Backward SDEs, Lecture II

Existence, Stability,and Numerical Methods

Coxeter Lectures, Fields Institute, Toronto

Nicole EL KAROUI

Paris VI University, Ecole Polytechnique,

nicole.elkaroui@cmap.polytechnique.fr

(with the financial support of the Fondation du risque,)

Backward Stochastic Differential Equation

- Standard filtred probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), 0 \leq t \leq T, \mathbb{P})$, supporting a standard BM $W \in \mathbb{R}^n$.
- A non anticipating coefficient $f(t, \omega, y, z)$ defined on $(\Omega \times \mathbb{R}^+, \mathbb{R}^d \times \mathbb{R}^{d \times n})$, a terminal condition $\xi_T \in \mathcal{F}_T$

Definition of BSDE solution

A solution of BSDE(f, ξ_T), is a par of non anticipating processes $(Y, Z) \in \mathbb{R}^d \times \mathbb{R}^{d \times n}$ such that

- $Y_t = \xi_T + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s,$
- or equivalently $-dY_t = f(t, Y_t, Z_t)ds - Z_t dW_t, \quad Y_T = \xi_T$
- with minimal integrability condition, $\int_0^T (|f(t, Y_t, Z_t)| + |Z_t|^2)dt < \infty$ *a.s.*

- Existence, Uniqueness ? : in which spaces of processes,...
- Properties ? : Stability, Comparison Theorem.....

Doob Inequalities

Notation for the running maximum : $\max |M|_T = \sup_{0,T} |M|_s$

Continuous Martingale : a priori estimates

- Doob inequalities :

  $\mathbb{E}[\max |M|^2_T] \leq c\mathbb{E}[|M_T|^2] \leq C\mathbb{E}[\max |M|^2_T]$

  Should be read in both directions ($A \leq B \leq C$)

  - $B \implies A$ is a Backward inequality
  - $C \implies B$ is a Forward inequality

- Burkholder, Davis Gundy inequalities

  Let $\langle M \rangle$ be the a quadratic variation of M, then for any $p > 0$

  $\mathbb{E}[\max |M|^p_T] \leq c_p\mathbb{E}[|M_T|^p/2] \leq C_p\mathbb{E}[\max |M|^p_T]$

Representation Theorem

A priori Forward or Backward Estimates

Weighted $\mathbb{H}^2_T$ space

- Forward $\mathbb{H}^2_c$, defined as $\mathbb{H}^2_T$ with the semi-norm $\|X\|^2_c = \max(e^{-2ct}\mathbb{E}[\max|X|^2_t])_T$
- Backward $\mathbb{H}^2_c$, defined as $\mathbb{H}^2_T$ with the semi-norm $\|X\|^2_\beta = \max(e^{2\beta t}\mathbb{E}[\max|X|^2_t])_T$

Estimates of $F^T_t = \int_t^T f_s ds$ a finite variation process.

- Forward $|F^T_t|^2 = |\int_t^T e^{sc/2}(e^{-sc/2} f_s)ds|^2 \leq e^{cT}\frac{1}{c}\int_t^T e^{-cs}|f_s|^2 ds$
- Backward $|F^T_t|^2 = |\int_t^T e^{-s\beta/2}(e^{s\beta/2} f_s)ds|^2 \leq e^{-\beta t}\frac{1}{\beta}\int_t^T e^{s\beta}|f_s|^2 ds$

Semimartingale Estimates

Let $x_T = x_t - \int_t^T f_s ds - \int_t^T \eta_s dW_s$ a Itô's semimartingale

- Forward

  Since $x_t = x_0 - \int_0^t f_s ds - \int_0^t \eta_s dW_s$, then
  $|x_t| \leq |y_0| + |F_0^t| + \max|\eta.W|_t$. By the Doob inequality,
  $$e^{-ct}\mathbb{E}[\max|x|_t^2] \leq \mathbb{E}\Big[e^{-ct}|x_0|^2 + \frac{1}{c}\int_0^t e^{-cs}(|f_s|^2 + |\eta_s|^2)ds\Big]$$
  $$\|x\|_c^2 \leq 2\mathbb{E}\Big[e^{-cT}|x_0|^2 + \frac{1}{c}\int_0^T e^{-sc}(|f_s|^2 + |\eta_s|^2)ds\Big]$$

- Backward

  By Doob inequality, since $|x_t| \leq \mathbb{E}[|x_T| + |F_t|\,|\mathcal{F}_t]$,
  $$e^{t\beta/}|x_t| \leq \mathbb{E}\Big[\Big(e^{T\beta/2}|x_T| + \frac{1}{\beta}\int_t^T e^{s\beta}|f_s|^2 ds\Big)^{1/2}|\mathcal{F}_t\Big]$$
  $$\begin{cases} \|x\|_\beta^2 & \leq 4\mathbb{E}[e^{T\beta}|x_T|^2 + \frac{1}{\beta}\int_0^T e^{s\beta}|f_s|^2 ds] \\ \|\eta.W\|_\beta^2 & \leq K\Big[\mathbb{E}[e^{T\beta}|x_T|^2 + \frac{1}{\beta}\int_0^T e^{s\beta}|f_s|^2 ds\Big] \end{cases}$$

Lipschitz Assumptions

Forward Assumptions

- $F(t,[x]_t)$, and $G(t,[x]_t)$ (path dependency) in $\mathbb{L}^2$
- Uniformly Lipschitz i.e, there exists $K > 0$ s.t a.s
  $|F(t,[x^1]_t) - F(t,[x^2]_t)| + |G(t,[x^1]_t) - G(t,[x^2]_t)| \leq K\,|[x_1 - x_2]|$

Backward Assumptions

- Standard data (f, ξ) : $\int_0^T |f(t,0,0)|^2 ds, \xi \in \mathbb{L}^2$
- f is uniformly lipschitz, i.e., there exists $C > 0$ s.t a.s
  $|f(t,y_1,z_1) - f(t,y_2,z_2)| \leq C\,(|y_1 - y_2| + |z_1 - z_2|)$

Notations : given two coefficients f^1, f^2,

- $\delta Y_t = Y_t^1 - Y_t^2$, $\delta Z_t = Z_t^1 - Z_t^2$
- $\delta_2 f_t = f^1(t,y_2,z_2) - f^2(t,y_2,z_2)$, $\delta_2 F_t = \delta_2 f_t(Y_t^2, Z_t^2)$

Solutions via Picard Approximations

- Forward Lipschitz SDE

  $dX_t =$

  $F(t,[X]_t)dt, + G(t,[X]_t)dW_t$

  - General filtration
  - Standard $\mathbb{L}^2$ multi-dim data (X_0, F, G), uniformly Lipschitz.
- Existence and Uniqueness
  - $\exists$ a unique solution in $\mathbb{H}^2_T$

- Backward Lipschitz SDE

  $-dY_t = f(t, Y_t, Z_t)dt, -Z_t.dW_t$,

  $Y_T = \xi_T$

  - Brownian Filtration
  - Standard $\mathbb{L}^2$ multi-dim data

    uniformly Lipschitz.
- Existence and Uniqueness
  - $\exists$ a unique pair $(Y, Z) \in \mathbb{H}^2$

In the both cases, the Picard sequence converges uniformly in the right $\mathbb{H}^2_T$ space to the solution with an exponential speed. The estimates are uniform in the boundary conditions.

