Backward SDEs with Financial Applications Part II

Emmanuel.Gobet@imag.fr

Laboratoire Jean Kuntzmann

Grenoble University (Ensimag)

Agenda

- Thursday: applications in finance + numerical methods
- Friday an saturday: numerical methods

1 Applications in finance

[Ref: El Karoui, Peng, Quenez '97 ; El Karoui, Quenez '97 ; Peng '03]

1.1 Pricing of European style contingent claims

Standard filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)0 \leq t \leq T, \mathbb{P})$, supporting a standard BM $W \in \mathbb{R}^q$.

Usual assumptions:

1. d risky assets: $\mathbf{dS^i_t = S^i_t(b^i_t + \sum_{j=1}^{q} \sigma^{i,j}_t dW^j_t)}$, $1 \leq i \leq d$.

   The appreciation rates $\mathbf{b^i}$ and volatilities $\sigma^{\mathbf{i,j}}$ are predictable and bounded.

2. A non risky asset (money market instrument): $\mathbf{dS^0_t = S^0_t r_t dt}$, where $\mathbf{r_t}$ is the short rate (predictable and bounded).

3. Existence of risk premium $\theta_{\mathbf{t}}$: predictable and bounded process such that $\mathbf{b_t - r_t 1} = \sigma_{\mathbf{t}}\theta_{\mathbf{t}}$ ($\mathbf{1}$ is the vector with all components equal to 1).

1.1.1 Self-financing strategy

$\phi_{\mathbf{t}}$: the row vector of amounts invested in each risky asset.

Here, **we do not assume any constraints on the strategy.**

The wealth process Y_t satisfies the self-financing condition:

$$\begin{aligned} dY_t &= \sum_{i=1}^{d} \phi_t^i \frac{dS_t^i}{S_t^i} + (Y_t - \sum_{i=1}^{d} \phi^i(t)) r_t dt \\ &= \phi_t(\sigma_t dW_t + b_t dt) + (Y_t - \phi_t \mathbf{1}) r_t dt \\ &= r_t Y_t dt + \phi_t \sigma_t \theta_t dt + \phi_t \sigma_t dW_t. \end{aligned}$$

If we set $\mathbf{Z_t} = \phi_{\mathbf{t}} \sigma_{\mathbf{t}}$, the self-financing condition writes

$$\mathbf{-dY_t = -r_t Y_t dt - Z_t \theta_t dt - Z_t dW_t.}$$

Up to the specification of the terminal value of Y_T, (Y, Z) solves a **Linear BSDE** (LBSDE), with a driver defined by $\mathbf{f(t, \omega, y, z) = -r_t y - z \theta_t}$.

The driver $f(t, \omega, y, z) = -r_t y - z\theta_t$ is globally Lipschitz in (y, z) (recall that r and θ are bounded).

⚠ Note that to safely come back to the hedging strategy, one has to invert the relation $\phi_t \mapsto Z_t = \phi_t \sigma_t$

$\rightsquigarrow$ **usually, the volatility matrix σ has to be invertible** $\leftrightarrow$ **complete market.**

1.1.2 Complete market without portfolio constraints

Replication of an option

Assume additionnally that

1. the volatility matrix σ has a full rank ($\mathbf{d} = \mathbf{q}$) and its inverse is bounded.

Consider a option maturing at T and payoff $\mathbf{\Phi(S_t : 0 \leq t \leq T)}$ (a path-dependent functional of S).

? Replication of the option? link with the risk-neutral valuation rule?

Answer: YES

Theorem. If $\Phi(S_t : 0 \leq t \leq T) \in \mathbb{L}_2(\mathbb{P})$, then there is a solution $(Y, Z) \in \mathbb{H}_2$ to the LBSDE and thus to the hedging problem.

In addition, the Y-component has a explicit representation has a conditional expectation.

Proof.

- For existence and uniqueness, apply standard BSDE results (see Jin Ma's minicourse).
- The hedging strategy is given by $\phi_t = \mathbf{Z_t}\sigma_{\mathbf{t}}^{-1}$.
- Finally, all LBSDE have an explicit representation (see [EPQ97]): the solution to $-\mathbf{dY_t} = [\varphi_t + \mathbf{Y_t}\beta_t + \mathbf{Z_t}\gamma_t]\mathbf{dt} - \mathbf{Z_t dW_t}$ and $Y_T = \xi \in L_2$ (with bounded (β, γ), $\varphi \in \mathbb{H}_2$ and $\xi \in \mathbb{L}_2$) is given by $\mathbf{Y_t} = \mathbb{E}[\xi\mathbf{\Gamma_T^t} + \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{\Gamma_t^s}\varphi_s\mathbf{ds}|\mathcal{F}_t]$ where $\Gamma_t^s = \exp(\int_t^s (\beta_r - \frac{1}{2}|\gamma_r|^2)dr + \int_t^s \gamma_r^* dW_r)$.
  $=$ **Rare situation where explicit solutions are known.**

In our setting of replicating an option, we have $\varphi_s = 0$, $\beta_t = -r_t$, $\gamma_t = -\theta_t$:

$$Y_t = \mathbb{E}_{\mathbb{P}}\Big[\exp(\int_t^T (-r_s - \frac{1}{2}|\theta_s|^2)ds - \int_t^T \theta_s^* dW_s)\xi|\mathcal{F}_t\Big]$$
$$= \mathbb{E}_{\mathbb{Q}}\Big[\exp(\int_t^T -r_s ds)\xi|\mathcal{F}_t\Big]$$

where $\mathbb{Q}|_{\mathcal{F}_t} = \exp(-\frac{1}{2}\int_0^t |\theta_s|^2)ds - \int_0^t \theta_s^* dW_s)\mathbb{P}|_{\mathcal{F}_t}$ defines the usual (unique) risk-neutral measure.

⚠ Solving this BSDE is done under the historical measure (with non risk-neutral simulations) and estimates under $\mathbb{P}$!

1.1.3 Complete market with portfolio constraints

Bid-ask spread for interest rates [Bergman '95, Korn '95, Cvitanic Karatzas '93]

The investor borrows money at interest rate R_t and lends at rate $\mathbf{r_t} < \mathbf{R_t}$.

$\rightsquigarrow$ Modification of the self-financing strategy:

$$dY_t = \sum_{i=1}^{d} \phi_t^i \frac{dS_t^i}{S_t^i} + (Y_t - \sum_{i=1}^{d} \phi^i(t))_+ r_t dt - (Y_t - \sum_{i=1}^{d} \phi^i(t))_- R_t dt$$

$$= \phi_t(\sigma_t dW_t + b_t dt) + (Y_t - \phi_t \mathbf{1}) r_t dt - (R_t - r_t)(Y_t - \phi_t \mathbf{1})_- dt$$

$$= \mathbf{r_t} Y_t dt + \phi_t \sigma_t \theta_{\mathbf{t}}^{\mathbf{r}} dt + \phi_t \sigma_t dW_t \quad \underbrace{-(R_t - r_t)(Y_t - \phi_t \mathbf{1})_-}_{\text{additional cost when borrowing}} \quad dt$$

where $\mathbf{b_t} - \mathbf{r_t 1} = \sigma_{\mathbf{t}} \theta_{\mathbf{t}}^{\mathbf{r}}$.

Similarly, with $\mathbf{b_t} - \mathbf{R_t 1} = \sigma_{\mathbf{t}} \theta_{\mathbf{t}}^{\mathbf{R}}$, we have

$$dY_t = \mathbf{R_t} Y_t dt + \phi_t \sigma_t \theta_{\mathbf{t}}^{\mathbf{R}} dt + \phi_t \sigma_t dW_t \quad \underbrace{-(R_t - r_t)(Y_t - \phi_t \mathbf{1})_+}_{\text{smaller portfolio appreciation when lending}} \quad dt.$$

Set $\mathbf{Z_t} = \phi_\mathbf{t}\sigma_\mathbf{t}$. Then, (Y, Z) solves a **non-linear** BSDE with the globally Lipschitz driver

$$\begin{aligned}\mathbf{f^{r,R}(t, y, z)} &= -r_t y - z\theta_t^r + (R_t - r_t)(y - z\sigma_t^{-1}\mathbf{1})_- \\ &= -R_t y - z\theta_t^R + (R_t - r_t)(y - z\sigma_t^{-1}\mathbf{1})_+.\end{aligned}$$

We focus on the dependence on (r, R) by denoting $(\mathbf{Y^{r,R}}, \mathbf{Z^{r,R}})$ the solution to the BSDE with a given terminal condition and driver $\mathbf{f^{r,R}}$.

Comparison of prices with/without different interest rates?

Lower bounds. The price with different interest rates is still larger than the price with fixed interest rates:

$$\mathbf{Y_t^{r,R}} \geq \max(\mathbf{Y_t^{r,r}}, \mathbf{Y_t^{R,R}})$$

for any $t \in [0, T]$.

Proof. Apply the comparison theorem within its strong version:

$$f^{r,R}(t, y, z) \geq \max(-r_t y - z\theta_t^r, -R_t y - z\theta_t^R) = \max(f^{r,r}(t, y, z), f^{R,R}(t, y, z)).$$

Upper bounds and equalities: examples in the Black-Scholes setting.

- **Call option:** $\Phi(S) = (S_T - K)_+$.
  From the Black-Scholes formula with a **single interest rate**, one knows that the amount in cash is always negative (money borrowing) $\rightsquigarrow$

$$\mathbf{f^{r,R}(t, Y_t^{R,R}, Z_t^{R,R})} = -R_t Y_t^{R,R} - Z_t^{R,R}\theta_t^R + (R_t - r_t)\underbrace{(Y_t^{R,R} - Z_t^{R,R}\sigma_t^{-1}\mathbf{1})_+}_{=0}$$

$$= \mathbf{f^{R,R}(t, Y_t^{R,R}, Z_t^{R,R})}.$$

Hence, $(Y^{R,R}, Z^{R,R})$ also solves the BSDE with the driver $f^{r,R}$. By uniqueness:

$$(\mathbf{Y^{r,R}}, \mathbf{Z^{r,R}}) = (\mathbf{Y^{R,R}}, \mathbf{Z^{R,R}}).$$

The price is obtained using the higher interest rate.

