z) instead of a the value at a point (or around a point).
- Recently, [HPW20] used (deep) Neural Networks to compute the above scheme. Namely the class Φ is given by NN with a fixed structure with parameters $\theta \in \mathbb{R}^q$ for some q big.

$\hookrightarrow$ we explain below how a to set up the regression method.

How to compute one step of the scheme?

- To compute conditional expectation $\mathbb{E}[\mathcal{U}(X_{n+1})|\mathcal{F}_{t_n}]$ where $\mathcal{U}(\cdot)$ is a measurable function s.t. $\mathcal{U}(X_{n+1}) \in \mathcal{L}^2$, one observes

$$\mathbb{E}[\mathcal{U}(X_{n+1})|\mathcal{F}_{t_n}] = \mathbb{E}[\mathcal{U}(X_{n+1})|X_n] = \operatorname{argmin}_{\gamma \in \mathcal{L}^2} \mathbb{E}\Big[|\mathcal{U}(X_{n+1}) - \gamma(X_n)|^2\Big]$$

  $\hookrightarrow$ In practice one has to restric the class of possible function γ: linear specification (leading to OLS) or non-linear specification (Neural network to train)

- One could compute two regression:

$$Z_n = \mathbb{E}_{t_n}\left[Y_{n+1}\frac{\Delta W_n}{h}\right] \text{ and } Y_n = \mathbb{E}_{t_n}[Y_{n+1}] + hf(Y_n, Z_n)$$

- Or observe that

$$(Y_n, Z_n) = \operatorname{argmin}_{y,z \in L^2(\mathcal{F}_{t_n})} \mathbb{E}\Big[|Y_{n+1} - (y - hf(X_n, y, z) + z\Delta W_n)|^2\Big].$$

Elements of proof

- We first observe that

$$Y_{n+1} = Y_n - hf(X_n, Y_n, Z_n) + Z_n\Delta W_n - \Delta M_n,$$

where $\mathbb{E}_{t_n}[\Delta M_n] = \mathbb{E}_{t_n}[\Delta M_n \Delta W_n] = 0, \mathbb{E}_{t_n}\left[|\Delta M_n|^2\right] < \infty$. So that

$$\begin{aligned}
&\mathbb{E}\left[|Y_{n+1} - (y - hf(X_n, y, z) + z\Delta W_n)|^2\right] \\
&= \mathbb{E}\left[|Y_n - hf(X_n, Y_n, Z_n) - \{y - hf(X_n, y, z)\} + (Z_n - z)\Delta W_n - \Delta M_n|^2\right] \\
&= \mathbb{E}\left[|Y_n - hf(X_n, Y_n, Z_n) - \{y - hf(X_n, y, z)\}|^2\right] + h\mathbb{E}\left[|Z_n - z|^2\right] + \mathbb{E}\left[|\Delta M_n|^2\right].
\end{aligned}$$

Obviously, (Y_n, Z_n) does achieve the minimum of the right side of the above equation.
Reciprocally, any optimal solution $(y^\star, z^\star)$ must satisfy $z^\star = Z_n$ from the second term of the right side in the above equality. Moreover, necessarily one has

$$\begin{aligned}
y^\star &= \mathbb{E}_{t_n}[Y_n - hf(X_n, Y_n, Z_n) + hf(X_n, y^\star, Z_n)] \\
&= \mathbb{E}_{t_n}[Y_{n+1} + hf(X_n, y^\star, Z_n)] \ .
\end{aligned}$$

By uniqueness of the scheme definition, we conclude $y^\star = Y_n$. □

The scheme in practice

- [HPW20] uses the class of feedforward Neural Network: functions $\varphi(\theta, \cdot)$ with parameters $\theta \in \mathbb{R}^q$ for some q large (θ representing biais and matrice weights for all layers).
- The scheme is as follows:
  - set at T, $\mathcal{U}^\star_N(\cdot) = g(\cdot)$
  - from step $n+1$ to n, knowing $\mathcal{U}^\star_{n+1}(\cdot)$, find $\theta^\star, \vartheta^\star \in \mathrm{argmin}_{\theta,\vartheta}$ of

$$\mathbb{E}\big[|\mathcal{U}^\star_{n+1}(X_{n+1}) - (\mathcal{U}(\theta, X_n) - hf(\mathcal{U}(\theta, X_n), \mathcal{V}(\vartheta, X_n)) + \mathcal{V}(\vartheta, X_n)\Delta W_n)|^2\big]$$

  $\hookrightarrow$ Then set $\mathcal{U}^\star_n(\cdot) = \mathcal{U}(\theta^\star, \cdot)$
- $\mathcal{U}^\star_n$ is then an approximation of $u(t_n, \cdot)$.
- In practice, the above optimisation problem is computed using SGD methods (training of Neural Networks).