General Markovian Setting

Let X be a diffusion process on a general filtered probability space, and $\mathcal{B}_e$ be the $\sigma-$ field on $\mathbb{R}^n$ generated by $\mathbb{E}\int_t^T \phi(s, X_s^{t,x})ds$ where ϕ is a continuous bounded. Let $(f, \Psi) \in \mathcal{B}_e$ be squared integrable $(\mathbb{E}\int_0^T f^2(s, X_s^{t,x})ds < +\infty$; $\mathbb{E}[\Psi^2(X_T^{t,x})] < +\infty,)$

- Markovian representation of the solution[CJPS]

  The semimartingale $Y_s^{t,x} = \mathbb{E}[\Psi(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x})dr|\mathcal{F}_s]$ admits a continuous version given by $u(s, X_s^{t,x})$ with $u(t,x) = Y_t^{t,x} \in \mathcal{B}_e$

- Markovian representation of the martingale Moreover, $u(t,x) + \int_t^s f(r, X_r^{t,x})dr + Y_s^{t,x} = U_s^{t,x}$ is an additive martingale with the following representation depending on $d(t,x) \in \mathcal{B}_e$,

  $U_s^{t,x} = \int_t^s \underbrace{d(r, X_r^{t,x})^* \sigma(r, X_r^{t,x})}_{Z_r^{t,x}*} dW_r$; $t \leq s$

Markovian BSDEs

Let X be a diffusion process and the associated BSDE :

$$-dY_s = f(s, X_s^{t,x}, Y_s, Z_s)ds - Z_s^* dW_s, \;\; Y_T = \Psi(X_T^{t,x})$$

- General setting : Thanks to Picard approximates, there exists $u(t,x), d(t,x) \in \mathcal{B}_e$ such that
  $Y_s = u(s, X_s^{t,x}), Z_s = d(s, X_s^{t,x})^* \sigma(s, X_s^{t,x})$.
- PDE solution in one dimensional case
  Let $\mathcal{L}^X$ the elliptic operator associated with the diffusion X.
  Then, under mild regularity assumptions, u is a viscosity solution of the HJB Type PDE
  $$\begin{cases} \partial_t u(t,x) & + & \mathcal{L}v(t,x) + f(t,x,u(t,x), \partial_x u(t,x)\sigma(t,x)) = 0 \\ u(T,x) & = & \Psi(x) . \end{cases}$$
  Then, $d(t,x)$ plays the role of $\partial_x u$ the gradiant of u.
  proof is provided by the strict comparison theorem.

Linear growth assumption, d=1

For simplicity, we assume that $f(t,0,0)=0$

Linear growth : $|f(t,y,z)| \leq g_\mu(y,z) = a|y| + \mu|z|S$ Let $\overline{Y}^\mu$ the solution of the Lipschitz BSDE with coefficient g_μ and $\underline{Y}^\mu$ the process $-\overline{Y}^\mu(-\xi_T)$. Uniform bounds

Then any square integrable solution (Y,Z) of BSDE(f) with linear growth satisfies

$$\underline{Y}^\mu \leq Y \leq \overline{Y}^\mu$$

Lepeltier,San Martin,'97

There exists a minimal (a maximal)solution to the BSDE with GL continuous coefficient.

General methodology

The different steps of the proof are the following

- Use a monotone Lipschitz regularisation f^n of f, with same linear growth
- Show that the solutions (Y^n, Z^n) are bounded in L^2, $\mathbb{E}[\int_0^T |Z_s|^2 ds] \leq C$
- Show the control of $\mathbb{E}[\int_0^T |\delta^{i,j} Z_s|^2 ds]$ by $\left(\mathbb{E}[\int_0^T |\delta^{i,j} Y_s|^2 ds]\right)^{1/2}$
- Use the motonocity of the sequence Y^n and the previuos estimates to show that Z^n converges strongly in $\mathbb{H}^2$ to Z, and so Y^n converges uniformly to Y
- The last step uses the property of the approximating seauence to show that $f^n(t, Y^n, Z^n)$ also converges to $f(t, Y, Z)$

Sketch of the proof

Regularisation by inf convolution

$f^n(x) = inf_{y\in\mathbb{R}^p}\{f(y) + n|x - y|\}$ is well defined for $n \geq \sup(a, \mu) = K$

Key inequality Denote by $Y^{i,j} = \delta^{i,j}Y$ the difference between Y^i and Y^j.

By Itos formula

$$|Y^{i,j}|_t^2 + \int_t^T |Z^{i,j}|_s^2 ds\mathbb{E}_t[\int_t^T |Z_s|^2 ds]$$

Reflected BSDEs around a regular obstacle

How to maintain a BSDE solution above a given regular obstacle ?

Assume $dO_t = U_t dt + V_t dW_t$

. Let (Y, Z) a solution of BSDE(f,ξ_T)

By comparison theorem, if $\xi_T \geq O_T$, and $f(t, O_t, V_t) + U_t \geq 0$, then $Y_t \geq O_t \forall t$

The idea is to push the solution above O_t by adding some "cash", when you need, $f(t, O_t, V_t) + U_t \leq 0$, in a minimal way. Working with $Y_t - O_t$, the problem may be rewritten as to push a solution of BSDE above 0.

Definition of Reflected BSDE Above 0

$$
\begin{cases}
Y_t = \Phi + \int_t^T f(s, Y_s, Z_s)ds + \mathbf{K_T} - \mathbf{K_t} - \int_t^T Z_s dW_s, \\
\mathbf{Y_t} \geq \mathbf{O_t}, \\
K \text{ is continuous, increasing, } K_0 = 0 \text{ and } \int_0^T Y_t dK_t = 0.
\end{cases}
$$

The above observation suggests to be looking for a process K absolutely continuous w.r. to $f(t, 0, 0)^- dt$,

$$
dK_t = \alpha_t \mathbf{1}_{\{Y_t = 0\}} f(t, 0, 0)^- dt, \ \alpha_t \in [0, 1]
$$

Transformation of the problem

The problem is now expressed in terms of α_t.

Regularization

Let ϕ^n a Lipschitz regularization of $\mathbf{1}_{\{y=0\}}$, bounded by 1, and decreasing.

• By the same method that above, one show the same properties holds true, for the BSDEs with $f^n = f + \phi^n(y)$ • to show that the sequence Y^n converges uniformly, and Z^n strongly in L^2 to a pair (Y, Z), with $Y \geq 0$.

• The only small difficulty is to show that dK_t^n converges to a solution with support $\{Y_t = 0\}$

Applications to optimal stopping problems

General obstacle Lower bound. For any stopping time $\tau \in \mathcal{T}_{t,T}$, one has

$$\begin{aligned} Y_t &= \mathbb{E}(Y_t + \int_t^T f(s, Y_s, Z_s)ds + K_\tau - K_t - \int_t^T Z_s dW_s | \mathcal{F}_t) \\ &\geq \mathbb{E}(O_t \mathbf{1}_{\tau<T} + \Phi \mathbf{1}_{\tau=T} + \int_t^\tau f(s, Y_s, Z_s)ds | \mathcal{F}_t), \end{aligned}$$

which implies

$$\mathbf{Y_t} \geq \operatorname*{ess\ sup}_{\tau \in \mathcal{T}_{\mathbf{t},\mathbf{T}}} \mathbb{E}(\mathbf{O}_\tau \mathbf{1}_{\tau<\mathbf{T}} + \text{˘}\mathbf{1}_{\tau=\mathbf{T}} + \int_{\mathbf{t}}^\tau \mathbf{f}(\mathbf{s}, \mathbf{Y_s}, \mathbf{Z_s})\mathbf{ds} | \mathcal{F}_{\mathbf{t}}).$$

Equality. The equality holds for $\tau^* = \inf\{u \in [t, T] : Y_u = O_u\} \wedge T$.

Numerical Point of view

New interest for these kind with the swing options, the real options.

$\Longrightarrow$ The regular obstacle method is very interesting for numerical methods since

- it gives an upper approximation (the penalisation app. gives a lower bound).
- the bounds on the approximated driver depends less on n than for the penalisation scheme.
- No available estimates on the rate of convergence w.r.t. n.

- Thanks to Emmanuel Gobet to allows me to use its beautiful presentation of the numerical aspect of BSDEs
- The complete presentation may be find on the following site :
- http ://www.cmap.polytechnique.fr ?euroschoolmathfi09
  Then, go to minicours
  Find the slides of E.Gobet and J.Ma on BSDEs

Our aim :

- to simulate Y and Z
- to estimate the error, in order to tune finely the convergence parameters.