- **Put option:** $\Phi(S) = (K - S_T)_+$.
  Similarly, with a **single interest rate**, one always lends money $\rightsquigarrow$

$$(\mathbf{Y^{r,R}}, \mathbf{Z^{r,R}}) = (\mathbf{Y^{r,r}}, \mathbf{Z^{r,r}}).$$

  The price is obtained with the lower interest rate.

- **Call Spread:** $\Phi(S) = (S_T - K_1)_+ - 2(S_T - K_2)_+$ $(K_1 < K_2)$.
  With probability 1, we have

$$\mathbf{Y_t^{r,R}} > \max(\mathbf{Y_t^{r,r}}, \mathbf{Y_t^{R,R}}) \quad \forall \mathbf{t} < \mathbf{T}.$$

  Proof by contradiction. Assume the equality on a set $A \in \mathcal{F}_t$. The comparison theorem implies the equality of drivers along $(Y_s^{r,r}, Z_s^{r,r})_{t\le s\le T}$ and $(Y_s^{R,R}, Z_s^{R,R})_{t\le s\le T}$ almost surely on $A \rightsquigarrow \mathbb{P}(\mathbf{A}) = \mathbf{0}$.

- **General payoff with deterministic coefficients** $(r_t)_t, (R_t)_t, (\sigma_t)_t, (b_t)_t$ **:** sufficient conditions in [EPQ97]. If

$$\mathbf{D_t \Phi(S) \sigma_t^{-1} 1 \ge \Phi(S)} \quad \mathbf{dt \otimes d\mathbb{P}} \ \mathbf{- a.e.},$$

  then $(\mathbf{Y^{r,R}}, \mathbf{Z^{r,R}}) = (\mathbf{Y^{R,R}}, \mathbf{Z^{R,R}})$.

Short sales constraints [Jouiny, Kallal '95...]

Difference of returns b_t^l and b_t^s when long and short positions in the risky assets.

Aim at modeling the existence of reposit rate for instance.

Similar story as before.

Leads to

- two risk premias $\theta_{\mathbf{t}}^{\mathbf{l}}$ and $\theta_{\mathbf{t}}^{\mathbf{s}}$.
- a BSDE with non-linear driver $\mathbf{f}(\mathbf{t}, \mathbf{y}, \mathbf{z}) = -\mathbf{r_t y} - \mathbf{z}\theta_{\mathbf{t}}^{\mathbf{l}} + [\mathbf{z}\sigma_{\mathbf{t}}^{-\mathbf{1}}]^{-}\sigma_{\mathbf{t}}(\theta_{\mathbf{t}}^{\mathbf{l}} - \theta_{\mathbf{t}}^{\mathbf{s}})$.

1.2 Incomplete markets

Suppose that $d < q$: number of tradable assets d smaller than the number of sources of risk q.

Examples:

- **trading restriction** on the assets.
- **stochastic volatilities model** like Heston model:

$$\begin{aligned} dS_t &= S_t(r_t dt + \sqrt{V_t} dW_t), \\ dV_t &= \kappa(\theta_t - V_t)dt + \xi_t \sqrt{V_t} dB_t, \\ d\langle W, B\rangle_t &= \rho_t dt. \end{aligned}$$

Here $d = 1$ (one can not trade the volatility) and $q = 2$.

Market incompleteness

Denote the associated amount ϕ_t^1 in the traded assets and the associated volatility $\sigma_t^1 \in \mathbb{R}^d \otimes \mathbb{R}^q$ w.r.t. the q-dimensional BM W.

The self-financing equation writes: $\mathbf{dY_t = r_t Y_t dt} + \phi_\mathbf{t}^\mathbf{1} \sigma_\mathbf{t}^\mathbf{1} \theta_\mathbf{t} \mathbf{dt} + \phi_\mathbf{t}^\mathbf{1} \sigma_\mathbf{t}^\mathbf{1} \mathbf{dW_t}$.

⚠ In general, there does not exist a strategy ϕ_t^1 such that $Y_T = \Phi(\mathbf{S})$.

Possible approaches:

1. **mean-variance hedging**
2. **super-replication**
3. ...
4. **local-risk minimization:** mean self-financing strategy + orthogonality of the cost process to the tradable martingale part

   $\rightsquigarrow$ Find a martingale M orthogonal to $(\int_0^t \sigma_s^1 dW_s)_t$ such that

   $$\mathbf{Y_T + M_T = \Phi(S)}\ \text{(}\textbf{[Föllmer-Schweizer decomposition '90]}\text{)}.$$

A BSDE-solution to the FS decomposition

Assumption: $\mathrm{rank}(\sigma_t^1) = d$ (non redundant tradable assets).

The FS strategy is obtained by solving a linear BSDE

$$\mathbf{dY_t = r_t Y_t dt + Z_t \theta_t^1 dt + Z_t dW_t}, \quad \mathbf{Y_T = \Phi(S)},$$

where

- $\sigma_t = \begin{pmatrix} \sigma_t^1 \\ \hline \sigma_t^2 \end{pmatrix} \in \mathbb{R}^q \otimes \mathbb{R}^q$ has a full rank q (we complete the market by *fictitious assets* with volatilities σ_t^2).
- $\theta_\mathbf{t}^\mathbf{1} = \mathrm{Proj}^{\perp}_{\mathrm{Range}([\sigma_\mathbf{t}^\mathbf{1}]^*)}(\theta_\mathbf{t})$ is the minimal risk premium.

(the solution of this LBSDE is the risk-neutral evaluation under the minimal martingale measure).

Proof by verification. (Y, Z) solves $dY_t = r_t Y_t dt + Z_t \theta^1_t dt + Z_t dW_t$ where $\boldsymbol{\theta^1_t = [\sigma^1_t]^* [\sigma^1_t \sigma^{1,*}_t]^{-1} \sigma^1_t \theta_t}$.

Define $[\mathbf{Z^1_t}]^* := \mathrm{Proj}^{\perp}_{\mathrm{Range}([\sigma^1_t]^*)}(\mathbf{Z^*_t}) = [\sigma^1_t]^* [\phi^1_t]^*$ and $\mathbf{Z^2_t := Z_t - Z^1_t}$.

Since $\mathbb{R}^q = \mathrm{Range}([\sigma^1_t]^*) \oplus \mathrm{Ker}(\sigma^1_t)$, one has $[Z^2_t]^* \in \mathrm{Ker}(\sigma^1_t)$: $\boldsymbol{\sigma^1_t [Z^2_t]^* = 0}$.

It follows

- $Z_t \theta^1_t = Z^1_t \theta^1_t + Z^2_t \theta^1_t = \phi^1_t \sigma^1_t \theta^1_t + \underbrace{Z^2_t [\sigma^1_t]^* [\sigma^1_t \sigma^{1,*}_t]^{-1} \sigma^1_t \theta_t}_{=0} = \phi^1_t \sigma^1_t \theta^1_t,$
- $Z_t dW_t = \phi^1_t \sigma^1_t dW_t + \underbrace{Z^2_t dW_t}_{=: dM_t}.$

Thus, $\mathbf{dY_t = r_t Y_t dt + \phi^1_t \sigma^1_t \theta^1_t dt + \phi^1_t \sigma^1_t dW_t + dM_t}$.

In addition, $\boldsymbol{< \int_0^\cdot \sigma^1_s dW_s, M >_t = 0 = \int_0^t \sigma^1_s [Z^2_s]^* ds}$

$\Longrightarrow$ M **is strongly orthogonal to** $(\int_0^t \sigma^1_t dW_t)_t$.

Uniqueness is proved similarly.

Other connections between pricing and BSDEs

- Superhedging via increasing sequence of non linear BSDEs (via penalization on the non tradable risks) **[Cvitanic, Karatzas '93; El Karoui, Quenez '95; El Karoui, Peng, Quenez '97]**
- Non linear pricing theory **[El Karoui, Quenez '97]**
- Large investor (fully coupled FBSDE) **[Cvitanic, Ma '96...]**.
- Recursive utility: driver quadratic in z **[Duffie, Epstein '92 ...]**.
- Exponential hedging and quadratic BSDE **[El Karoui, Rouge '01; Sekine '06 ...]** : $V(x) = \sup_{\phi\in\mathcal{A}} E(U(X_T^{x,\phi} - F))$ with U exponential utility.
- g-expectations and dynamically consistent evaluations/expectations **[Peng '03 ...]**
- American options **[El Karoui, Kapoudjian, Pardoux, Peng, Quenez '97]**

1.3 Dynamically consistent evaluation

An operator $\mathcal{E}_{s,t} : \mathbb{L}_2(\mathcal{F}_t) \mapsto \mathbb{L}_2(\mathcal{F}_s)$ is a dynamically consistent non linear evaluation if it satisfies:

A1) **Monotonicity**: $X \geq Y \Longrightarrow \mathcal{E}_{s,t}(X) \geq \mathcal{E}_{s,t}(Y)$.

A2) **Constant-preserving**: $\mathcal{E}_{t,t}(X) = X$ for $X \in \mathbb{L}_2(\mathcal{F}_t)$.

A3) **Time-consistency**: $\mathcal{E}_{r,s}(\mathcal{E}_{s,t}(X)) = \mathcal{E}_{r,t}(X)$ for all $r \leq s \leq t$.

A4) **0-1 law**: $\forall A \in \mathcal{F}_s$ and $X \in \mathbb{L}^2(\mathcal{F}_t)$ with $s \leq t$, one has $\mathbf{1}_A \mathcal{E}_{s,t}(X) = \mathbf{1}_A \mathcal{E}_{s,t}(\mathbf{1}_A X)$.

Consider a Lipschitz driver g and for $X \in \mathbb{L}^2(\mathcal{F}_t)$, denote by $(Y^g_{s,t}(X))_{s \leq t}$ the solution to

$$\mathbf{Y_s = X + \int_s^t g(r, Y_r, Z_r)dr - \int_s^t Z_r dW_r.}$$

Then $Y^g_{s,t}(X) = \mathcal{E}_{s,t}(X)$ **defines a dynamically consistent non linear evaluation.**

Proof. Follows from standard comparison and flow properties of BSDEs.

Converse property for dominated non linear evaluation

Consider a Brownian filtration and a dynamically consistent non linear evaluation operator $\mathcal{E}_{s,t}(.)$.