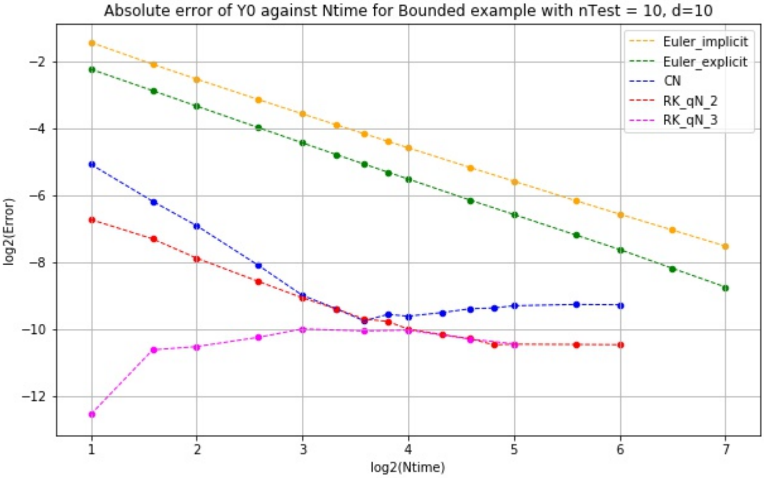
Numerical illustration [CCF22] - a toy model

- We consider the following example borrowed from [HPW20] (essentially brownian setting): For $d = 10$, $T = 1$, $t \in [0, T]$, let

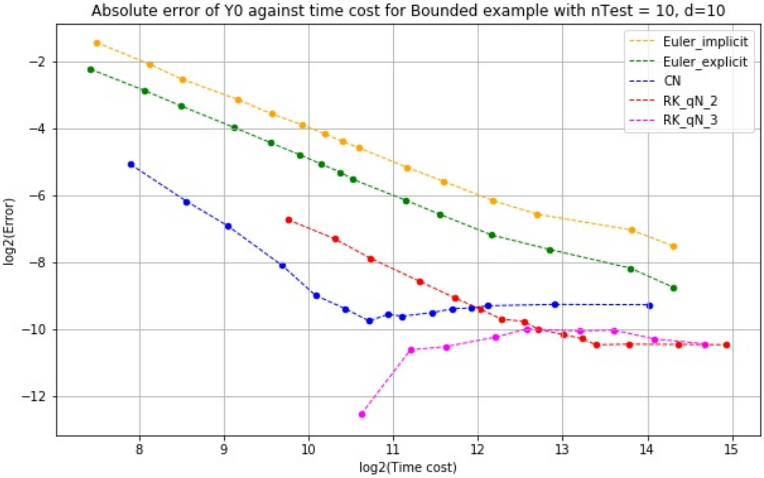
$$
\begin{aligned}
\mathrm{d}\mathcal{X}_t &= \frac{0.2}{d}\mathbf{1}_d\,\mathrm{d}t + \frac{1}{\sqrt{d}}\mathbf{n}_d\,\mathrm{d}W_t, \quad \mathcal{X}_0 = \mathbf{1}_d,\\
f(t,x,y,z) &= (\cos(\bar{x}) + 0.2\sin(\bar{x}))\,e^{\frac{T-t}{2}} - \frac{1}{2}(\sin(\bar{x})\cos(\bar{x})e^{T-t})^2\\
&\quad + \frac{1}{2d}\left(y\left(z\cdot\mathbf{1}_d\right)\right)^2,\\
g(x) &= \cos(\bar{x}), \qquad \text{where } \bar{x} = \sum_{i=1}^{d} x_n.
\end{aligned}
$$

- The theoretical solution of this BSDE is $Y_t = \cos(\bar{X}_t)e^{\frac{T-t}{2}}$ and $Z_t^i = -\frac{1}{\sqrt{d}}\sin(\bar{X}_t)e^{\frac{T-t}{2}}, i = 1, \cdots, d.$. (Apply Ito's formula...)

Numerical illustration [CCF22] - rate of convergence



Numerical illustration [CCF22] - complexity control



A shooting method: The Deep BSDE solver [HJW18]

- Consider, for $y \in \mathbb{R}$ and $Z \in \mathcal{H}_2$, the controlled process

$$Y_t^{y,Z} = y - \int_0^t f(Y_s^{y,Z}, Z_s)\,\mathrm{d}s + \int_0^t Z_s\,\mathrm{d}W_s \tag{39}$$

- Introduce the following optimisation problem

$$V := \min_{(y,Z)\in\mathbb{R}\times\mathcal{H}_2} \mathbb{E}\Big[|g(\mathcal{X}_T) - Y_T^{y,Z}|^2\Big] \tag{40}$$

- In the Lipschitz setting, the solution of the above optimisation is $V = 0$ and the argmin is given by $(\mathcal{Y}_0, \mathcal{Z})$ solution to the BSDE.
- Main idea: solve numerically the optimisation problem (40) to get an approximation of the BSDE.
- Remark: we observe that this numerical method aims to solve directly the stochastic target problem!