Quite intricate and demanding since

- it is a non-linear problem (the current process dynamics depen on the future evolution of the solution).
- it involves various deterministic and probabilistic tools (also from statistics).
- the estimation of the convergence rate is not easy because of the non-linearity, of the loss of independance (mixing of independent simulations).

Strong approximation.

$(X_t^N)_{0\leq t\leq T}$ is a strong approximation of $(X_t)_{0\leq t\leq T}$ if

$$\sup_{t\leq T}\|X_t^N - X_t\|_{\mathbb{L}_p}\to 0 \text{ (or } \|\sup_{t\leq T}|X_t^N - X_t|\|_{\mathbb{L}_p}\to 0\text{) as } N \text{ goes to } \infty.$$

Weak approximation. For any test function (smooth or non smooth), one has

$$\mathbb{E}[f(X_T^N)] - \mathbb{E}[f(X_T)]\to 0 \text{ as } N \text{ goes to } \infty.$$

Examples.

Approximation of SDE : $X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$.

Time discretization using **Euler scheme**. Define $t_k = k\frac{T}{N} = kh$.

$$X_0^N = x,\ X_{t_{k+1}}^N = X_{t_k}^N + b(t_k, X_{t_k}^N)h + \sigma(t_k, X_{t_k}^N)(W_{t_{k+1}} - W_{t_k}).$$

The simplest scheme to use. Converges at rate $\frac{1}{2}$ for strong approximation and 1 for weak approximation.

Milshtein scheme (not available for arbitrary σ) : rate 1 for both strong and weak approximations.

The BSDE case

We focus mainly on Markovian BSDE :

$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s,$ where X is a forward SDE. We know that $Y_t = u(t, X_t)$ and $Z_t = \nabla_x u(t, X_t)\sigma(t, X_t)$, where u solves a semi-linear PDE

$\Longrightarrow$ to approximate Y, Z, we need to approximate the function $u(\cdot)$, the gradiant of u and the process X

- $Y_t^N = u^N(t, X_t^N)$,
- in practice, X^N is always random,
- although u is deterministic, u^N may be random (e.g. Monte Carlo approximations) : **the randomness may come from two different objects**.

Formal error analysis

$$\begin{aligned}\mathbb{E}|Y_t^N - Y_t| &\le \mathbb{E}|u^N(t, X_t^N) - u(t, X_t^N)| + \mathbb{E}|u(t, X_t^N) - u(t, X_t)| \\ &\le |u^N(t,\cdot) - u(t,\cdot)|_{\mathbb{L}_\infty} + \|\nabla u\|_{\mathbb{L}_\infty}\mathbb{E}|X_t^N - X_t|.\end{aligned}$$

Two source of error :

- strong error related to $\mathbb{E}|X_t^N - X_t|$.
  For the Euler scheme $\mathbb{E}|X_t^N - X_t| = O(N^{-1/2})$.
- weak error related to $|u^N(t,\cdot) - u(t,\cdot)|_{\mathbb{L}_\infty}$.
  Indeed, to see that this is a weak-type error, take $f \equiv 0$, $u(t,x) = \mathbb{E}[\Phi(X_T)|X_t = x]$, and the Euler scheme to approximate the conditional law of X_T : from $[BT96]$, one knows

The grid

Time grid :

$$\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$$

with non uniform time step : $|\pi| = max_i(t_{i+1} - t_i)$.

We write $\Delta t_i = t_{i+1} - t_i$ and $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$.

Heuristic derivation

From $Y_{t_i} = Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s)ds - \int_{t_i}^{t_{i+1}} Z_s dW_s$, we derive

$$Y_{t_i} = \mathbb{E}[Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s)ds|\mathcal{F}_{t_i}],$$

$$\mathbb{E}[\int_{t_i}^{t_{i+1}} Z_s ds|\mathcal{F}_{t_i}] = \mathbb{E}[(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s)ds)\Delta W^*_{t_i}|\mathcal{F}_{t_i}]$$

Discrete backward iteration.

$$\Longrightarrow \begin{cases} \mathbf{Z}^N_{t_i} = \frac{1}{\Delta \mathbf{t_i}}\mathbb{E}[\mathbf{Y}^N_{t_{i+1}} \Delta \mathbf{W}^*_{t_i}|\mathcal{F}_{t_i}], \\ \mathbf{Y}^N_{t_i} = \mathbb{E}[\mathbf{Y}^N_{t_{i+1}} + \Delta \mathbf{t_i f}(\mathbf{t_i}, \mathbf{X}^N_{t_i}, \mathbf{Y}^N_{t_{i+1}}, \mathbf{Z}^N_{t+i})|\mathcal{F}_{t_i}] \text{ and } \mathbf{Y}^N_T = \breve{\ }(\mathbf{X}^N_T). \end{cases}$$

The scheme is of **explicit** type.

Implicit scheme

More closely related to the idea of discret BSDE.

$$(\mathbf{Y}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Z}^{\mathbf{N}}_{\mathbf{t_i}}) = \arg \min_{(Y,Z)\in\mathbb{L}_2(\mathcal{F}_{t_i})} \mathbb{E}[\mathbf{Y}^{\mathbf{N}}_{\mathbf{t_{i+1}}} + ´\mathbf{t_i f}(\mathbf{t_i}, \mathbf{X}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Y}, \mathbf{Z}) - \mathbf{Y} - \mathbf{Z}´\mathbf{W_{t_i}}]^2,$$

with $Y^N_{t_N} = \Phi(X^N_{t_N})$.

$$\boxed{\to \begin{cases} \mathbf{Z}^{\mathbf{N}}_{\mathbf{t_i}} = \frac{1}{´\mathbf{t_i}}\mathbb{E}[\mathbf{Y}^{\mathbf{N}}_{\mathbf{t_{i+1}}}´\mathbf{W}^*_{\mathbf{t_i}}|\mathcal{F}_{t_i}], \\ \mathbf{Y}^{\mathbf{N}}_{\mathbf{t_i}} = \mathbb{E}[\mathbf{Y}^{\mathbf{N}}_{\mathbf{t_{i+1}}}|\mathcal{F}_{t_i}] + ´\mathbf{t_i f}(\mathbf{t_i}, \mathbf{X}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Y}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Z}^{\mathbf{N}}_{\mathbf{t_i}}). \end{cases}}$$

Needs a Picard iteration procedure to compute $Y^N_{t_i}$.

Well defined for $|\pi|$ small enough (f Lipschitz).

Define the measure of the squared error

$$\mathcal{E}(Y^N - Y, Z^N - Z) = \max_{0\le i\le N} \mathbb{E}|Y^N_{t_i} - Y_{t_i}|^2 + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}|Z^N_{t_i} - Z_t|^2 dt.$$

Theorem. For a Lipschitz driver w.r.t. (x, y, z) and $\frac{1}{2}$-Holder w.r.t. t, one has

$$\mathcal{E}(Y^N - Y, Z^N - Z) \le C(\mathbb{E}|\Phi(X^N_T) - \Phi(X_T)|^2 + \sup_{i\le N} \mathbb{E}|X^N_{t_i} - X_{t_i}|^2 + |\pi| + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}|Z_t - \bar{Z}_{t_i}|^2 dt).$$

where $\bar{Z}_{t_i} = \frac{1}{\Delta t_i}\mathbb{E}(\int_{t_i}^{t_{i+1}} Z_s ds|\mathcal{F}_{t_i})$

Error Analysis

→ Different error contributions :

- Strong approximation of the forward SDE (depends on the forward scheme and not on the BSDE-problem)
- Strong approximation of the terminal conditions (depends on the forward scheme and on the BSDE-data Φ)
- L^2-regularity of Z (intrinsic to the BSDE-problem).

Diffusion approximation

The easy part : using the Euler scheme

- $\sup_{i \leq N} |X^N_{t_i} - X_{t_i}|_{\mathbb{L}_2} = O(N^{-1/2})$.
- If σ does not depend on x, rate $O(N^{-1})$.
- Overwise, Milshtein scheme to get N^{-1}-rate.