Define $g_\mu(y,z) = \mu|y| + \mu|z|$.

In addition, assume that for some $(k_t)_t$ and $\mu > 0$, one has

- $Y_{s,t}^{-g_\mu+k}(X) \leq \mathcal{E}_{s,t}(X) \leq Y_{s,t}^{g_\mu+k}(X)$ for all $X \in \mathbb{L}^2(\mathcal{F}_t)$,
- $\mathcal{E}_{s,t}(X) - \mathcal{E}_{s,t}(X') \leq Y_{s,t}^{g_\mu}(X - X')$ for all $X, X' \in \mathbb{L}^2(\mathcal{F}_t)$.

Then, there exits a standard driver with $g(t,0,0) = k_t$ such that

$$\mathcal{E}_{\mathbf{s},\mathbf{t}}(\mathbf{X}) = \mathbf{Y}^{\mathbf{g}}_{\mathbf{s},\mathbf{t}}(\mathbf{X}).$$

Extension to a domination by quadratic BSDEs **[Hu, Ma, Peng, Yao '08...]**

Qualitative properties on g tranfer to the $Y^g_{s,t}(X)$: sub-additivity, positive homogeneity, convexity... See **[Barrieu, El Karoui '09...]**

1.4 Reflected BSDEs and American options [EKP+97]

? $\exists$ solution $(\mathbf{Y}, \mathbf{Z}, \mathbf{K})$ to

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

Assumptions:

- standard Lipschitz driver f + augmented Brownian filtration
- $\Phi \in \mathbb{L}^2(\mathcal{F}_T)$
- The obstacle O is continuous adapted process, satisfying $\Phi \geq O_T$ and $\mathbb{E} \sup_{t \leq T} S_t^2 < \infty$.

Theorem. There is a unique triplet solution (Y, Z, K).

Applications to American options **[El Karoui, Kapoudjian, Pardoux, Peng, Quenez '97]**, to **switching problems** **[Hamadene, Jeanblanc '07...]**.

Applications to optimal stopping problems

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

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

which implies $\mathbf{Y_t} \geq \operatorname{ess}\sup_{\tau \in \mathcal{T}_{\mathbf{t},\mathbf{T}}} \mathbb{E}(\mathbf{O}_\tau \mathbf{1}_{\tau<\mathbf{T}} + \boldsymbol{\Phi} \mathbf{1}_{\tau=\mathbf{T}} + \int_{\mathbf{t}}^\tau \mathbf{f}(\mathbf{s}, \mathbf{Y_s}, \mathbf{Z_s})\mathbf{ds} | \mathcal{F}_{\mathbf{t}})$.

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

American options

Consider a linear driver $f(t, y, z) = -r_t y - z\theta_t$ (self-financing condition without constraints).

Theorem. Y_t is the price at time t of the American option with payoff $P_t = \mathbf{1}_{t=T}\Phi + \mathbf{1}_{t<T} O_t$: $\mathbf{Y_t} = \operatorname{ess}\sup_{\tau \in \mathcal{T}_{\mathbf{t},\mathbf{T}}} \mathbb{E}_{\mathbb{Q}}(\mathbf{e}^{-\int_{\mathbf{t}}^\tau \mathbf{r_s ds}} \mathbf{P}_\tau | \mathcal{F}_{\mathbf{t}})$.

Methods of construction of a solution

1. **Picard iteration + Snell envelops.**

   ⚠ Does not lead to a practical numerical method.

2. **Penalized BSDEs.** Consider the sequence of standard BSDEs $(Y^n, Z^n)_{n\geq 0}$ defined by

$$Y_t^n = \Phi + \int_t^T f(s, Y_s^n, Z_s^n)ds + \mathbf{n}\int_{\mathbf{t}}^{\mathbf{T}} (\mathbf{Y_s^n} - \mathbf{O_s})_{-}\mathbf{ds} - \int_t^T Z_s^n dW_s.$$

   - By comparison theorem, $Y^n \leq Y^{n+1}$, hence it converges to a process Y $\rightsquigarrow$ **lower approximation.**
   - We can prove that $Y_t \geq O_t$.
   - By setting $K_t^n = n\int_0^t (Y_s^n - O_s)_{-}ds$, one can prove that (Z^n, K^n) is a Cauchy sequence that the limit-triplet (Y^n, Z^n, K^n) converges to the RBSDE.

   ☺ The penalization approach can be turned into a numerical method.

   ☹ The driver and its Lispchitz constant increases like n!!

Methods of construction of a solution (Cont'd)

3. **Specific representation of the local time K.** **[Bally, Caballero, Fernandez, El Karoui '02]**

 Assume that the obstacle O has the Ito decomposition:

$$dO_t = U_t dt + V_t dW_t + dA_t^+$$

 with A^+ is a continuous increasing process, with dA_t^+ singular w.r.t. dt.

 Examples: call, put, convex payoffs...

 Then, one has

 - **smooth-fit condition**:

$$Z_t = V_t \textbf{ on the set } \{Y_t = O_t\}.$$

 - **absolute continuity of K**:

$$dK_t = \alpha_t \mathbf{1}_{Y_t = O_t}(f(t, O_t, V_t) + U_t)^- dt \textbf{ for some } \alpha_t \in [0, 1].$$

 Proof. The Ito decompositions of $d(Y_t - O_t)$ and $d(Y_t - O_t)_+$ coincide!! Proceed by identification.

An alternative representation of reflected BSDE [BCFK02]

$\exists$ solution $(\mathbf{Y}, \mathbf{Z}, \alpha)$ to

$$\begin{cases} Y_t = \Phi + \int_t^T f(s, Y_s, Z_s)ds + \int_{\mathbf{t}}^{\mathbf{T}} \alpha_{\mathbf{s}} \mathbf{1}_{\mathbf{Y_s = O_s}} (\mathbf{f}(\mathbf{s}, \mathbf{O_s}, \mathbf{V_s}) + \mathbf{U_s})_{-} \mathbf{ds} - \int_t^T Z_s dW_s, \\ \mathbf{Y_t} \geq \mathbf{O_t}. \end{cases}$$

Theorem. There is a unique solution (Y, Z, α) and $0 \leq \alpha \leq 1$.

α is uniquely determined only on $\{(s, \omega) : \mathbf{1}_{Y_s = O_s}(f(s, O_s, V_s) + U_s)_{-} > 0\}$.

By setting $K_t = \int_0^t \alpha_s \mathbf{1}_{Y_s = O_s}(f(s, O_s, V_s) + U_s)_{-} ds$, this proves that (Y, Z, K) is solution to the standard RBSDE.

Solving

$$Y_t = \Phi + \int_t^T f(s, Y_s, Z_s)ds + \int_t^T \alpha_s \mathbf{1}_{Y_s = O_s}(f(s, O_s, V_s) + U_s)_- ds - \int_t^T Z_s dW_s$$

The solution is obtained as follows:

- define a smooth function φ^n such that $\mathbf{1}_{[0,2^{-n}]} \leq \varphi^n \leq \mathbf{1}_{[0,2^{-(n-1)}]}$.
- consider the solution (Y^n, Z^n) of the standard BSDE with driver
$$\mathbf{f^n(s,\omega,y,z) = f(s,\omega,y,z) + \varphi^n(y - O_t)(f(s,O_s,V_s) + U_s)_-}.$$
- show that (Y^n, Z^n) converges to (Y, Z) and that α^n converges to $\alpha \mathbf{1}_{Y=O}$.

Then, Y^n is a **decreasing sequence converging to Y**.

$\Longrightarrow$ Very interesting for numerical methods since

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

2 Numerical methods

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 depend 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)...

2.1 Intricate combination of weak and strong approximations

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

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

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

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

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 $t_k = k\frac{T}{N} = kh$.

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

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 Brownian SDE (later, jumps could be included in X).

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 $u(.)$ 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

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

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

$$|u^N(t, .) - u(t, .)| = |\mathbb{E}(f(X_T)|X_t = x) - \mathbb{E}(f(X_T^N)|X_t^N = x)| = O(N^{-1})$$

$\Longrightarrow$ it seems that simulating accurately the underlying SDE in the strong approximation sense is necessary (stated later).

2.2 Resolution by dynamic programming equation

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

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

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

$$\Longrightarrow \begin{cases} \mathbf{Z}_{\mathbf{t_i}}^{\mathbf{N}} = \dfrac{\mathbf{1}}{\mathbf{\Delta t_i}}\mathbb{E}(\mathbf{Y}_{\mathbf{t_{i+1}}}^{\mathbf{N}} \mathbf{\Delta W}_{\mathbf{t_i}}^{*}|\mathcal{F}_{\mathbf{t_i}}), \\ \mathbf{Y}_{\mathbf{t_i}}^{\mathbf{N}} = \mathbb{E}(\mathbf{Y}_{\mathbf{t_{i+1}}}^{\mathbf{N}} + \mathbf{\Delta t_i f(t_i, X_{t_i}^N, Y_{t_{i+1}}^N, Z_{t_i}^N)}|\mathcal{F}_{\mathbf{t_i}}) \text{ and } \mathbf{Y_T^N = \Phi(X_T^N)}. \end{cases}$$

This is a discrete backward iteration. The scheme is of **explicit** type.

Implicit scheme

More closely related to the idea of discrete BSDE.

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

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

$$\rightsquigarrow \begin{cases} Z^N_{t_i} = \dfrac{1}{\Delta t_i}\mathbb{E}(Y^N_{t_{i+1}} \Delta W^*_{t_i} | \mathcal{F}_{t_i}), \\ \mathbf{Y}^{\mathbf{N}}_{\mathbf{t_i}} = \mathbb{E}(\mathbf{Y}^{\mathbf{N}}_{\mathbf{t_{i+1}}} | \mathcal{F}_{\mathbf{t_i}}) + \boldsymbol{\Delta}\mathbf{t_i}\mathbf{f}(\mathbf{t_i}, \mathbf{X}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Y}^{\mathbf{N}}_{\mathbf{t_i}}, \mathbf{Z}^{\mathbf{N}}_{\mathbf{t_i}}). \end{cases}$$

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

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

Rates of convergence of explicit and implicit schemes coincide for Lipschitz driver.