Discrete problem

- The controlled process $Y^{y,Z}$ is discretised using an Euler scheme on π: $\bar{Y}_{n+1} = \bar{Y}_n + hf(\bar{Y}_n, Z_n) + Z_n \Delta W_n$ and $\bar{Y}_0 = y$.
  $\hookrightarrow \mathcal{X}$ is also approximated by an Euler scheme X (if need be).
- The random variable (Z_n) must be discretised also in some sense.
  1. Non-linear specification: for some $\Theta \in \mathbb{R}^K$ where Θ stands for the coefficients of a Neural Network φ_{NN} and $Z_n = \varphi_{\mathrm{NN}}(\Theta, X_n)$
  2. Linear specification: $Z_n = \varphi_{\mathrm{L}}(\theta, X_{t_n})$ where

$$\varphi_L(\theta, \cdot) = \sum_{k=1}^{K} \theta_k \phi_k(\cdot), \quad \theta \in \mathbb{R}^{dK}$$

     where $(\phi_k)_{1 \leqslant k \leqslant K}$ are some basis functions.
- The discrete optimisation problem is now given by

$$\bar{V} := \inf_{\upsilon \in \mathbb{R}^{1+\bar{K}}} \mathbb{E}\big[|g(X_T) - \bar{Y}_T^{\upsilon}|^2\big] \neq 0$$

  where $\upsilon = (y, \Theta_0, \ldots, \Theta_{N-1})$ for the non-linear specification, or $\upsilon = (y, \theta_0, \ldots, \theta_{N-1})$ for the linear specification.

Comments on the discrete optimisation problem

- [HL20] proves two types of error control.
  - A posteriori error (for any $\upsilon \in \mathbb{R}^{1+\bar{K}}$)

$$\sup_{t\in[0,T]} \mathbb{E}\Big[|\mathcal{Y}_t - Y^{\upsilon}_{n[t]}|^2\Big] + \mathbb{E}\left[\int_0^T |\mathcal{Z}_t - Z^{\varphi}_{n[t]}|^2 \,\mathrm{d}t\right] \leqslant C\big(h + \mathbb{E}\big[|g(X_N) - Y^{y,\upsilon}_N|^2\big]\big)$$

(for $t \in [0, T]$ s.t. $t_k \leqslant t < t_{k+1}$ one sets $n[t] = k$)

  - Control by best approximation error:

$$\inf_{\upsilon\in\mathbb{R}^{1+\bar{K}}} \mathbb{E}\big[|g(X_T) - Y^{\upsilon}_T|^2\big] \leqslant C \left(h + \inf_{\theta} \mathbb{E}\left[\sum_{n=0}^{N-1} |\bar{z}_n(X_n) - \varphi(\theta_n, X_n)|^2\right]\right)$$

with $\bar{z}_n(X_n) = \frac{1}{h}\mathbb{E}\Big[\int_{t_n}^{t_n+1} \mathcal{Z}_t \,\mathrm{d}t | X_n\Big]$

- In practice, the discrete optimisation problem is solved by SGD. Thanks to the *A posteriori error control*, a small loss in the SGD procedure guarantees that one has a good estimation of the solution. (see next section for numerical illustration)

Outline

Introduction

- Let $u(\cdot)$ be a smooth solution to (denoting $v = \sigma(x,u)\partial_x u$)

$$\partial_t u + b(x,u,v)\partial_x u + \frac{1}{2}\sigma^2(x,u)\partial^2_{xx} u + f(x,u,v) = 0 \text{ and } u(T,\cdot) = g(\cdot)$$

- Solve $\mathrm{d}\mathcal{X}_t = b(\mathcal{X}_t, u(t,\mathcal{X}_t), v(t,\mathcal{X}_t))\,\mathrm{d}t + \sigma(\mathcal{X}_t, u(t,\mathcal{X}_t))\,\mathrm{d}W_t$
  then setting $(\mathcal{Y}_t, \mathcal{Z}_t) := (u(t,\mathcal{X}_t), v(t,\mathcal{X}_t))$, one gets,

$$\mathcal{X}_t = + \int_0^t b(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s + \int_0^t \sigma(\mathcal{X}_s, \mathcal{Y}_s)\,\mathrm{d}W_s \tag{41}$$

$$\mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s \tag{42}$$

- Equations (41)-(42) is a system of *fully coupled FBSDE*. They can represent the optimal control of an optimisation problem (see section I).
- I will give below an application to carbon markets modelling where the system (41)-(42) will describe the equilibrium price of allowances.

Numerical methods

- The fact that the system (41)-(42) is fully coupled renders its approximation quite intricate. We will consider the following (slightly) simplified system to present the numerical methods

$$\mathcal{X}_t = x + \int_0^t \sigma(\mathcal{Y}_s)\,\mathrm{d}W_s \text{ and } \mathcal{Y}_t = g(\mathcal{X}_T) + \int_t^T f(\mathcal{Z}_s)\,\mathrm{d}s - \int_t^T \mathcal{Z}_s\,\mathrm{d}W_s$$

- The first step in the numerical approximation will be to decoupled the system in a way.
- We will consider three methods:
  1. a global Picard Iteration method
  2. a forward method (the shooting method revisited)
  3. a layer method (the backward method revisited)

Picard Iteration method

- Initialise iteration with $\mathcal{Y}^0 = g(x)$ and compute from step $m-1$ to m:

$$\mathcal{X}_t^m = x + \int_0^t \sigma(\mathcal{Y}_s^{m-1})\,\mathrm{d}W_s \;\;\&\;\; \mathcal{Y}_t^m = g(\mathcal{X}_T^m) + \int_t^T f(\mathcal{Z}_s^m)\,\mathrm{d}s \;-\int_t^T \mathcal{Z}_s^m\,\mathrm{d}W_s$$