Strong approximation of the terminal condition

- If Φ Lipschitz, then $\mathbb{E}|\Phi(X_T^N) - \Phi(X_T)|^2 \leq L_\Phi^2 \mathbb{E}|X_T^N - X_T|^2$.
- New result if Φ is irregular, using the approximation theory
  Some results of Avikainen [*Avi*09] for discontinuous function $\Phi(x) = \mathbf{1}_{x \leq a}$.
- Possible generalization to functions with bounded variation [Avikainen '09]
- For intermediare regularity functions, open questions.

$\mathcal{E}^Z(\pi) = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|Z_{t_i}^N - Z_t|^2 dt$. **Theorem. [Convergence to 0]**

Theorem. [Ma, Zhang '02 '04]

Assume a Lipschitz driver f and a Lipschitz terminal condition Φ.

Then Z is a continous process and $\mathcal{E}^Z(\pi) = O(|\pi|)$ for any time-grid π.

No ellipticity assumption.

Key fact : Z can be represented via a linear BSDE ! ! It is proved using the Malliavin calculus representation of Z component.

The basics of Malliavin calculus :

Sensitivity of Wiener functionals w.r.t. the BM

For $\xi = \xi(W_t : t \geq 0)$, its Malliavin derivative $(\mathcal{D}_t\xi)_{t\geq 0}$ $\in \mathbb{L}_2(\mathbb{R}^+ \times \Omega, dt \otimes d\mathbb{P})$ is defined as

$$"\mathcal{D}_\mathbf{t}\xi = \partial_{\mathbf{dW_t}}\xi(\mathbf{W_t} : \mathbf{t} \geq \mathbf{0})."$$

Basic rules.

- If $\xi = \int_0^T h_t dW_t$ with $h \in \mathbb{L}_2(\mathbb{R}^+)$, $\mathcal{D}_t\xi = h_t \mathbf{1}_{t\leq T}$.
- For smooth random variables $X = g(\int_0^T h_t^1 dW_t, \dots, \int_0^T h_t^n dW_t)$, $\mathcal{D}_t X = \sum_{i=1}^n \partial_i g(\dots) h_t^i \mathbf{1}_{t\leq T}$.
- Duality relation with adjoint operator $\mathcal{D}^*$:

Malliavin derivatives of (Y, Z) for smooth data

Theorem.

If $Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s$, then for $\theta \le t \le T$

$$\mathcal{D}_\theta Y_t = \Phi'(X_T)\mathcal{D}_\theta X_T + \int_t^T [f'_x(s, X_s, Y_s, Z_s)\mathcal{D}_\theta X_s$$

$$+ f'_y(s, X_s, Y_s, Z_s)\mathcal{D}_\theta Y_s + f'_z(s, X_s, Y_s, Z_s)\mathcal{D}_\theta Z_s]ds - \int_t^T \mathcal{D}_\theta Z_s dW_s$$

$\Longrightarrow (\mathcal{D}_\theta Y_t, \mathcal{D}_\theta Z_t)_{t\in[0,T]}$ solves a linear BSDE (for fixed θ).

In addition :

- Viewing the BSDE as FSDE, one has $\mathbf{Z_t} = \mathcal{D}_\mathbf{t}\mathbf{Y_t}$.
- Due to $\mathcal{D}_\theta \mathbf{X_t} = \nabla \mathbf{X_t}[\nabla \mathbf{X}_\theta]^{-1}\sigma(\theta, \mathbf{X}_\theta)$, we get

$$(\mathcal{D}_\theta \mathbf{Y_t}, \mathcal{D}_\theta \mathbf{Z_t}) = (\nabla \mathbf{Y_t}[\nabla \mathbf{X}_\theta]^{-1}\sigma(\theta, \mathbf{X}_\theta), \nabla \mathbf{Z_t}[\nabla \mathbf{X}_\theta]^{-1}\sigma(\theta, \mathbf{X}_\theta)),$$

where

$$\nabla Y_t = \Phi'(X_T)\nabla X_T + \int_t^T [f'_x(s, X_s, Y_s, Z_s)\nabla X_s$$
$$+ f'_y(s, X_s, Y_s, Z_s)\nabla Y_s + f'_z(s, X_s, Y_s, Z_s)\nabla Z_s]ds - \int_t^T \nabla Z_s dW_s.$$

The explicit representation of the LBSDE yields **[Ma, Zhang '02]**

$$Z_t = \nabla Y_t[\nabla X_t]^{-1}\sigma(t, X_t)$$
$$= \mathbb{E}[\Phi'(X_T)\nabla X_T \Gamma_T^t + \int^T f'x(s, X_s, Y_s, Z_s)\nabla X_s \Gamma_T^s ds | \mathcal{F}_t][\nabla X_t]^{-1}\sigma(t, X_t$$

Z-regularity

$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|Z_t - \bar{Z}_{t_i}|^2 dt$

Following from this representation, to Ito-decomposition of Z contains :

- an absolutely continuous part (in dt) $\rightarrow$ **easy to handle.**
- a martingale part M (in dW_t) :

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|M_t - \bar{M}_{t_i}|^2 dt \leq |\pi| \mathbb{E}(M_T^2 - M_0^2)!!$$

Possible extensions to **$\mathbb{L}_\infty$-functionals [Zhang '04]**, to **jumps [Bouchard, Elie '08]**, to **RBSDE [Bouchard, Chassagneux '06]**, to **BSDE with random terminal time [Bouchard, Menozzi '09]**

Other methods :Gobet and alii

- The case of irregular function $\Phi(X_T)$, with strict ellipticity
- Error expansion for smooth data and uniform grid **[G.,Labart '07]**
- Resolution by Picard's iteration, as limit of linear BSDE : [Bender, Denk '07] ; **[G.,Labart '09]** with adaptive control variates. Smaller errors propagation compared to the dynamic programming equation.

Computations of the conditional expectations

Our objective : to implement the dynamic programming equation = to compute the conditional expectations $\rightarrow$ the crucial step ! !

Different points of view :

- the conditional expectation is a projection operator : if $Y \in \mathbb{L}_2$, then

$$\mathbb{E}(Y|X) = \text{Arg} \min_{m \in \mathbb{L}_2(\mathbb{P}^X)} \mathbb{E}(Y - m(X))^2.$$

$\rightarrow$ this is a least-squares problem. What for ?

- To simulate the random variable $m(X)$? one only needs its law.

- To compute the regression function m ? finding a function of dimension = $\dim(X)$ $\rightarrow$ curse of dimensionality.

- How many regression function to compute ?

  Answer. For the DPE of BSDEs, N regression functions and $N \to \infty$.

$$\begin{cases} v^N(t_i,x) &= \frac{1}{\Delta t_i}\mathbb{E}(u^N(t_{i+1},X^N_{t_{i+1}})\Delta W^N_{t_i} = x), \\ u^N(t_i,x) &= \mathbb{E}(u^N(t_{i+1},X^N_{t_{i+1}}) + \Delta t_i f(t_i,x,u^N(t_{i+1},X^N_{t_{i+1}}), v^N(t_{i+1},x)|X^N_{t_i} = x \\ u^N(T,x) &= \Phi(x). \end{cases}$$

- In which points $X \in \mathbb{R}^d$?

  Answer. Potentially, many ...

All is a question of global efficiency = balance between accuracy and computational cost

Markovian setting

Based on $\mathbb{E}(g(X_{t_{i+1}})|X_{t_i}) = \int g(x)\mathbb{P}_{X_{t_{i+1}}|X_{t_i}}(dx) = m(X_{t_i})$.

If $m(.)$ are required at only few values of $X_{t_i} = x_1, \ldots, x_n$:

- one can simulate M independant paths of $X_{t_{i+1}}$ starting from $X_{t_i} = x_1, \ldots, x_n$ and average them out (usual Monte Carlo procedures).
- but if needed for many i, exponentially growing tree ! !

How to put constraints on the complexity ?

One possibility for one-dimensional BM (or Geometric BM) : replace the true dynamics by that of a Bernoulli random walk (**binomial tree**).