The explicit scheme is the simplest one, and presumably sufficient for Lipschitz driver.

2.3 Error analysis

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

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

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

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

Remarks on generalized BSDEs

Forward jump SDE:

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s + \int_0^t \int_E \beta(s, X_{s^-}, e)\tilde{\mu}(ds, de),$$

Generalized BSDE (with Lipschitz driver):

$$-\mathrm{d}Y_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t - dL_t, \qquad Y_T = \Phi(X_T),$$

where L is càdlàg martingale orthogonal to W **[Barles, Buckdhan, Pardoux '97; El Karoui, Huang '97]**.

Then,

- the same dynamic programming equation holds to compute (Y, Z).
- error estimates are unchanged **[Lemor, G. '05]**.

Proof for the Y-component

$Y_{t_i} - Y^N_{t_i} = \mathbb{E}_{t_i}(Y_{t_{i+1}} - Y^N_{t_{i+1}}) + \mathbb{E}_{t_i}\int_{t_i}^{t_{i+1}}\{f(s,X_s,Y_s,Z_s) - f(t_i, X^N_{t_i}, Y^N_{t_{i+1}}, Z^N_{t_i})\}ds.$
Then, use Young inequality $\mathbf{(a+b)^2 \leq (1+\gamma\Delta t_i)a^2 + (1+\frac{1}{\gamma\Delta t_i})b^2}$ to get

$$\mathbb{E}|Y_{t_i} - Y^N_{t_i}|^2 \leq (1+\gamma\Delta t_i)\mathbb{E}|\mathbb{E}_{t_i}(Y_{t_{i+1}} - Y^N_{t_{i+1}})|^2 + (1+\frac{1}{\gamma\Delta t_i})4L_f^2\Delta t_i\mathbb{E}\int_{t_i}^{t_{i+1}}|Z_s - Z^N_{t_i}|^2\mathrm{d}s$$

$$+ (1+\frac{1}{\gamma\Delta t_i})4L_f^2\Delta t_i({\Delta t_i}^2 + \int_{t_i}^{t_{i+1}}\mathbb{E}|X_s - X^N_{t_i}|^2\mathrm{d}s + \int_{t_i}^{t_{i+1}}\mathbb{E}|Y_s - Y^N_{t_{i+1}}|^2\mathrm{d}s).$$

Gronwall's lemma? $\gamma =?$

- $\mathbb{E}\int_{t_i}^{t_{i+1}}|Z_s - Z^N_{t_i}|^2\mathrm{d}s = \mathbb{E}\int_{t_i}^{t_{i+1}}|Z_s - \overline{Z}_{t_i}|^2\mathrm{d}s + \Delta t_i\mathbb{E}|\overline{Z}_{t_i} - Z^N_{t_i}|^2.$
- $\Delta t_i\mathbb{E}|\overline{Z}_{t_i} - Z^N_{t_i}|^2 \leq$
  $C\{\mathbb{E}|Y_{t_{i+1}} - Y^N_{t_{i+1}}|^2 - \mathbb{E}|\mathbb{E}_{t_i}(Y_{t_{i+1}} - Y^N_{t_{i+1}})|^2\} + C\Delta t_i\mathbb{E}\int_{t_i}^{t_{i+1}}f(s,X_s,Y_s,Z_s)^2ds.$
- $\mathbb{E}|X_s - X^N_{t_i}|^2 \leq 2\mathbb{E}|X_{t_i} - X^N_{t_i}|^2 + 2\mathbb{E}|X_s - X_{t_i}|^2 \leq 2\mathbb{E}|X_{t_i} - X^N_{t_i}|^2 + C\Delta t_i.$
- $\mathbb{E}|Y_s - Y^N_{t_{i+1}}|^2 \leq$
  $3\mathbb{E}|Y_{t_{i+1}} - Y^N_{t_{i+1}}|^2 + 3\mathbb{E}\int_{t_i}^{t_{i+1}}|Z_s|^2ds + 3\Delta t_i\mathbb{E}\int_{t_i}^{t_{i+1}}f(s,X_s,Y_s,Z_s)^2ds.$

After simplifications, we obtain:

$$\mathbb{E}|Y_{t_i} - Y_{t_i}^N|^2 \leq (1 + C\Delta t_i)\mathbb{E}|Y_{t_{i+1}} - Y_{t_{i+1}}^N|^2 + C\Delta t_i{}^2 + C\Delta t_i \max_{0\leq i\leq N} \mathbb{E}|X_{t_i} - X_{t_i}^N|^2$$
$$+ C\mathbb{E}\int_{t_i}^{t_{i+1}} |Z_s - \overline{Z}_{t_i}|^2 ds + C\Delta t_i \mathbb{E}\int_{t_i}^{t_{i+1}} (f(s, X_s, Y_s, Z_s)^2 + |Z_s|^2) ds.$$

Discrete Gronwall's lemma yields

$$\max_{0\leq k\leq N} \mathbb{E}|Y_{t_i}^N - Y_{t_i}|^2 \leq C|\pi| + C \max_{0\leq i\leq N} \mathbb{E}|X_{t_i} - X_{t_i}^N|^2$$
$$+ C\sum_{i=0}^{N-1} \mathbb{E}\int_{t_i}^{t_{i+1}} |Z_s - \overline{Z}_{t_i}|^2 ds + C \underbrace{\mathbb{E}|Y_T^N - Y_T|^2}_{=\mathbb{E}|\Phi(X_T^N) - \Phi(X_T)|^2} .$$

2.4 Strong approximation $\sup_{i\leq N}\mathbb{E}|X^N_{t_i} - X_{t_i}|^2$

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 $O(N^{-1})$.
- Otherwise, Milshtein scheme to get N^{-1}-rate.

2.5 Strong approximation of the terminal condition

- If Φ Lipschitz, then $\mathbb{E}|\Phi(X_T^N) - \Phi(X_T)|^2 \leq L_\Phi^2 \mathbb{E}|X_T^N - X_T|^2$.
- If Φ is irregular

Some results of Avikainen [Avi09] for discontinuous function ($\Phi(x) = \mathbf{1}_{x \leq a}$).

Also useful for the Multi-Level Monte Carlo methods of Giles [Gil08].

Theorem. If X_T has a bounded density $p(.)$, then for any $p > 0$

$$\sup_{\mathbf{a} \in \mathbb{R}} \mathbb{E}|\mathbf{1}_{\mathbf{X_T^N} < \mathbf{a}} - \mathbf{1}_{\mathbf{X_T} < \mathbf{a}}| \leq \mathbf{9} \left(|\mathbf{p}|_{\mathbb{L}_\infty} \|\mathbf{X_T^N} - \mathbf{X_T}\|_{\mathbb{L}_\mathbf{p}} \right)^{\frac{\mathbf{p}}{\mathbf{p}+\mathbf{1}}}.$$

Optimal inequalities:

- if $\mathbb{E}|\mathbf{1}_{\hat{X}<a} - \mathbf{1}_{X<a}| \leq C(X, a, p, r)\|\hat{X} - X\|_{\mathbb{L}_p}^r$ for any r.v. X with bounded density, then $r \leq \frac{p}{p+1}$.
- if $\mathbb{E}|\mathbf{1}_{\hat{X}<a} - \mathbf{1}_{X<a}| \leq C(X, p_0)\|\hat{X} - X\|_{\mathbb{L}_p}^{\frac{p}{p+1}}$ for any $p \geq p_0$, any a and any $\hat{X}$, then X has a bounded density.

$\Longrightarrow$

$$\begin{aligned}\mathbb{E}|\Phi(X_T^N) - \Phi(X_T)|^2 &= \mathbb{E}|\mathbf{1}_{X_T^N \le a} - \mathbf{1}_{X_T \le a}|^2 \\ &\le C_p(\|X_T^N - X_T\|_{\mathbb{L}_p})^{p/(p+1)} \\ &\le C'_p N^{-\frac{1}{2}\frac{p}{p+1}}.\end{aligned}$$

Hence, the convergence rate decreases from N^{-1} to $N^{-\frac{1}{2}+\epsilon}$ for any $\epsilon > 0$.

(under a non degeneracy assumptions on the SDE).

Possible generalization to functions with bounded variation **[Avikainen '09]**.

For intermediare regularity functions, open questions.

2.6 The L_2-regularity of Z

$\mathbb{L}_2$-regularity of Z-component

Define $\mathcal{E}^{\mathbf{Z}}(\pi) = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|\mathbf{Z_t} - \bar{\mathbf{Z}}_{\mathbf{t_i}}|^2 \mathbf{dt}$.

Theorem. **[Convergence to 0]** Since the $\bar{Z}$ is the a L_2-projection of Z, in full generality one has

$$\lim_{|\pi|\to 0} \mathcal{E}^Z(\pi) = 0.$$

Theorem. **[Ma, Zhang '02 '04]** Assume a Lipschitz driver f and a Lipschitz terminal condition Φ.

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

No ellipticity assumption.

Sketch of proof

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

$$'' \mathcal{D}_{\mathbf{t}}\xi = \partial_{\mathbf{dW_t}}\xi(\mathbf{W_t} : \mathbf{t} \geq \mathbf{0}). \text{ ``}$$

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 $X = g(\int_0^T h_t^1 dW_t, \cdots, \int_0^T h_t^n dW_t)$,
$$\mathcal{D}_t X = \sum_{i=1}^n \partial_i g(\cdots) h_t^i \mathbf{1}_{t\leq T}.$$
- chain rule for $\xi = g(X)$ with smooth g: $\mathcal{D}_t\xi = g'(X)\mathcal{D}_t X$.

- duality relation with adjoint operator $\mathcal{D}^*$: $\mathbb{E}(\int_{\mathbb{R}^+} u_t.\mathcal{D}_t\xi \, dt) = \mathbb{E}(\mathcal{D}^*(u)\xi)$ (known as integration by parts formula).

  If u is adapted and in $\mathbb{L}_2$, then $\mathcal{D}^*(u) = \int_0^T u_t dW_t$ (usual stochastic Ito-integral).

- Clark-Ocone's formula: if $\xi \in \mathbb{L}_2(\mathcal{F}_T)$ and in $\mathbb{D}_{1,2}$:

$$\xi = \mathbb{E}(\xi) + \int_0^T \mathbb{E}(\mathcal{D}_t\xi|\mathcal{F}_t)dW_t.$$

  Provides a representation of the Z when the driver is null.