- Under structural conditions, the above scheme converges. To compute it in practice, one approximate the function u^m s.t. $\mathcal{Y}_t^m = u^m(t,\mathcal{X}_t^m)$. [BZ08] uses a sequence of Euler scheme to perform this approximation.
- Namely, Initialise with $\bar{u}_n^0(\cdot) = g(\cdot)$, for $0 \leqslant n \leqslant N$ and compute from step $m-1$ to m on the grid π:
  - An Euler scheme for the forward part: $X_{n+1}^m = X_n^m + \sigma(\bar{u}_n^{m-1}(X_n^m))\Delta W_n$
  - and the backward iteration on π: $Y_n^m = \mathbb{E}\left[Y_{n+1}^m|\mathcal{F}_{t_n}\right] + hf(Z_n^m)$, $Z_n^m = \mathbb{E}\left[Y_{n+1}^m\frac{\Delta W_n}{h}|\mathcal{F}_{t_n}\right]$

  $\hookrightarrow$ set then $\bar{u}_n^m$ s.t. $Y_n^m = \bar{u}_n^m(X_n^m)$.
- Computing the conditional expectation in practice allows to get an approximation of $\bar{u}_n^m$.

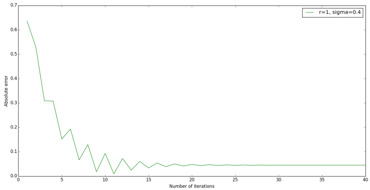
Numerical Illustration

- We choose $d = 4$ using an from [BZ08] (they run $d = 10$ in the paper)

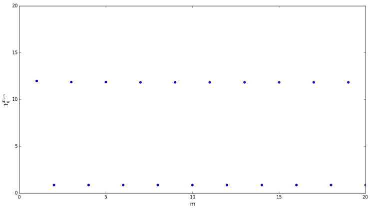
$$X_{i,t} = x_{i,0} + \int_0^t \sigma Y_s \, \mathrm{d}W_{i,s}, \quad 1 \leqslant i \leqslant d\,,$$
$$Y_t = \sum_{k=1}^{d} \sin(X_{k,T}) + \int_t^T -rY_s + \frac{1}{2} e^{-3r(T-s)} \sigma^2 \left(\sum_{k=1}^{d} \sin(X_{k,s}) \right)^3 \mathrm{d}s\,. \tag{43}$$

One verifies that (43) decouples via $Y_t = e^{-r(T-t)} \sum_{k=1}^{d} \sin(X_{k,t})\,.$

- We plot the sequence of Picard Iteration $m \mapsto Y_0^m$



(a) $r = 1, \sigma = 0.4$



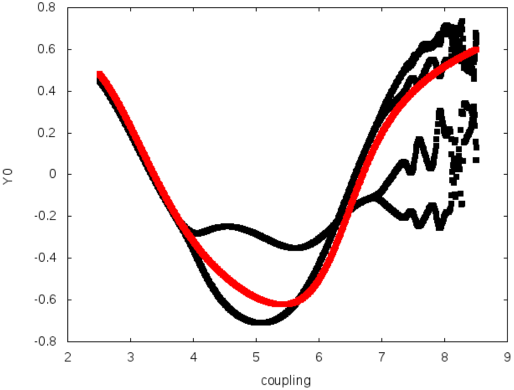
(b) $r = 0, \sigma = 0.5$

Numerical bifurcation

- The model:

$$\mathrm{d}X_t = \rho \cos(Y_t)\,\mathrm{d}t + \mathrm{d}W_t \text{ and } X_0 = x \in \mathbb{R}\ ,$$
$$dY_t = Z_t\,\mathrm{d}W_t \text{ and } Y_1 = \sin(X_1)\ .$$

- The important parameter is the coupling parameter ρ that will vary.



Deep BSDE solver

- The principle is the same as in the case of decoupled BSDEs.
- The controlled process is now given by

$$X_t^{y,Z} = x + \int_0^t \sigma(Y_s^{y,Z})\,\mathrm{d}W_s$$
$$Y^{y,Z} = y - \int_0^t f(Z_s)\,\mathrm{d}s + \int_0^t Z_s\,\mathrm{d}W_s$$

- Under some structural condition (see [HL20]), the optimisation

$$V := \min_{(y,Z)\in\mathbb{R}\times\mathcal{H}_2} \mathbb{E}\Big[|g(X_T^{y,Z}) - Y_T^{y,Z}|^2\Big] \tag{44}$$

  has unique argmin given by $(\mathcal{Y}_0, \mathcal{Z})$ (solution of the fully coupled FBSDE)
- [HL20] proposes a discrete time version of the optimisation solved using SGD (extending [HJW17, HJE18])

Numerical illustration

• This is example of Bender and Zhang now in dimension $d = 100$... (the figure below comes directly from [HL20]!)