3.2 Representation of conditional expectation using Malliavin calculus

[Fournié, Lasry, Lebuchoux, Lions '01 ; Bouchard, Touzi '04 ; Bally, Caramellino, Zanette '05 . . .]

Theorem. [integration by parts formula] Suppose that for any smooth f, one has

$$\mathbb{E}(f^k(F)G) = \mathbb{E}(f(F)H_k(F,G))$$

for some r.v. $H_k(F,G)$, depending on F, G, on the multi-index k but not on f.

Then, one has

$$\mathbb{E}(G|F=x) = \frac{\mathbb{E}(\mathbf{1}_{F_1 \le x_1,\ldots,F_d \le x_d} H_{1,\ldots,1}(F,G))}{\mathbb{E}(\mathbf{1}_{F_1 \le x_1,\ldots,F_d \le x_d} H_{1,\ldots,1}(F,1))}.$$

Formal proof $(\mathbf{d}=\mathbf{1})$:

$\mathbb{E}(G|F=x)$ $\quad \mathbb{E}(G\delta_x(X))$ $\quad \mathbb{E}(G(\mathbf{1}_{F<x})')$ $\quad \mathbb{E}(\mathbf{1}_{F<x}H_1(F,G))$

- The H are obtained using Malliavin calculus, or a direct integration by parts when densities are known.
- Actually, we look for $H(F,G) = G\tilde{H}(F,G)$. Representation with factorization not so immediate to obtain (possible for SDE).
- In practice, large variance $\to$ need some extra localization procedures.
- For non trivial dynamics, the computational time needed to simulate H may be high.
- For BSDEs, available rates of convergence w.r.t. N and M **[Bouchard, Touzi '04]** using N independent set of simulated paths.

Statistical regression model : $Y = m(X) + \epsilon$, with $\mathbb{E}[\epsilon|X] = 0$.

X is called the (random) design.

Large literature on statistical tools to approximate $\mathbb{E}[Y|X]$.

References **[Hardle '92 ; Bosq, Lecoutre '87 ; Gyorfi, Kohler, Krzyzak, Walk '02]**

Problem : compute $m(\cdot)$ using M independent (?) samples $(Y_i, X_i)_{1\leq i\leq M}$.

Usually estimation errors in the literature are not sufficient for our purpose :

- the law X may not have a density w.r.t. Lebesgue measure.
- the support of the law of the X is never bounded ! !

Discussions of non parametric regression tools from theoretical/practical points of view

3.3.1. Kernel estimators

$$\mathbb{E}[Y|W=x] \approx \frac{\frac{1}{h^d}\sum_{i=1}^{M} K(\frac{x-X_i}{h})Y_i}{\frac{1}{h^d}\sum_{i=1}^{M} K(\frac{x-X_i}{h})} = m_{M,h}(x), \text{ where}$$

- the kernel function is defined on the compact support $[-1,1]$, bounded, even, non-negative, C_p^2 and $\int_{\mathbb{R}^d} K(u)du = 1$,
- $h > 0$ is the bandwith.

Non-integrated $\mathbb{L}_2$-error estimates available.

3.3.1. Projection on a set of functions

Set of functions : $(\phi_k)_{0\leq k\leq K}$.

$$\mathbb{E}(Y|X) = \arg\min_g \mathbb{E}(Y - g(X))^2 \approx \arg\min_{\sum_{k=1}^K \alpha_k\phi_k(\cdot)} (Y - \sum_{k=1}^K \alpha_k\phi_k(X))^2.$$

Computations of the optimal coefficients $(\alpha_k)_k$: it solves the normal equation

$$A\alpha = \mathbb{E}(Y\phi), \text{ where } A_{i,j} = \mathbb{E}(\phi_i(X)\phi_j(X)), \ [\mathbb{E}(Y\phi)]_i = \mathbb{E}(Y\phi_i(X)).$$

- For simplisity, one should have a system of orthonormal functions (w.r.t the law of X).

- If the system is not orthonormal, one should compute A and invert it.

  Its dimensions is expected to be very large : $K \to \infty$ to ensure convergent approximations.

  Presumably big instabilities (ill-conditioned matrix) to solve this least-squares problem **[Golub, Van Loan '96]**.

- In practice, A is computed using simulations, as well $\mathbb{E}[Y\phi]$.

  Equivalent to solve the **empirical least-squares problem** :

$$(\alpha_k^M)_k = \arg\min_{\alpha} \frac{1}{M} \sum_{m=1}^{M} (Y^m - \sum_{k=1}^{K} \alpha_k \phi_k(X^m))^2.$$

CLT At fixed K, if A is invertible, one has $\lim_{M\to\infty} \sqrt{M}(\alpha^M - \alpha) = \mathcal{N}(0, \ldots)$.

The case of polynomial functions

- Popular choice.
- Smooth approximation.
- Global approximation : within few polynomials, a smooth $m(.)$ can be very well approximated.
- But show convergence for non smooth functions (non-linear BSDEs may lead non-smooth functions).
- Do projections on polynomials converge to $m(.)$? $\oplus_{k\geq 0}(P)_k(X) = \mathbb{L}_2(X)$? If for some $a > 0$ one has $\mathbb{E}(e^{a|X|}) < \infty$, then polynomials are dense in $\mathbb{L}_2$-functions. **Proof.** Related to the moment problems. Is a r.v. characterized by its polynomial moment ? In particular, if X is log-normal, ortonomials of X are not dense in $\mathbb{L}_2$ (Feller counter-exemple) ! !

The case of local approximation

Piecewise constant approximations $\phi_k = \mathbf{1}_{\mathcal{C}_k}$, where the subsets $(\mathcal{C}_k)_k$ forms a tesselation of a part of $\mathbb{R}^d$: $\mathcal{C}_k \cap \mathcal{C}_l = \emptyset$ for $l \neq k$.

$$\arg \inf_{g=\sum_k \alpha_k \mathbf{1}_{\mathcal{C}_k}} \mathbb{E}(Y - g(X))^2 \text{ or } \arg \inf_{g=\sum_k \alpha_k \mathbf{1}_{\mathcal{C}_k}} \mathbb{E}^M(Y - g(X))^2 ?$$

The "matrix"

$A = (\mathbb{E}(\phi_i(X)\phi_j(X)))_{i,j}$ is diagonal : $A = \mathrm{Diag}(\mathbb{P}(X \in \mathcal{C}_i)_i)) \implies$

$$\alpha_k = \begin{cases} \frac{\mathbb{E}(Y \mathbf{1}_{X \in \mathcal{C}_k})}{\mathbb{P}(X \in \mathcal{C}_k)} = \mathbb{E}(Y | X \in \mathcal{C}_k) & \text{if } \mathbb{P}(X \in \mathcal{C}_k) > 0, \\ 0 & \text{if } \mathbb{P}(X \in \mathcal{C}_k) = 0, \end{cases}$$

$$\alpha_k^M = \begin{cases} \frac{1}{\#\{m; X^m \in \mathcal{C}_k\}} \sum_{m: X^m \in \mathcal{C}_k} Y^m & \text{if } \#\{m : X^m \in \mathcal{C}_k\} > 0, \end{cases}$$

Rate of approximations of a Lipschitz regression function $m(.)$

Size of the tesselation : $|\mathcal{C}| \leq \sup_l \sup_{(x,y)\in\mathcal{C}_l} |x-y|$.

Given a probability measure μ : $\mu = \mathbb{P}_X$ or $\mu = \frac{1}{M}\sum_{m=1}^{M} \delta_{X^m}(.)$.

$$\inf_{g=\sum_k \alpha_k \mathbf{1}_{\mathcal{C}_k}} \int_{\mathbb{R}^d} |g(x)-m(x)|^2 \mu(dx)$$

$$\leq \sum_k \int_{\mathcal{C}_k} |m(x_k)-m(x)|^2 \mu(dx) + \int_{[\cup_k \mathcal{C}_k]^c} m^2(x)\mu(dx)$$

$$\leq \sum_k |\mathcal{C}|^2 \mu(\mathcal{C}_k) + |m|^2_\infty \mu([\cup_k \mathcal{C}_k]^c) \leq |\mathcal{C}|^2 + |m|^2_\infty \mu([\cup_k \mathcal{C}_k]^c).$$

- We expect the tesselation size to be small.
- The complementary $\mu([\cup_k \mathcal{C}_k]^c)$ has to be small (tail estimates).