- if $X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$, then for $r \leq t$

$$\begin{aligned}\mathcal{D}_\mathbf{r}\mathbf{X}_\mathbf{t} &= \int_r^t b'(s, X_s)\mathcal{D}_r X_s ds + \int_r^t \sigma'(s, X_s)\mathcal{D}_r X_s dW_s + \sigma(r, X_r)\\ &= \nabla\mathbf{X}_\mathbf{t}[\nabla\mathbf{X}_\mathbf{r}]^{-1}\sigma(\mathbf{r}, \mathbf{X}_\mathbf{r}).\end{aligned}$$

- $\mathcal{D}_\mathbf{t}\mathbf{X}_\mathbf{t} = \sigma(\mathbf{t}, \mathbf{X}_\mathbf{t})$.

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 $\theta \leq t \leq T$

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

$\Longrightarrow (\mathcal{D}_\theta Y_t, \mathcal{D}_\theta Z_t)_{t\in[\theta,T]}$ 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]**

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

Application to the study of the $\mathbb{L}_2$-regularity of Z:
$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathbb{E}|Z_t - \bar{Z}_{t_i}|^2 dt$

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

- an absolutely continuous part (in dt) $\rightsquigarrow$ **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]**.

The case of irregular terminal function $\Phi(X_T)$ [G., Makhlouf '09]

In the following, we assume strict ellipticity.

If not, Z can be discontinuous at some points [Zha05] ...

Sketch of proof.

1. We study the case with $f \equiv 0$. It gives the significative contribution.
2. We study the BSDE-difference $(Y^{f\neq 0} - Y^{f=0}, Z^{f\neq 0} - Z^{f=0})$. The L_2-regularity of $Z^{f\neq 0} - Z^{f=0}$ is still nicer, since it has zero terminal condition.

The BSDE with null driver

We first approximate $\Phi(X_T) \in \mathbb{L}_2$ by a sequence of bounded terminal conditions $\Phi^M(S_T) = M \wedge \Phi(X_T) \vee -M \xrightarrow{\mathbb{L}_2} \Phi(X_T)$ and then deduce by stability results.

$u(t,x) := \mathbb{E}\left[\Phi(X_T)|X_t = x\right]$ solves

$$\partial_t u(t,x) + \sum_{i=1}^{d} b_i(t,x)\partial_{x_i} u(t,x) + \frac{1}{2}\sum_{i,j=1}^{d} [\sigma\sigma^*]_{i,j}(t,x)\partial^2_{x_i,x_j} u(t,x) = 0 \text{ for } t < T,$$
$$u(T,x) = \Phi(x)$$

From Itô's formula, we can identify the solution $(\mathbf{y}, \mathbf{z})$ to the BSDE

$$y_t = \Phi(X_T) - \int_t^T z_s dW_s.$$

$\rightsquigarrow y_t = u(t, X_t)$ and $z_t = \nabla_x u(t, X_t)\sigma(t, X_t)$

The index α to measure the regularity

For $\alpha \in (0, 1]$, set

$$K^{\alpha}(\Phi) := \mathbb{E}|g(X_T)|^2 + \sup_{t\in[0,T)} \frac{\mathbb{E}(\Phi(X_T) - \mathbb{E}(\Phi(X_T)|\mathcal{F}_t))^2}{(T-t)^{\alpha}}$$

and define

$$\mathbb{L}_{2,\alpha} = \{\Phi \text{ s.t. } K^{\alpha}(\Phi) < +\infty\}.$$

It measures the rate of decreasing of the integrated conditional variance of $\Phi(X_T)$.

The index α is also called **fractional regularity** (introduced by Geiss...).

Some examples:

1. Lipschitz$\Longrightarrow \Phi \in \mathbb{L}_{2,\alpha=1}$;
2. α-Holder $\Longrightarrow \Phi \in \mathbb{L}_{2,\alpha}$;
3. indicator function $\Longrightarrow \Phi \in \mathbb{L}_{2,\alpha=\frac{1}{2}}$.

Fractional regularity for indicator functions (digital payoffs)

Proof. Let $\Phi(x) = \mathbf{1}_{[0,\infty)}(x)$ and $(X_t) \equiv (W_t)$. One has

$$\mathbb{E}[\Phi(X_T) - \mathbb{E}(\Phi(X_T)|\mathcal{F}_t)]^2 = \mathbb{E}\int_t^T |u_x^{'}(s, W_s)|^2 ds.$$

Then

$$u(t,x) = \mathbb{P}(x + W_T - W_t \geq 0),$$

$$u'_x(t,x) = \frac{1}{\sqrt{2\pi(T-t)}} \exp - \frac{x^2}{2(T-t)},$$

$$\mathbb{E}|u'_x(t, W_t)|^2 = \frac{1}{2\pi\sqrt{T+t}\sqrt{T-t}}$$

$\Longrightarrow \alpha = \frac{1}{2}$.

$\mathbb{L}_{2,\alpha}$= interpolation space between $\mathbb{L}_2$ and $\mathbb{D}_{1,2}$

Following **[Geiss, Hujo '07]**, one defines:

- the K-functional by

$$K(\Phi,\lambda;\mathbb{L}_2,\mathbb{D}_{1,2}) = \inf\{|\Phi^0|_{\mathbb{L}_2} + \lambda|\Phi^1|_{\mathbb{D}_{1,2}} \text{ such that } \Phi = \Phi^0 + \Phi^1\}.$$

- the space $(\mathbb{L}_2,\mathbb{D}_{1,2})_{\alpha,\infty}$ by the elements Φ such that

$$|\Phi|_{(\mathbb{L}_2,\mathbb{D}_{1,2})_{\alpha,\infty}} := \sup_{\lambda>0} \lambda^{-\alpha} K(\Phi,\lambda;\mathbb{L}_2,\mathbb{D}_{1,2}) < \infty.$$

In the BM case, possible in terms of sequences using the chaos decomposition.

Such Wiener chaos expansion enables to provide a Φ such that $\Phi(W_1) \notin \bigcup_{\alpha\in(0,1]} \mathbb{L}_{2,\alpha}$.

Equivalent estimates on u and its derivatives

Assume uniform ellipticity.

Lemma. Let $\alpha \in (0,1]$. Then the three following assertions are equivalent:

i) $\Phi \in \mathbb{L}_{2,\alpha}$.

ii) For some constant $C > 0$, $\forall t \in [0,T)$, $\int_0^t \mathbb{E}\left|D^2 u(s,X_s)\right|^2 ds \le \frac{C}{(T-t)^{1-\alpha}}$.

iii) For some constant $C > 0$, $\forall t \in [0,T)$, $\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 \le \frac{C}{(T-t)^{1-\alpha}}$.

And, if $\Phi \in \mathbb{L}_{2,\alpha}$, one can take C in i) and ii) proportional to $K^\alpha(\Phi)$.

If $\alpha < 1$ (resp. $\alpha = 1$), the previous three assertions are also equivalent to (resp. lead to) the following one:

iv) For some constant $C > 0$, $\forall t \in [0,T)$, $\mathbb{E}\left|D^2 u(t,X_t)\right|^2 \le \frac{C}{(T-t)^{2-\alpha}}$.

A general upper bound in $\mathbb{L}_{2,\alpha}$

For Φ in some $\mathbb{L}_{2,\alpha}$ ($\alpha \in (0,1]$), one has

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|z_t - \bar{z}_{t_i}|^2 dt \leq C(|\pi| K^\alpha(\Phi) T^\alpha + \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} (t_{k+1} - r)\mathbb{E}|D^2 u(r, X_r)|^2 dr)$$

Corollary. Assume $\Phi \in \mathbb{L}_{2,\alpha}$ ($\alpha \in (0,1]$). Then, for the uniform time grid,

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|z_t - \bar{z}_{t_i}|^2 dt = O(N^{-\alpha}).$$

The rate is optimal: for each $\alpha \in (0,1]$, one can exhibit a Φ achieving exactly this rate [GT01].

Theorem. Assume that $\Phi \in \mathbb{L}_{2,\alpha}$, for some $\alpha \in (0,1]$.
Now, take $\beta = 1$, if $\alpha = 1$, and $\beta < \alpha$ otherwise. Then, $\exists C > 0$ such that, for any time net $\pi = \{t_k, k = 0...N\}$,

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|z_t - \bar{z}_{t_i}|^2 dt \leq CK^\alpha(\Phi)T^\alpha|\pi| + CK^\alpha(\Phi)T^{\alpha-\beta} \sup_{k=0...N-1} \left(\frac{t_{k+1} - t_k}{(T - t_k)^{1-\beta}} \right).$$

Corollary. For $\alpha < 1$, the non-uniform grid

$$\pi^{(\beta)} := \{t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \leq k \leq N\}.$$

with $\beta < \alpha$ yields an error as N^{-1} for the L_2-regularity of Z.

By adapting the grid to the payoff regularity, we can maintain the rate $\frac{1}{N}$ for the $\mathbb{L}_2$-regularity of Z.