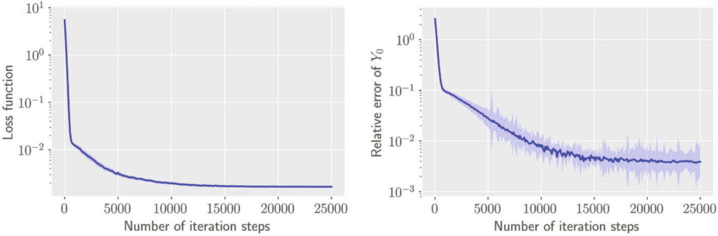


Fig. 1 Loss function (left) and relative approximation error of Y_0 (right) against the number of iteration steps in the case of Example 1 (100-dimensional). The proposed deep BSDE method achieves a relative error of size 0.39%. The shaded area depicts the mean ± the standard deviation of the associated quantity in 5 runs

Probabilistic layer methods

- These methods have been introduced in e.g. [MT00] (see also [Dou72]). They amount to solve the PDE on each time interval.
- **Definition:**
  - initialization: $u_N(\cdot) := g(\cdot)$ and $v_N(\cdot) = 0$
  - the transition from step $n+1$ to step n is as follows

$$X^{t_n,x}_{n+1} := x + \sigma(x, u_{n+1}(x))\Delta W_n$$

and

$$v_n(x) = \mathbb{E}\left[u_{n+1}(X^{t_n,x}_{n+1})\frac{\Delta W_n}{h}\right]$$
$$u_n(x) = \mathbb{E}\left[u_{n+1}(X^{t_n,x}_{n+1})\right] + hf(v_n(x))$$

$\hookrightarrow$ We observe that the decoupling is done using the predictor u_{n+1}.

- **Convergence:** Delarue and Menozzi [DM06] obtains convergence with a rate essentially if σ is uniformly elliptic.
- In **practice**, it requires the introduction of a discretization grid in space.

Carbon markets

- Carbon dioxide (CO_2) emission have a negative impact on the environment.
- Carbon markets are implemented to 'price' this and hopefully carbon emission reduction could be achieved
- Since 2005, the EU has had its own emissions trading system (ETS): an example of *cap-and-trade scheme*
  - A central authority set a limit on pollutant emission during a given period. Allowances are allocated to participating installations (via auctioning).
  - The total amount of allowances is the aggregated *cap*.
  - At the end of the period, each participating installation has to surrender an allowance for each unit of emission or pay a penalty.
  - During the period, participants can trade the allowances.

Main features

- Model based on FBSDEs see e.g. [CDET13, CD13, HS12]
- Three main processes on one period $[0, T]$.

1. The spot allowance price Y: we assume that the market is frictionless and arbitrage-free and that there is a probability such that $(e^{-rt}Y_t)_{0\leqslant t\leqslant T}$ is a martingale, namely

$$\mathrm{d}Y_t = rY_t\,\mathrm{d}t + Z_t\,\mathrm{d}W_t$$

   r is the interest rate, Z is a square integrable process.

2. Auxiliary process P:

$$\mathrm{d}P_t = b(P_t)\,\mathrm{d}t + \sigma(P_t)\,\mathrm{d}W_t$$

   Represent state variables that trigger the emission process (Electricity price or demand & fuel prices etc.) Fundamentals that are linked to goods emitting CO_2.

3. Emission process E: cumulative process with impact from the allowance price

$$\mathrm{d}E_t = \mu(P_t, Y_t)\,\mathrm{d}t$$

   $\hookrightarrow$ μ is decreasing in Y to take into account feedback of the allowance price

Results for one-period model

- From Carmona and Delarue [CD13], there exists a unique solution to:

$$\begin{aligned} \mathrm{d}P_t &= b(P_t)\,\mathrm{d}t + \sigma(P_t)\,\mathrm{d}W_t, && \text{(forward)} \\ \mathrm{d}E_t &= \mu(P_t, Y_t)\,\mathrm{d}t, && \text{(forward)} \\ \mathrm{d}Y_t &= rY_t\,\mathrm{d}t + Z_t\,\mathrm{d}W_t, && \text{(backward)} \end{aligned}$$

  with terminal condition: $\phi(E_T) = \rho \mathbf{1}_{\{E_T > \Lambda\}} \leqslant Y_T \leqslant \rho \mathbf{1}_{\{E_T \geqslant \Lambda\}} =: \phi_+(E_T)$.
  There exists a decoupling field s.t. $Y_t = u(t, P_t, E_t)$ for $t < T$.
- The decoupling field u is the "entropy" solution to

$$\partial_t u + \mu(p, u)\partial_e u + \mathcal{L}_p u = ru, \quad \text{and } u(T, e, p) = \phi(e) \tag{45}$$

  $\hookrightarrow$ u is Lipschitz in p and non decreasing in e.
  $\hookrightarrow$ $\partial_e u$ explodes at T near Λ, we only know $|\partial_e u(t, p, e)| \leqslant \frac{C}{T-t}$
  $\hookrightarrow$ Set $\mu(p, u) = -u$ and $\mathcal{L}_p = 0, r = 0$. One obtains a 'backward' inviscid Burgers equation...
- Multi-period model (finite or infinite number of period) [CCC23].