Efficient choice of tesselation ?

Given $x \in \mathbb{R}^d$, how to locate efficiently the $\mathcal{C}_k$ such that $x \in \mathcal{C}_k$?

- **Voronoi tesselations** associated to a sample $(X^K)_{1 \leq k \leq K}$ of the underlying r.v. $X : \mathcal{C}_k = \{z \in \mathbb{R}^d : |z - X^k| = min_l |z - X^l|\}$.

  Closed to quantization ideas.

  Theorically, there exists searching algorithms with a cost $\mathcal{O}(log(K))$.

- **Regular grid (hepercubes).**

  $k = (k_1, \ldots, k_d) \in \{0, \ldots, K_1 - 1\} \times \ldots \times \{0, \ldots, K_d - 1\}$ define

  $\mathcal{C}_k = [-x_{1,min} + \Delta x_1 k_1, -x_{1,min} + \Delta x_1 (k_1 + 1)[\times \cdots \times [-x_{d,min} + \Delta x_d k_d, -x_{d,min} + \Delta x_d (k_d + 1)$

  Tesselation size $= \mathcal{O}(\max_i \Delta x_i)$

  Quick search formula :

3.4 Model-free estimation of the regression error [GKKW02]

In the BSDEs framework, see **[Lemor, G., Warin '06]**.

Working assumptions :

- $Y = m(X) + \epsilon$ with $\mathbb{E}(\epsilon|X) = 0$.
- Data : sample of independant copies $(X_1, Y_1), \ldots, (X_n, Y_n)$.
- $\sigma^2 = \sup_x \mathrm{Var}(Y|X = x) < \infty$
- $F_n = \mathrm{Span}(f_1, \ldots, f_{K_n})$ a linear vector space of dimension K_n, which may depend on the data !

Notations : $|f|_n^2 = \frac{1}{n}\sum_{i=1}^n f^2(X_i)$. Write μ^n for the empirical measure associated to $(X_1, \ldots, X_n)$.

$$\hat{m}_n(.) = \arg\min_{f \in F_n} \frac{1}{n}\sum_{i=1}^n |f(X_i) - Y_i|^2.$$

Proof

W.l.o.g., we can assume that

- $(f_1, \ldots, f_{K_n})$ is orthonormal family in $\mathbb{L}_2(\mu^n)$: $\frac{1}{n}\sum_i f_k(X_i) f_l(X_i) = \delta_{k,l}$.

$\Longrightarrow$ The solution of $\arg\min_{f \in F_n} \frac{1}{n}\sum_{i=1}^n |f(X_i) - Y_i|^2$ is given by

$$\hat{\mathbf{m}}_{\mathbf{n}}(.) = \sum_{\mathbf{j}} \alpha_{\mathbf{j}} \mathbf{f}_{\mathbf{j}}(.) \text{ with } \alpha_{\mathbf{j}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i}} \mathbf{f}_{\mathbf{j}}(\mathbf{X}_{\mathbf{i}}) \mathbf{Y}_{\mathbf{i}}.$$

Lemma. Denote $\mathbb{E}^*(.) = \mathbb{E}(.|X_1, \ldots, X_n)$. Then $\mathbb{E}^*(\tilde{m}_n(.))$ is the least-squares solution of $\arg\min_{f \in F_n} \frac{1}{n}\sum_{i=1}^n |f(X_i) - m(X_i)|^2 = \arg\min_{f \in F_n} |f - m|_n^2$.

Proof.

- The above least-squares solution is given by $\sum_j \alpha_j^* f_j(.)$ with $\alpha_j^* = \frac{1}{n}\sum_i f_j(X_i) m(X_i)$.

Pythagore theorem : $|\tilde{m}_n - m|^2 = |\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n))|_n^2 + |\mathbb{E}^*(\tilde{m}_n) - m|_n^2$.

$$\begin{aligned}\text{Then, } \mathbb{E}^*|\tilde{m}_n - m|_n^2 &= \mathbb{E}^*|\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n)|_n^2 + |\mathbb{E}^*(\tilde{m}_n) - m|_n^2 \\ &= \mathbb{E}^*|\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n)|_n^2 + \min_{f\in F_n} |f - m|_n^2.\end{aligned}$$

Since $(f_j)_j$ is orthonormal in $\mathbb{L}_2(\mu_n)$, we have

$$|\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n)|_n^2 = \sum_j |\alpha_j - \mathbb{E}^*(\alpha_j)|^2.$$

Thus, using $\alpha_j - \mathbb{E}^*(\alpha_j) = \frac{1}{n}\sum_i f_j(X_i)(Y_i - m(X_i))$, we have

$$\begin{aligned}\mathbb{E}^*|\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n)|_n^2 &= \sum_j \frac{1}{n^2}\mathbb{E}^* \sum_{i,l} f_j(X_i)f_j(X_l)(Y_i - m(X_i))(Y_l - m(X_l)) \\ &= \sum_j \frac{1}{n^2} \sum_i f_j^2(X_i)\mathrm{Var}(Y_i|X_i)\end{aligned}$$

since **the $(\epsilon_i)_i$ conditionnaly on $(X_1, \ldots, X_n)$ are centered.**

$$\Longrightarrow \mathbb{E}^*|\tilde{m}_n - \mathbb{E}^*(\tilde{m}_n)|_n^2 \le \sigma^2 \sum \frac{1}{n^2} \sum f_j^2(X_i) = \sigma^2 \frac{K_n}{n}.$$

Uniform law of large numbers

$Z_{1:n} = (Z_1, \ldots, Z_n)$ a i.i.d. sample of size n.

For $\mathcal{G} \subset \{g : \mathbb{R}^d \mapsto [0, B]\}$, one needs to quantify

$$\mathbb{P}[\forall g \in \mathcal{G} : |\frac{1}{n}\sum_{i=1}^{n} g(Z_i) - \mathbb{E}g(Z)| > \epsilon]$$

as a function of ϵ and n ?

By Borel-Cantelli lemma, may lead to uniform laws of large numbers :

$$\sup_{g \in \mathcal{G}} |\frac{1}{n}\sum_{i=1}^{n} g(Z_i) - \mathbb{E}g(Z)| \to 0 \text{ a.s.}$$

ϵ-cover of $\mathcal{G}$

Definition. For a class of functions $\mathcal{G}$ and a given empirical measure μ^n associated to n points $Z_{1:n} = (Z_1, \dots, Z_n)$, we define a **$\epsilon$-cover of $\mathcal{G}$** w.r.t. $\mathbb{L}_1(\mu^n)$ by a **collection $(g_1, \dots, g_N)$ in $\mathcal{G}$** such that

$$\text{for any } g \in \mathcal{G}, \text{ there is a } j \in \{1, \dots, N\} \text{ s.t. } |g - g_j|_{\mathbb{L}_1(\mu^n)} < \epsilon.$$

Set $\mathcal{N}_1(\epsilon, \mathcal{G}, \mathbf{Z_{1:n}})$=the simplest size N of ϵ-cover of $\mathcal{G}$ w.r.t. $\mathbb{L}_1(\mu^n)$.