Back to the initial BSDE

We define the BSDE-difference

$$Y_t^0 := Y_t - y_t, \qquad Z_t^0 := Z_t - z_t.$$

solution in $\mathbb{L}_2$ of the BSDE with **null terminal condition** and **singular generator**

$$f^0(t,x,y,z) := f\left(t,x,y+u(t,x), z+\nabla_x u(t,x)\sigma(t,x)\right),$$

i.e.

$$Y_t^0 = \int_t^T f^0(s,X_s,Y_s^0,Z_s^0)ds - \int_t^T Z_s^0 dW_s.$$

Theorem. We have $Z_t - z_t = U_t\sigma(t,X_t)$ where (U,V) the solution of the following linear BSDE

$$U_t = \int_t^T \{a_r^0 + U_r\big(b_r^0 I_d + \nabla_x b(r, X_r) + \sum_{j=1}^q c_{j,r}^0 \nabla_x \sigma_j(r, X_r)\big) + \sum_{j=1}^q V_r^j \big(c_{j,r}^0 I_d + \sigma'_{j,r}\big)\}dr$$
$$- \sum_{j=1}^q \int_t^T V_r^j dW_r^j,$$

where we have set $f^0(t,x,y,z) = f\big(t, x, y + u(t,x), z + \nabla_x u(t,x)\sigma(t,x)\big)$ and

$$a_r^0 := \nabla_x f^0(r, X_r, Y_r^0, Z_r^0);$$
$$b_r^0 := \nabla_y f^0(r, X_r, Y_r^0, Z_r^0);$$
$$c_r^0 := \nabla_z f^0(r, X_r, Y_r^0, Z_r^0).$$

Proof. In general for $\Phi \in \bigcup_{\alpha\in(0,1]} \mathbb{L}_{2,\alpha}$, we have $\int_0^T \mathbb{E}|a_r^0|^2 dr = \infty$, but we can prove $\mathbf{\int_0^T |a_r^0|_{\mathbb{L}_2} dr < \infty}$ (one needs results from **[Briand, Delyon, Hu, Pardoux, Stoica '03]**)

Key point: to establish that the usual representation of Z^0 using Malliavin derivatives holds (**not trivial!!**)

Corollary. Assume that $g \in \mathbb{L}_{2,\alpha}$ ($\alpha \in (0,1]$). Then

$$|Z_t - z_t| \le C \int_t^T \frac{\sqrt{\mathbb{E}\left[\left(\Phi(X_T) - \mathbb{E}[\Phi(X_T)|\mathcal{F}_s]\right)^2 |\mathcal{F}_t\right]}}{T-s} ds + C(T-t).$$

$\Longrightarrow$

1. **$\mathbb{L}_2$-bounds:**

$$\mathbb{E}\,|Z_t - z_t|^2 \le C K^\alpha(\Phi)(T-t)^\alpha + C(T-t)^2.$$

2. **Pointwise bounds:** when Φ is α-Hölder continuous, it yields

$$|Z_t - z_t| \le C(T-t)^{\frac{\alpha}{2}} + C(T-t).$$

The $\mathbb{L}_2$-regularity of z (without driver) controls the $\mathbb{L}_2$-regularity of Z (with driver)

Corollary. Assume that $\Phi \in \mathbb{L}_{2,\alpha}$ ($\alpha \in (0,1]$). Then

$$\frac{1}{2}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}|z_t - \bar{z}_{t_i}|^2 dt + O(|\pi|) \le \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}|Z_t - \bar{Z}_{t_i}|^2 dt$$
$$\le 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}|z_t - \bar{z}_{t_i}|^2 dt + O(|\pi|).$$

To achieve the rate N^{-1} with N-points grid, one should choose,

- if $\alpha = 1$, uniform grids
- if $\alpha < 1$, the non-uniform grid

$$\pi^{(\beta)} := \{t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \le k \le N\}.$$

with an index $\beta < \alpha$.

Error expansion for smooth data and uniform grid [G., Labart '07]

Instead of upper bounds on $Y - Y^N$ and $Z - Z^N$ in L_2 norm, we expand the error.

Dynamic programming equation on the value function

Due to the Markov property of the Euler scheme $(X^N_{t_i})_i$, one has $Y^N_{t_i} = u^N(t_i, X^N_{t_i})$ and $Z^N_{t_i} = v^N(t_i, X^N_{t_i})$ where

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

Approximation result of weak type

Theorem. Assuming smooth data b, σ, f, Φ, one has

$$|u^N(t_i, x) - u(t_i, x)| \leq \frac{C(1 + |x|^k)}{N}$$

and

$$|v^N(t_i, x) - \nabla_x u(t_i, x)\sigma(t_i, x)| \leq \frac{C(1 + |x|^k)}{N}.$$

Proof. Adaptation of the Malliavin calculus approach of Kohatsu-Higa [KH01].

Global expansion

Corollary.

$$Y_{t_i}^N - Y_{t_i} = \nabla_x u(t_i, X_{t_i})(X_{t_i} - X_{t_i}^N) + O(|X_{t_i} - X_{t_i}^N|^2) + O(N^{-1})$$

and

$$Z_{t_i}^N - Z_{t_i} = [\nabla_x [\nabla_x u\sigma]^*(t_i, X_{t_i})(X_{t_i} - X_{t_i}^N)]^* + O(|X_{t_i} - X_{t_i}^N|^2) + O(N^{-1}).$$

Proof of corollary.

$$\begin{aligned} Y_{t_i}^N - Y_{t_i} &= u^N(t_i, X_{t_i}^N) - u(t_i, X_{t_i}) \\ &= u^N(t_i, X_{t_i}^N) - u(t_i, X_{t_i}^N) + u(t_i, X_{t_i}^N) - u(t_i, X_{t_i}) \\ &= O(N^{-1}) + \nabla u(t_i, X_{t_i})(X_{t_i} - X_{t_i}^N) + O(|X_{t_i} - X_{t_i}^N|^2). \end{aligned}$$ □

$\Longrightarrow$ **Strong approximation of the forward SDE is crucial.**

$\Longrightarrow$ At time 0, $\mathbf{Y_0^N - Y_0 = O(N^{-1})}$!!
First proved by Chevance [Che97] when f does not depend on z.

2.7 Resolution by Picard's iteration

BSDE = limit of a sequence of linear BSDE

$Y_t^{n+1} = u^{n+1}(t, X_t) = \mathbb{E}(\Phi(X_T) + \int_t^T f(s, X_s, Y_s^n, Z_s^n) ds | X_t)$

and

$Z_t^{n+1} = \nabla_x u^{n+1}(t, X_t)\sigma(t, X_t)$.

Allow (t, x) to play similar roles.

[Bender, Denk '07]; **[G., Labart '09]** with adaptive control variates.

Smaller errors propagation compared to the dynamic programming equation.

3 Computations of the conditional expectations

Our objective: to implement the dynamic programmin equation = to compute the conditional expectations $\rightsquigarrow$ the crucial step!!
Different points of view:

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

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

  $\rightsquigarrow$ 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= $\dim(X)$ $\rightsquigarrow$ curse of dimensionality.
- Markovian setting: $\mathbb{E}(g(X_{t_{i+1}})|X_{t_i})$ with $(X_{t_i})_i$ Markov chain.
  - To compute the transition operator from X_{t_i} to $X_{t_{i+1}}$? to compute the integral of g w.r.t. $\mathbb{P}_{X_{t_{i+1}}|X_{t_i}}(dx)$?

– To simulate the transition?

- How many regression functions to compute?

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

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

- 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 $X_{t_i} = x_1, ..., x_n$:

- one can simulate M independant paths of $X_{t_{i+1}}$ starting from $X_{t_i} = x_1, \cdots, x_n$ 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**).

The size of the tree grows linearly with N since **it recombines**.

In practice, feasible in dimension 1. Convergence: see **[Ma, Protter, San Martin, Torres '02]**.

Available for Ornstein-Uhlenbeck process (trinomial tree).

3.1 For more general dynamics: quantization [Graf, Luschgy '00]

Step 1. To discretize *optimally* the law of X_{t_j} for each j $\rightsquigarrow$ quantization.

Step 2. To use this quantized level to implement the dynamic programming equation.

Step 1. Computation of the grids. Fix the number of points M_j $(\to \infty)$.

Min. of the $\mathbb{L}_2$-distortion: $\mathcal{X}^{\mathbf{j}} = \{\mathbf{x}^{\mathbf{j}}_{\mathbf{m}} : \mathbf{1} \leq \mathbf{m} \leq \mathbf{M_j}\} = \text{argmin } \mathbb{E}(\min_{\mathbf{l}} |\mathbf{X}_{\mathbf{t_j}} - \mathbf{x}^{\mathbf{j}}_{\mathbf{l}}|^{\mathbf{2}})$.

- Existence of stochastic algorithm to compute these points (Kohonen algorithm).
- Quite slow. Better to compute them off-line.
- Suitable for $\mathbb{L}_2$-approximations (and Lipschitz functions).
- Grid already known in the case of Gaussian r.v. for various dimensions and various number of points **[see Pages' website]**.
- Rate of convergence available on the distorsion (Zador theorem: $M_j^{1/d}$) of the optimal grid.

Define Voronoi tesselations: $\mathcal{C}_k(\mathcal{X}^j) = \{z \in \mathbb{R}^d : |z - x_k^j| = \min_l |z - x_l^j|\}$.

Step 2. Computation of conditional expectations.

$$\mathbb{E}(g(X_{t_{j+1}})|X_{t_j} = x_k^j) = \sum_{l=1}^{M_{j+1}} \alpha_{k,l} g(x_l^{j+1}).$$

Weights $\alpha_{k,l}^j =?$ $\rightsquigarrow \alpha_{k,l}^j \approx \dfrac{\mathbb{P}(X_{t_j} \in \mathcal{C}_k(\mathcal{X}^j), X_{t_{j+1}} \in \mathcal{C}_l(\mathcal{X}^{j+1}))}{\mathbb{P}(X_{t_j} \in \mathcal{C}_k(\mathcal{X}^j))}$.

Computed by Monte Carlo simulations of X (also done off-line).

To sum up:

- ☺ deterministic approximations
- ☺ many computations are made off-line
- ☹ require the pre-computations of quantified grids of weights

For RBSDEs (with f independent of z), see **[Bally, Pages '03]** .

3.2 Representation of conditional expectations 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

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

Formal proof (d=1): $\mathbb{E}(G|F=x) = \frac{\mathbb{E}(G\delta_x(X))}{\mathbb{E}(\delta_x(X))} = \frac{\mathbb{E}(G(\mathbf{1}_{F\le x})')}{\mathbb{E}((\mathbf{1}_{F\le x})')} = \frac{\mathbb{E}(\mathbf{1}_{F\le x}H_1(F,G)))}{\mathbb{E}(\mathbf{1}_{F\le x}H_1(F,1))}$.

- 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 variances $\rightsquigarrow$ 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.

3.3 The approach using projections and regressions

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 $m(.)$ 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!
- ...

In addition, the samples are not independant (since one has N-times iteration in the discrete BSDE).

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

3.3.1 Kernel estimators

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

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 L_2-error estimates available.

Remaining problems with the non-compact support of X (partially solved recently in **[G., Labart '09]** using weighted Sobolev space estimates).