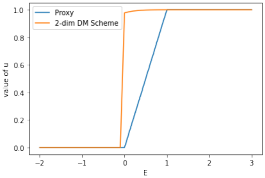
A numerical Toy model

- One-period Toy model ($r = 0$), dimension $d + 1$, $\sigma > 0$:
$$\mathrm{d}P_t = \sigma \,\mathrm{d}W_t, \;\; \mathrm{d}E_t = \left(\frac{1}{\sqrt{d}} \sum_{\ell=1}^{d} P_t^\ell - Y_t\right) \mathrm{d}t, \;\; \mathrm{d}Y_t = Z_t \cdot \mathrm{d}W_t,$$
and "$Y_T = \mathbf{1}_{[1,\infty)}(E_T)$".
- The quasi-linear pde associated is:
$$\partial_t u + \left(\frac{1}{\sqrt{d}} \sum_{\ell=1}^{d} p^\ell - u\right) \partial_e u + \frac{\sigma^2}{2} \sum_{\ell=1}^{d} \partial^2_{p_\ell p_\ell} u = 0$$
- Reduced to one dimension via $u(t, p, e) = \omega(t, e + (T - t)\frac{1}{\sqrt{d}} \sum_{\ell=1}^{d} p^\ell)$ with
$$\partial_t \omega - \omega \partial_\xi \omega + \frac{\sigma^2 (T - t)^2}{2} \partial^2_{\xi\xi} \omega = 0 \text{ and } \omega(T, \xi) = \mathbf{1}_{\{\xi \geqslant 1\}}$$

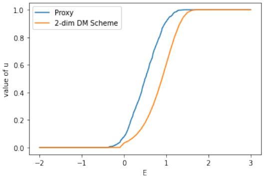
$\hookrightarrow$ Particle method associated to scalar conservation law can be used (Bossy, Jourdain, Tallay...) to get a proxy for the true solution: $e \mapsto u(0, 0, e)$.

Methods for fully coupled FBSDEs

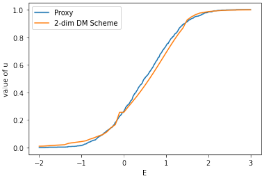
- Results for Delarue-Menozzi scheme [DM06]: (probabilistic layer method)



(c) $\sigma = 0.01$

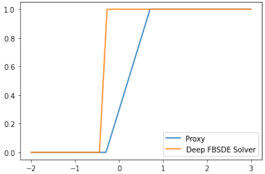


(d) $\sigma = 0.3$

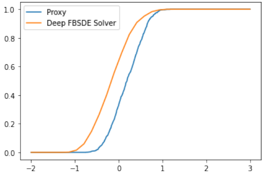


(e) $\sigma = 1.0$

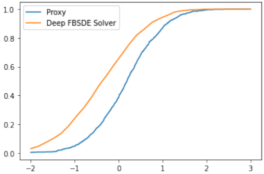
- Results for the deep FBSDE solver (learning error is small):



(f) $\sigma = 0.01$



(g) $\sigma = 0.3$



(h) $\sigma = 1.0$

A splitting scheme

- The numerical methods above fail to capture the correct weak solution.
- This comes from the degeneracy in e and the irregularity of the final condition. Many PDE methods would work, however the dimension of P is too 'big' in applications.
- We use a splitting scheme to treat both problem: on a time grid $\pi = (t_n)_{0 \leqslant n \leqslant N}$
  we iterate a *transport operator* (fixing p) and a *diffusion operator* (fixing e)
- The transport part is implemented using methods designed for discontinuous solution.
- Results [CY22b]:
  1. we prove the convergence of the splitting scheme with rate $\frac{1}{2}$, in the setting of existence and uniqueness for singular FBSDEs.
  2. we test the splitting scheme using various approximations of the transport operator and the diffusion part (regression).

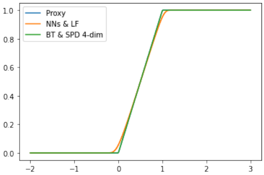
Implementation

- The transport operator is implemented using finite difference schemes: Upwind scheme or Lax-Friedrichs scheme, with J steps in space.
- The regression to estimate functions from $\mathbb{R}^d \to \mathbb{R}^J$ is computed using NN. (simple version of scheme in [HPW20])
- We develop also an alternative scheme: the regression is computed on a tree and transport operator approximated by a particles system. This works well for $d \leqslant 4$ and $P_t := f(t, W_t)$. (Convergence with a rate is proven in [CY22a]).
- We test also a multiplicative model:

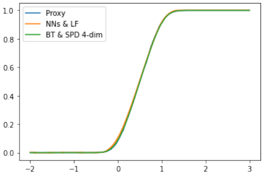
$$\mathrm{d}P_t^\ell = \mu P_t^\ell \,\mathrm{d}t + \sigma P_t^\ell \,\mathrm{d}W_t^\ell,\ P_0^\ell = 1, \text{ and } \mathrm{d}E_t = \tilde{\mu}(Y_t, P_t)\,\mathrm{d}t \tag{46}$$

with $\tilde{\mu}(y,p) = \left(\prod_{\ell=1}^d p^\ell\right)^{\frac{1}{\sqrt{d}}} e^{-\theta y}$, for some $\theta > 0$ and $\phi(p,e) = \mathbf{1}_{\{e \geqslant 0\}}$.
$\hookrightarrow$ *it can be reduced to a 2-dimensional model*!

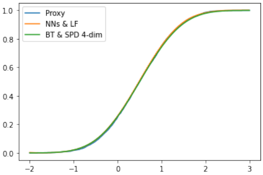
Some numerics on the Toy model



(i) $\sigma = 0.01$



(j) $\sigma = 0.3$

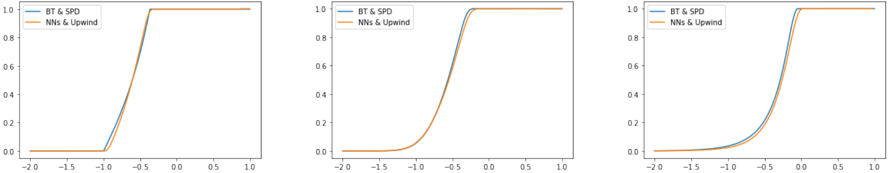


(k) $\sigma = 1.0$

Figure: Linear Toy Model: Comparison of the three methods:
- Neural Nets & Lax-Friedrichs (NN&LF) with $d = 10$
- an alternative scheme (BT&SPD) with $d = 4$
- The Proxy solution given by particle method.