Theorem. For $\mathcal{G} \subset \{g : \mathbb{R}^d \mapsto [-B, B]\}$. For any n and any $\epsilon > 0$, one has

$$\mathbb{P}(\forall \mathbf{g} \in \mathcal{G} : |\frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{g}(\mathbf{Z_i}) - \mathbb{E}\mathbf{g}(\mathbf{Z})| > \epsilon) \leq \mathbf{8}\mathbb{E}(\mathcal{N}_\mathbf{1}(\epsilon/\mathbf{8}, \mathcal{G}, \mathbf{Z_{1:n}})) \exp(-\frac{\mathbf{n}\epsilon^\mathbf{2}}{\mathbf{512B^2}}).$$

Theorem. If $\mathcal{G} = \{-B \vee \sum_k \alpha_k \phi_k(.) \wedge B : (\alpha_1, \dots, \alpha_K) \in \mathbb{R}^K\}$, then

$$\mathcal{N}_\mathbf{1}(\mathbf{e}, \mathcal{G}, \mathbf{Z_{1:n}}) \leq \mathbf{3}\Big(\frac{\mathbf{4eB}}{\ } \log(\frac{\mathbf{4eB}}{\ })\Big)^{\mathbf{K+1}}$$

3.5 Applications to numerical solution of BSDEs using empirical simulations [LGW06]

Regular time grid with time step $h = \frac{T}{N}$+Lipschitz f, Φ, b and σ.

Towards an approximation of the regression operators

Truncation of the tails using a threshold $R = (R_0, \ldots, R_d)$:

$$[\Delta W_{l,k}]_w = (-R_0\sqrt{h}) \vee \Delta W_{l,k} \wedge (R_0\sqrt{h}),$$

$$f^R(t,x,y,z) = f(t, -R_1 \vee x_1 \wedge R_1, \ldots, -R_d \vee x_d \wedge R_d, y, z),$$

$$\Phi^R(x) = \Phi(-R_1 \vee x_1 \wedge R_1, \ldots, -R_d \vee x_d \wedge R_d).$$

$\rightarrow$ Localized BSDEs

Define $Y_T^{N,R}(X_{t_k}^N) = \Phi^R(X_{t_k}^N)$ and

Proposition. For some **Lipschitz** functions $y_k^{N,R}(\bullet)$ and $z_k^{N,R}(\bullet)$, one has :

$$\begin{cases} Z_{l,t_k}^{N,R} & = \frac{1}{h}\mathbb{E}(Y_{t_{k+1}}^{N,R}[\Delta W_{l,k}]_\omega|\mathcal{F}_{t_k}) = z_{l,k}^{N,R}(X_{t_k}^N). \\ Y_{t_k}^{N,R} & = \mathbb{E}(Y_{t_{k+1}}^{N,R} + hf^R(t_k, X_{t_k}^N, Y_{t_{k+1}}^{N,R}, Z_{t_k}^{N,R})|\mathcal{F}_{t_k}) = y_k^{N,R}(X_{t_k}^N). \end{cases}$$

a) The Lipschitz constants of $y_k^{N,R}(\bullet)$ and $N^{-1/2}z_k^{N,R}(\bullet)$ are uniform in N and R.

b) **Bounded functions :** $sup_N(\| y_k^{N,R}(\bullet) \|_\infty + N^{-1/2} \| z_k^{N,R}(\bullet) \|_\infty) = C_\star < \infty$

Proposition. (Convergence as $|R| \uparrow \infty$) For h small enough, one has

$$\max_{0\leq k\leq N} \mathbb{E}|Y_{t_k}^{N,R} - Y_{t_k}^N|^2 + h\mathbb{E}\sum_{k=0}^{N-1}|Z_{t_k}^{N,R} - Z_{t_k}^N|^2$$

$$\leq C\mathbb{E}|\Phi(X_{t_n}^N) - \Phi^R(X_{t_N}^N)|^2 + C\frac{1+R^2}{h}\sum_{k=0}^{N-1}\mathbb{E}(|\Delta W_k|^2\mathbf{1}_{|\Delta W_k|\geq R_0\sqrt{h}})$$

$$+ Ch\mathbb{E}\sum_{k=0}^{N-1}|f(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N) - f^R(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N)|^2.$$

Approximation of $y_k^{N,R}(\bullet)$ and $z_k^{N,R}(\bullet)$

Projection on a finite dimensional space :

$$\mathbf{y}_{\mathbf{k}}^{\mathbf{N},\mathbf{R}}(\bullet) \approx \alpha_{\mathbf{0},\mathbf{k}} . \mathbf{p}_{\mathbf{0},\mathbf{k}}(\bullet), \qquad \mathbf{z}_{\mathbf{l},\mathbf{k}}^{\mathbf{N},\mathbf{R}}(\bullet) \approx \alpha_{\mathbf{l},\mathbf{k}} . \mathbf{p}_{\mathbf{l},\mathbf{k}}(\bullet).$$

(for instance, hypercubes as presented before).

Coefficients will be computed by extra M independent simulations of $(X_{t_k}^N)_k$ and $(\Delta W_k)_k \to \{(X_{t_k}^{N,m})_k\}_m$ and $\{(\Delta W_k^m)_k\}_m$ (**only one set of simulated paths**).

In addition, we impose **boundedness properties** :

$$\mathbf{y}_{\mathbf{k}}^{\mathbf{N},\mathbf{R},\mathbf{M}}(\bullet) = [\alpha_{\mathbf{0},\mathbf{k}}^{\mathbf{M}} . \mathbf{p}_{\mathbf{0},\mathbf{k}}(\bullet)]_{\mathbf{y}}, \qquad \mathbf{z}_{\mathbf{l},\mathbf{k}}^{\mathbf{N},\mathbf{R},\mathbf{M}}(\bullet) \approx [\alpha_{\mathbf{l},\mathbf{k}}^{\mathbf{M}} . \mathbf{p}_{\mathbf{l},\mathbf{k}}(\bullet)]_{\mathbf{z}},$$

where $[\psi]_y = -C_y \vee \psi \wedge C_y$, $[\psi]_z = -C_z N^{1/2} \vee \psi \wedge C_z N^{1/2}$

The final algorithm

$\rightarrow$ Initialization : for $k = N$ take $y_N^{N,R}(\cdot) = \Phi^R(\cdot)$.

$\rightarrow$ Iteration : for $k = N-1, \cdots, 0$, solve the q least-squares problems :

$$\alpha_{l,k}^M = \arg\inf_{\alpha} \frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m})\frac{[\Delta W_{l,k}^m]_\omega}{h} - \alpha \cdot p_{l,k}(X_{t_k}^{N,m})|^2$$

Then compute $\alpha_{0,k}^M$ as the minimizer of

$$\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) + hf^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}), [\alpha_{l,k}^M \cdot p_{l,k}(X_{t_k}^{N,m})]_z) - \alpha \cdot p_{0,k}(X_{t_k}^{N,m}$$

Then define $y_k^{N,R,M}(\bullet) = [\alpha_{0,k}^M \cdot p_{0,k}(\bullet)]_y$, $z_{l,k}^{N,R,M}(\bullet) = [\alpha_{l,k}^M \cdot p_{l,k}(\bullet)]_z$.

Error analysis

Robust error bounds

Theorem. Under Lipschitz conditions (only !), one has

$$\max_{0\le k\le N} \mathbb{E}|Y^{N,R}_{t_k} - y^{N,R,M}_k(S^N_{t_k})|^2 + h\sum_{k=0}^{N-1}\mathbb{E}|Z^{N,R}_{t_k} - z^{N,R,M}_k(S^N_{t_k})|^2$$

$$\le C\,\frac{C_\star^2\log(M)}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}\mathbb{E}(K^M_{l,k}) + Ch$$

$$+ C\sum_{k=0}^{N-1}\{\inf_\alpha \mathbb{E}|y^{N,R}_k(S^N_{t_k}) - \alpha\cdot p_{0,k}(S^N_{t_k})|^2 + \sum_{l=1}^{q}\{\inf_\alpha \mathbb{E}|\sqrt{h}z^{N,R}_{l,k}(S^N_{t_k}) - \alpha\cdot p_{l,k}(S^N_{t_k})|^2$$

$$+ C\,\frac{C_\star^2}{h}\sum_{k=0}^{N-1}\{\mathbb{E}[K^M_{0,k}\exp(-\frac{Mh^3}{72C_\star^2K^M_{0,k}})\exp(CK_{0,k+1}\log\frac{CC_\star(K^M_{0,k})^{\frac{1}{2}}}{h^{\frac{3}{2}}})]$$

$$+ h\mathbb{E}[K^M_{l,k}\exp(-\frac{Mh^2}{72C_\star^2R_0^2K^M_{l,k}})\exp(CK_{0,k+1}\log\frac{CC_\star R_0(K^M_{l,k})^{\frac{1}{2}}}{h})]$$

$$+ \exp(CK_{0,k}\log\frac{CC_\star}{h^{\frac{3}{2}}})\exp(-\frac{Mh^3}{72C_\star^2})\}.$$

Convergence of the parameters in the cases of HC functions

For a global squared error of order $\epsilon = \frac{1}{N}$, choose :

1. Edge of the hypercube : $\delta \sim \frac{C}{N}$.
2. Number of simulations : $M \sim N^{3+2d}$.

Available for a large class of models on X, which depend essentially on $\mathbb{L}_2$ bounds on the solution (no ellipticity condition, with or without jump...).