Computational efficiency: to compute $m_{M,h}$ at one point, M computations.

3.3.2 Projection on a set of functions

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

$$\mathbb{E}(Y|X) = \operatorname{Arg}\min_g \mathbb{E}\left(Y - g(X)\right)^2$$

$$\approx \operatorname*{Argmin}_{\sum_{k=1}^K \alpha_k \phi_k(.)} \mathbb{E}(Y - \sum_{k=1}^K \alpha_k \phi_k(X))^2.$$

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

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

- For simplicity, one should have a system of orthonormal functions (w.r.t. the law of X).
- In practice, impossible except in few cases (Gaussian case using Hermitte polynomials, ...).
- In many situations, the law of X is not explicitely known.

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

  Its dimension 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**:

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

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

- Which set of functions leads to quick/efficient computations of (α_k^M)?

- How to prove convergence rates of $\alpha.\phi(.) - m(.)$ as $M \to \infty$ and $K \to \infty$ (for general laws for (X, Y))?

The case of polynomial functions

- Popular choice.
- Smooth approximation.

Global approximation: within few polynomials, a smooth $m(.)$ can be very well approximated.

But slow convergence for non smooth functions (non-linear BSDEs may lead to non-smooth functions).

Do projections on polynomials converge to $m(.)$? $\oplus_{k\geq 0}\mathcal{P}_k(X) = \mathbb{L}_2(X)$?

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? $\square$

In particular, if X is log-normal, olynomials of X are not dense in $\mathbb{L}_2$ (Feller counter-exemple)!! Compare with Longstaff-Schwartz algorithm [LS01].

In the good cases, convergence rates?

The case of local approximation

Piecewise constant approximations. $\boldsymbol{\phi_k = 1_{\mathcal{C}_k}}$ 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: $A = \text{Diag}(\mathbb{P}(X \in \mathcal{C}_i)_i)$.

$\Longrightarrow$

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

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

Possible easy **extensions to piecewise affine functions** (or polynomials).

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

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

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

$$\begin{aligned}
&\inf_{g=\sum_k \alpha_k \mathbf{1}_{\mathcal{C}_k}} \int_{\mathbb{R}^d} |g(x)-m(x)|^2 \mu(dx) \\
&\qquad \leq \sum_k \int_{\mathcal{C}_k} |m(x_k)-m(x)|^2\mu(dx) + \int_{[\cup_k \mathcal{C}_k]^c} m^2(x)\mu(dx) \\
&\qquad \leq \sum_k |\mathcal{C}|^2 \mu(\mathcal{C}_k) + |m|_\infty^2 \mu([\cup_k \mathcal{C}_k]^c) \\
&\qquad \leq |\mathcal{C}|^2 + |m|_\infty^2 \mu([\cup_k \mathcal{C}_k]^c).
\end{aligned}$$

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

Efficient choice of tesselations?

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 $O(\log(K))$.

- **Regular grid (hypercubes).**

  $k = (k_1, ..., k_d) \in \{0, .., K_1 - 1\} \times ... \times \{0, .., K_d - 1\}$ 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$=O(\max_i \Delta x_i)$.

  Quick search formula:

  $$x \in \mathcal{C}_k \text{ with } k = (k_1, ..., k_d) \text{ if } x_{i,\min} \leq x_i < x_{i,\max} \text{ and } k_i = \lfloor \frac{x_i - x_{i,\min}}{\Delta x_i} \rfloor.$$

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 $(X_1, Y_1), \cdots, (X_n, Y_n)$.
- $\sigma^2 = \sup_x \mathrm{Var}(Y|X = x) < \infty$
- $F_n = \mathrm{Span}(f_1, ...f_{K_n})$ 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 $(X_1, \cdots, X_n)$.

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

Theorem. $\mathbb{L}_2(\mu^n)$-error: $\mathbb{E}(|\hat{\mathbf{m}}_\mathbf{n} - \mathbf{m}|_\mathbf{n}^2 | \mathbf{X_1}, \cdots, \mathbf{X_n}) \leq \sigma^2 \frac{\mathbf{K_n}}{\mathbf{n}} + \min_{\mathbf{f} \in \mathbf{F_n}} |\mathbf{f} - \mathbf{m}|_\mathbf{n}^2$.

A little extra work would give bounds in $\mathbb{L}_2(\mu)$.

Proof

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

- $(f_1, ...f_{K_n})$ 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

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

Lemma. Denote $\mathbb{E}^*(.) = \mathbb{E}(.|X_1, \cdots, X_n)$. 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 $\alpha_j^* = \frac{1}{n}\sum_i f_j(X_i)m(X_i)$.
- As a conditional expectation, $\mathbb{E}^*(\tilde{m}_n(.)) = \sum_j \mathbb{E}^*(\alpha_j)f_j(.)$.

Then, $\mathbb{E}^*(\alpha_j) = \frac{1}{n}\sum_i f_j(X_i)\mathbb{E}^*(Y_i) = \frac{1}{n}\sum_i f_j(X_i)\mathbb{E}(m(X_i) + \epsilon_i|X_1, \cdots, X_n) = \alpha_j^*$.

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

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

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 $(X_1, \cdots X_n)$ are centered.**

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

Uniform law of large numbers

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

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

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

as a function of ϵ and n?

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

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

as $n \to \infty$.

ϵ-cover of $\mathcal{G}$

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

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

Set $\mathcal{N}_1(\epsilon, \mathcal{G}, \mathbf{Z_{1:n}})$=the smallest 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 $\mathcal{G} = \{-B \vee \sum_k \alpha_k \phi_k(.) \wedge B : (\alpha_1, \cdots, \alpha_K) \in \mathbb{R}^K\}$, then

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

☺ Enables to replace an empirical mean by its expectation, up to error ϵ with high probability (explicitely quantified).

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 $R = (R_0, \cdots, R_d)$:

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

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

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

$\rightsquigarrow$ Localized BSDEs

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

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

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

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

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

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

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

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

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

$\rightsquigarrow$ **Small impact of the threshold R. But more numerical stability.**

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

Projection on a finite dimensional space:

$$\mathbf{y_k^{N,R}}(\bullet) \approx \alpha_{\mathbf{0,k}}.\mathbf{p_{0,k}}(\bullet), \qquad \mathbf{z_{l,k}^{N,R}}(\bullet) \approx \alpha_{\mathbf{l,k}}.\mathbf{p_{l,k}}(\bullet).$$

(for instance, hypercubes as presented before).

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

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

$$\mathbf{y_k^{N,R,M}}(\bullet) = [\alpha^{\mathbf{M}}_{\mathbf{0,k}}.\mathbf{p_{0,k}}(\bullet)]_{\mathbf{y}}, \quad \mathbf{z_{l,k}^{N,R,M}} \approx [\alpha^{\mathbf{M}}_{\mathbf{l,k}}.\mathbf{p_{l,k}}(\bullet)]_{\mathbf{z}},$$

where $[\psi]_y = -C_\star \vee \psi \wedge C_\star, \quad [\psi]_z = -C_\star N^{1/2} \vee \psi \wedge C_\star N^{1/2}$.

$$\boxed{\rightsquigarrow Y_{t_k} \approx y_k^{N,R,M}(X_{t_k}^N), \quad Z_{l,t_k} \approx z_{l,k}^{N,R,M}(X_{t_k}^N).}$$

The final algorithm

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

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

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

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

$$\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) + h f^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}), [\alpha_{l,k}^M \cdot p_{l,k}(X_{t_k}^{N,m})]_z) - \alpha \cdot p_{0,k}(X_{t_k}^{N,m})|^2.$$

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

Error analysis

1. $M = \infty$: quite easy to analyse.
2. For fixed N and fixed set of functions, Central Limit Theorem on α as $M \to \infty$.
3. Non asymptotic estimates? **hard** because dependent regression operators.

Robust error bounds

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

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

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

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

$$+C\frac{C_\star^2}{h}\sum_{k=0}^{N-1}\Big\{\mathbb{E}\Big(K^M_{0,k}\exp(-\frac{Mh^3}{72C_\star^2K^M_{0,k}})\exp(CK_{0,k+1}\log\frac{C\ C_\star(K^M_{0,k})^{\frac12}}{h^{\frac32}})\Big)$$

$$+h\mathbb{E}\Big(K^M_{l,k}\exp(-\frac{Mh^2}{72C_\star^2R_0^2K^M_{l,k}})\exp(CK_{0,k+1}\log\frac{C\ C_\star R_0(K^M_{l,k})^{\frac12}}{h})\Big)$$

$$+\exp(CK_{0,k}\log\frac{C\ C_\star}{h^{\frac32}})\exp(-\frac{Mh^3}{72C_\star^2})\Big\}.$$

Convergence of the parameters in the case 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	$N=5$, $\delta=5$ $D=[60,140]$	$N=20$, $\delta=1$ $D=[60,200]$	$N=50$, $\delta=0.5$ $D=[40,200]$
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)	2.96(0.01)

Table 1: Results for the combination of Calls using **HC**.

Global polynomials (GP)

Polynomials of d variables with a maximal degree.

M	$N=5$ $d_y=1$, $d_z=0$	$N=20$ $d_y=2$, $d_z=1$	$N=50$ $d_y=4$, $d_z=2$	$N=50$ $d_y=9$, $d_z=9$
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(0.04)	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 2: Results for the calls combination using **GP**.