Lax-Friedrichs scheme implemented with discretization of space $J = 1500, 1000, 500$, for $\sigma = 0.01, 0.3, 1$ respectively and number of time step $K = 30$. The number of time step for the splitting is $N = 64$. For $BT\&SPD$, the number of particles is $M = 3500$ and the number of time steps $N = 20$.

On the multiplicative model



(a) $\sigma = 0.01$ (b) $\sigma = 0.3$ (c) $\sigma = 1.0$

Figure: A multiplicative model in dimension $d = 10$. Comparison of two methods:
- Neural nets & Upwind scheme
- the alternative scheme on equivalent 4-dimensional model (BT&SPD).

The Upwind scheme used discretization of space $J = 100, 400, 500$ respectively for $\sigma = 1, 0.3, 0.01$ and number of time step $K = 20$. The number of time step for the splitting is $N = 32$. For $BT\&SPD$, the number of particles is $M = 3500$, and the number of time steps $N = 20$.

Conclusion

- Forward-Backward SDEs can represent the solution of various type of stochastic control problems. They also yield probabilistic representation of various class of non-linear PDEs
- Designing numerical probabilistic methods can be a good alternative to PDE methods: especially if the dimension of the state space is large.
- They appear to be very natural in the mean field setting (where the state space is infinite dimensional)
- They can also handle non markovian specification [BL14] (not discussed here)
- Time discretization combined with regression methods is now well understood even in the fully coupled case.
- The question of the curse of dimensionality is still present. Beyond the use of deep neural networks [AJK$^+$23], some methods are promising, for example: Multi-Level Picard [HJK$^+$19] or branching particles[HLTT14] (not discussed here).
- Challenges certainly remain for McKean-Vlasov FBSDEs and the approximation of PDEs with non linearity on the second order derivatives.

Numerical stability analysis for BSDEs schemes [CR15]

References I

- Julia Ackermann, Arnulf Jentzen, Thomas Kruse, Benno Kuckuck, and Joshua Lee Padgett, *Deep neural networks with relu, leaky relu, and softplus activation provably overcome the curse of dimensionality for kolmogorov partial differential equations with lipschitz nonlinearities in the lp-sense*, arXiv preprint arXiv:2309.13722 (2023).
- Bruno Bouchard, Ivar Ekeland, and Nizar Touzi, *On the malliavin approach to monte carlo approximation of conditional expectations*, Finance and Stochastics **8** (2004), no. 1, 45–71.
- Philippe Briand and Céline Labart, *Simulation of bsdes by wiener chaos expansion*, The Annals of Applied Probability **24** (2014), no. 3, 1129–1171.
- V. Bally and G. Pagès, *A quantization algorithm for solving multidimensional discrete-time optimal stopping problems*, Bernoulli **9** (2003), no. 6, 1003–1049.
- Bruno Bouchard and Nizar Touzi, *Discrete-time approximation and monte-carlo simulation of backward stochastic differential equations*, Stochastic Processes and their applications **111** (2004), no. 2, 175–206.
- Christian Bender and Jianfeng Zhang, *Time discretization and markovian iteration for coupled fbsdes*, The Annals of Applied Probability **18** (2008), no. 1, 143–177.

References II

- Jean-François Chassagneux, Hinesh Chotai, and Dan Crisan, *Modelling multiperiod carbon markets using singular forward-backward sdes*, Mathematics of Operations Research **48** (2023), no. 1, 463–497.
- Jean-François Chassagneux, Junchao Chen, and Noufel Frikha, *Deep runge-kutta schemes for bsdes*, arXiv preprint arXiv:2212.14372 (2022).
- René Carmona and François Delarue, *Singular FBSDEs and scalar conservation laws driven by diffusion processes*, Probability Theory and Related Fields **157** (2013), no. 1-2, 333–388.
- René Carmona, François Delarue, Gilles-Edouard Espinosa, and Nizar Touzi, *Singular forward–backward stochastic differential equations and emissions derivatives*, The Annals of Applied Probability **23** (2013), no. 3, 1086–1128.
- Jean-François Chassagneux, *Linear multistep schemes for bsdes*, SIAM Journal on Numerical Analysis **52** (2014), no. 6, 2815–2836.
- Dan Crisan and Konstantinos Manolarakis, *Solving backward stochastic differential equations using the cubature method: application to nonlinear pricing*, SIAM Journal on Financial Mathematics **3** (2012), no. 1, 534–571.

References III

Dan Crisan, Konstantinos Manolarakis, and Nizar Touzi, *On the monte carlo simulation of bsdes: An improvement on the malliavin weights*, Stochastic Processes and their Applications **120** (2010), no. 7, 1133–1158.