Complexity/accuracy

Global complexity : $\mathcal{C} \sim \epsilon^{-\frac{1}{4+2d}}$.

Techniques of local duplicating of paths : $\mathcal{C} \sim \epsilon^{-\frac{1}{4+d}}$.

3.6 Numerical results (mainly due to J.P. Lemor)

Ex.1 : bid-ask spread for interest rates

- Black-Scholes model and $\Phi(\mathbf{S}) = (S_T - K_1)_+ - 2(S_T - K_2)_+$.
- $f(t,x,y,z) = -\{yr + z\theta - (y - \frac{z}{\sigma})_-(R-r)\}$, $\theta = \frac{\mu - r}{\sigma}$.
- Parameters :

μ	σ	r	R	T	S_0	K_1	K_2
0.05	0.2	0.01	0.06	0.25	100	95	105

M	$N=5,\ \delta=5$ $D=[60,140]$	$N=20,\ \delta=1$ $D=[60,200]$	$N=50,\ \delta=0.5$ $D=[60,200]$
128	3.05(**0.27**)	3.71(**0.95**)	3.69(**4.15**)
512	2.93(0.11)	3.14(0.16)	3.48(0.54)
2048	2.92(0.05)	3.00(0.03)	3.08(0.12)
8192	2.91(0.03)	2.96(0.02)	2.99(0.02)
32768	2.90(0.01)	**2.95**(0.01)	[illegible]

Global polynomials (GP)

Polynomials of d variables with a maximal degree.

M	$N=5$ $d_y=1,\ d_z=0$	$N=20$ $d_y=2,\ d_z=1$	$N=50$ $d_y=4,\ d_z=2$	$N=50$ $d_y=9,\ d_z=$
128	2.87(0.39)	3.01(0.24)	3.02(0.22)	3.49(**1.57**)
512	2.82(0.20)	2.94(0.12)	2.97(0.09)	3.02(0.1)
2048	2.78(0.07)	2.92(0.07)	2.92(0.0.04)	2.97(0.03)
8192	2.78(0.05)	2.92(0.04)	2.92(0.02)	2.96(0.01)
32768	2.79(0.03)	2.91(0.02)	2.91(0.01)	2.95(0.01)

Table: Results for the calls combination using **GP.**

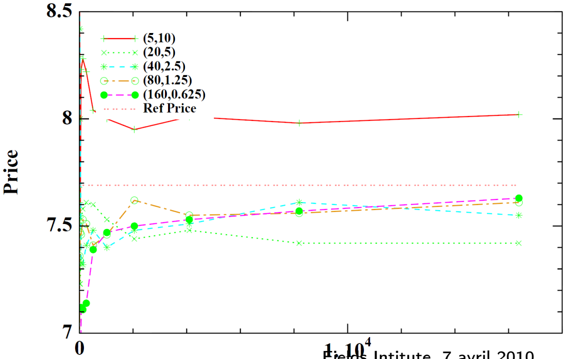
Ex.2 : locally-risk minimizing strategies (FS decomposition)

Heston stochastic volatility models **[Heath,Platen,Schweizer '02]** :

$$\frac{dS_t}{S_t} = \gamma Y_t^2 dt + Y_t dW_t, \qquad dY_t = (\frac{c_0}{Y_t} - c_1 Y_t)dt + c_2 dB_t.$$

Functions **HC**,

parameters (N, δ).



American options via RBDSDEs : several approaches

1. Talking the **max** with obstacle $\rightarrow$ Bermuda options (**lower approximation**)

$$Y^n_{t_k} = \max(\Phi(t_k, S^N_{t_k}), \mathbb{E}(Y^N_{t_{k+1}} | \mathcal{F}_{t_k}) + hf(t_k, S^N_{t_k}, Y^N_{t_k}, Z^N_{t_k})),$$

$$Z^N_{l,t_k} = \frac{1}{h}\mathbb{E}(Y^N_{t_{k+1}} \Delta W_{l,k} | \mathcal{F}_{t_k}).$$

2. **Penalization.** Obtained as the limit of standard BSDEs with driver $f(s, S_s, Y_s, Z_s) + \lambda(Y_s - \Phi(s, S_s))_-$ with $\lambda \uparrow +\infty$.

Lower approximation.

3. **Regularization** of the increasing process : when

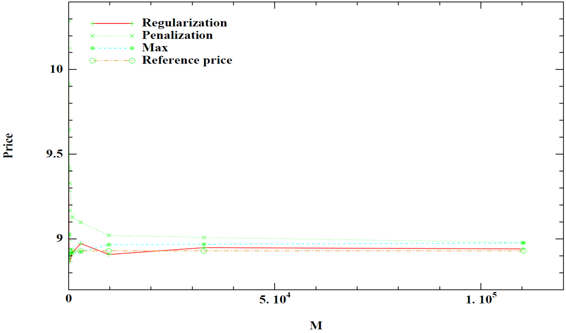
$$d\Phi(t, S_t) = U_t dt + V_t dW_t + dA^+_t,$$

Ex.3 : American options on tree assets

- Payoff $g(x) = (K - (\prod_{i=1}^{3} x_i)^{\frac{1}{3}})^+$.
- Black-Scholes parameters :

T	r	σ	K	S_0^i	d
1	0.05	0.4Id	100	100	1

- Reference price **8.93** (PDE method).



Functions **HC(1,0)** with local polynomials of degree 1 for Y and 0 for Z.

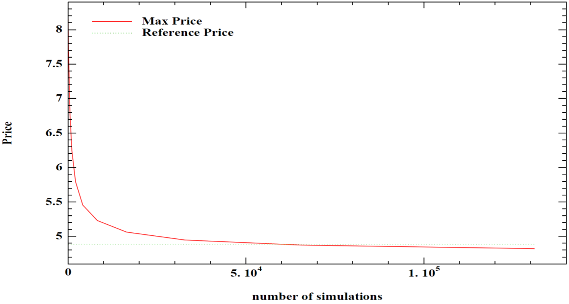
Regularisation : $N = 32$, $\delta = 9$, $\lambda = 2$.

Max : $N = 44$, $\delta = 7$.

Penalization : $N = 60$, $\delta = 2$, $\lambda = 2$.

Ex.4 : American options on ten assets

- $d = 10 = 2p$. Multidimensional Black-Scholes model : $\frac{dS_t^l}{S_t^l} = (r - \mu_l)dt + \sigma_l dW_t^l$.
- Payoff : $\max(x_1 \cdots x_p - x_{p+1} \cdots x_{2p}, 0)$.
- $r = 0$, dividend rate $\mu_1 = -0.05$, $\mu_l = 0$ for $l \geq 2$. $\sigma_l = \frac{0.2}{\sqrt{d}}$. $T = 0.5$. $S_0^i = 40^{\frac{2}{d}},\ 1 \leq i \leq p$. $S_0^i = 36^{\frac{2}{d}},\ p+1 \leq i \leq 2p$.
- Reference price **4.896**, obtained with a PDE method **[Villeneuve, Zanette 2002]**.
- Price with quantization algorithm : 4.9945 **[Bally-Pages-Printemps 2005]**.



Functions **HC(1,0)**.

Max : $N = 60$, $\delta = 0.6$.

Computational time :

15 seconds.

References