Jean-François Chassagneux and Adrien Richou, *Numerical stability analysis of the euler scheme for bsdes*, SIAM Journal on Numerical Analysis **53** (2015), no. 2, 1172–1193.

Jean-François Chassagneux and Camilo A Garcia Trillos, *Cubature methods to solve BSDEs: Error expansion and complexity control*, arXiv preprint arXiv:1702.00999 (2017).

Jean-François Chassagneux and Mohan Yang, *Convergence of particles and tree based scheme for singular fbsdes*, arXiv preprint arXiv:2212.11917 (2022).

______, *Numerical approximation of singular forward-backward sdes*, Journal of Computational Physics **468** (2022), 111459.

François Delarue and Stéphane Menozzi, *A forward–backward stochastic algorithm for quasi-linear pdes*, The Annals of Applied Probability **16** (2006), no. 1, 140–184.

Avron Douglis, *Layering methods for nonlinear partial differential equations of first order*, Annales de l'institut Fourier, vol. 22, 1972, pp. 141–227.

References IV

Nicole El Karoui, Shige Peng, and Marie Claire Quenez, *Backward stochastic differential equations in finance*, Mathematical finance **7** (1997), no. 1, 1–71.

Emmanuel Gobet, José G López-Salas, Plamen Turkedjiev, and Carlos Vázquez, *Stratified regression monte-carlo scheme for semilinear pdes and bsdes with large scale parallelization on gpus*, SIAM Journal on Scientific Computing **38** (2016), no. 6, C652–C677.

Emmanuel Gobet, Jean-Philippe Lemor, and Xavier Warin, *A regression-based monte carlo method to solve backward stochastic differential equations*, The Annals of Applied Probability **15** (2005), no. 3, 2172–2202.

Emmanuel Gobet and Plamen Turkedjiev, *Approximation of backward stochastic differential equations using malliavin weights and least-squares regression*, Bernoulli **22** (2016), no. 1, 530–562.

———, *Linear regression mdp scheme for discrete backward stochastic differential equations under general conditions*, Mathematics of Computation **85** (2016), no. 299, 1359–1391.

References V

E Gobet and P Turkedjiev, *Adaptive importance sampling in least-squares monte carlo algorithms for backward stochastic differential equations*, Stochastic Processes and their applications **127** (2017), no. 4, 1171–1203.

J. Han, A. Jentzen, and W. E, *Solving high-dimensional partial differential equations using deep learning*, Proceedings of the National Academy of Sciences **115** (2018), no. 34, 8505–8510.

Martin Hutzenthaler, Arnulf Jentzen, Thomas Kruse, et al., *On multilevel picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations*, Journal of Scientific Computing **79** (2019), no. 3, 1534–1571.

Jiequan Han, Arnulf Jentzen, and E Weinan, *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, Communications in Mathematics and Statistics **5** (2017), no. 4, 349–380.

Jiequn Han, Arnulf Jentzen, and E Weinan, *Solving high-dimensional partial differential equations using deep learning*, Proceedings of the National Academy of Sciences **115** (2018), no. 34, 8505–8510.

References VI

Jiequan Han and Jihao Long, *Convergence of the deep bsde method for coupled fbsdes*, Probability, Uncertainty and Quantitative Risk **5** (2020), no. 1, 1–33.

Pierre Henry-Labordere, Xiaolu Tan, and Nizar Touzi, *A numerical algorithm for a class of bsdes via the branching process*, Stochastic Processes and their Applications **124** (2014), no. 2, 1112–1140.

Côme Huré, Huyên Pham, and Xavier Warin, *Deep backward schemes for high-dimensional nonlinear pdes*, Mathematics of Computation **89** (2020), no. 324, 1547–1579.

Sam Howison and Daniel Schwarz, *Risk-neutral pricing of financial instruments in emission markets: a structural approach*, SIAM Journal on Financial Mathematics **3** (2012), no. 1, 709–739.

Peter Eris Kloeden, Eckhard Platen, and Henri Schurz, *Numerical solution of SDE through computer experiments*, Springer Science & Business Media, 2012.

Jean-Philippe Lemor, Emmanuel Gobet, and Xavier Warin, *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Bernoulli **12** (2006), no. 5, 889–916.

References VII

G Milstein and M Tretyakov, *Numerical algorithms for semilinear parabolic equations with small parameter based on approximation of stochastic equations*, Mathematics of computation **69** (2000), no. 229, 237–267.

Etienne Pardoux and Shige Peng, *Adapted solution of a backward stochastic differential equation*, Systems & Control Letters **14** (1990), no. 1, 55–61.

______, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Stochastic partial differential equations and their applications, Springer, 1992, pp. 200–217.

Gilles Pagès and Abass Sagna, *Improved error bounds for quantization based numerical schemes for bsde and nonlinear filtering*, Stochastic Processes and their Applications **128** (2018), no. 3, 847–883.

Marjon J Ruijter and Cornelis W Oosterlee, *A fourier cosine method for an efficient computation of solutions to bsdes*, SIAM Journal on Scientific Computing **37** (2015), no. 2, A859–A889.

Jianfeng Zhang, *A numerical scheme for bsdes*, the annals of applied probability **14** (2004), no. 1, 459–488.

References VIII

Jianfeng Zhang and Jianfeng Zhang, *Backward stochastic differential equations*, Springer, 2017.