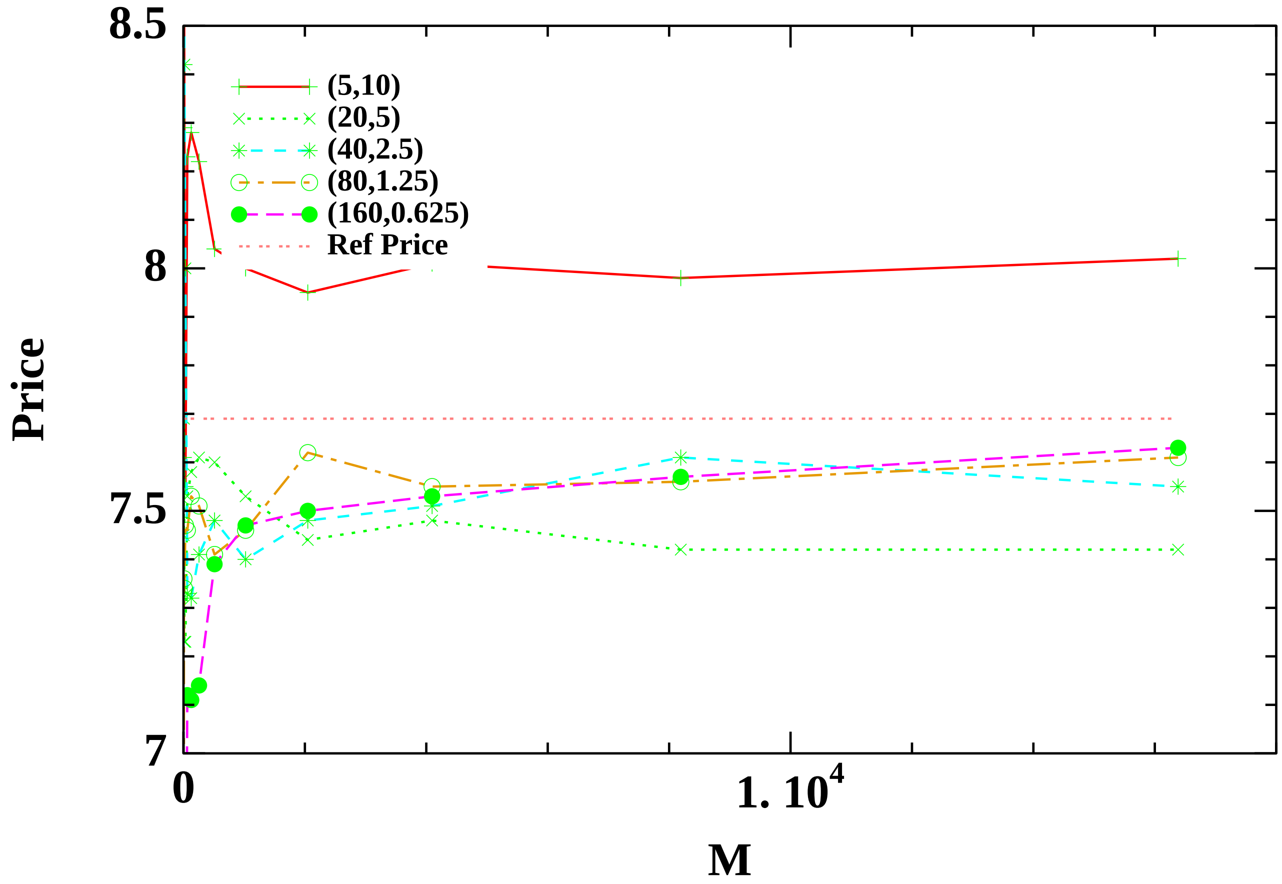
Large standard error $\rightsquigarrow$ GP not appropriate

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

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

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

Functions **HC**, parameters (N, δ).



American options via RBSDEs: several approaches

1. Taking the **max** with obstacle $\rightsquigarrow$ 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

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

   then $$dK_t = \alpha_t \mathbf{1}_{Y_t=\Phi(t,S_t)}(f(t, S_t, \Phi(t, S_t), V_t) + U_t)_- dt \text{ with } \alpha_t \in [0, 1].$$

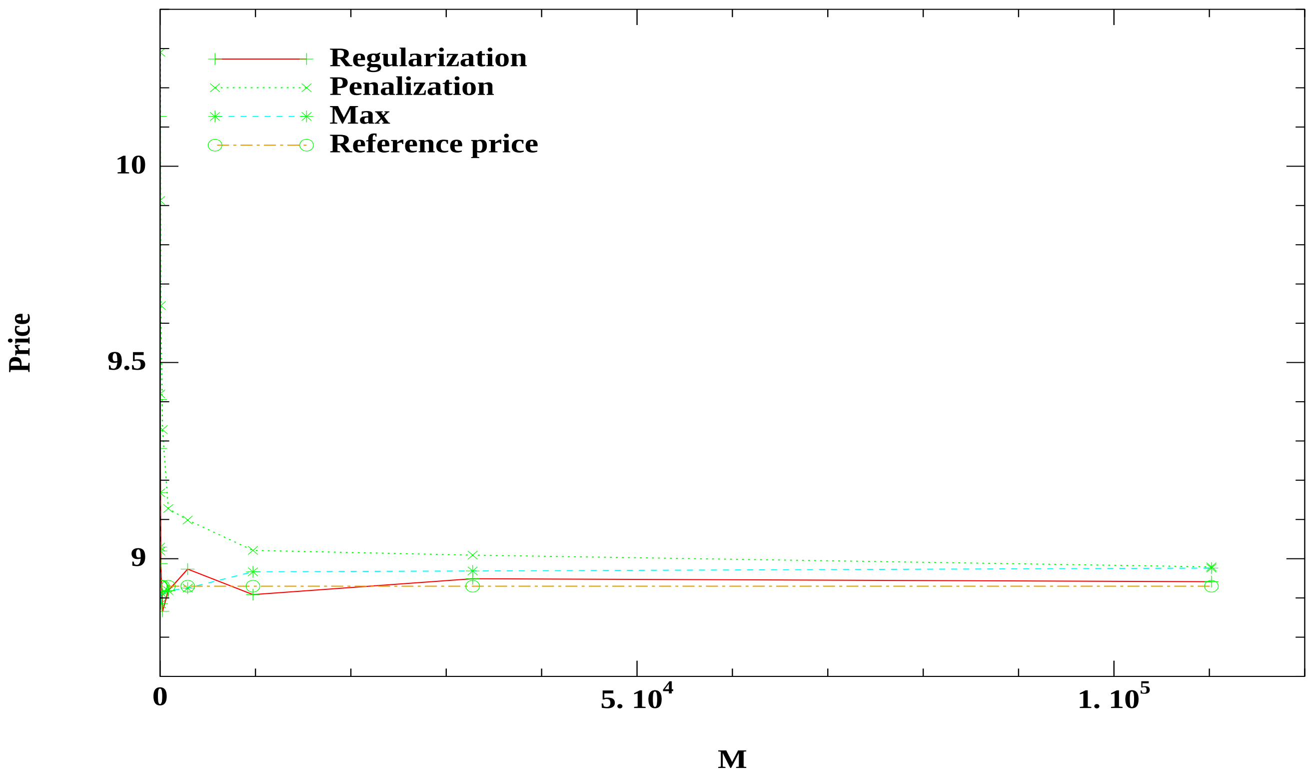
   Obtained as a limit of standard BSDEs with driver $f(s, S_s, Y_s, Z_s) + \rho_\lambda(Y_s - \Phi(s, S_s))(f(s, S_s, \Phi(s, S_s), V_s) + U_s)_-$ etc...
   Upper approximation.

Ex.3 : American option on three assets

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

T	r	σ	K	S_0^i	d
1	0.05	0.4 Id	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.

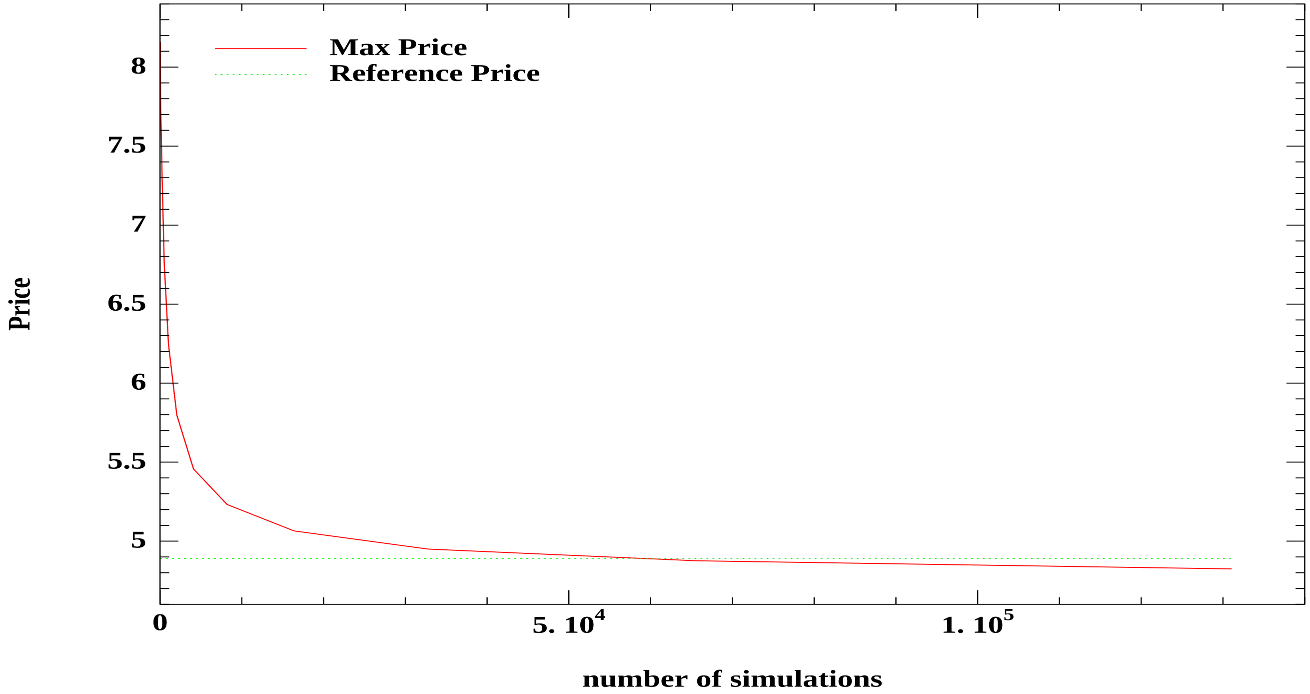
Regularization: $N = 32$, $\delta = 9$, $\lambda = 2$.

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

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

Ex.4 : American option 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

[Avi09] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13:381–401, 2009.

[BBP97] G. Barles, R. Buckdahn, and E. Pardoux. Backward stochastic differential equations and integral-partial differential equations. *Stochastics Stochastics Rep.*, 60(1-2):57–83, 1997.

[BC08] B. Bouchard and J.F. Chassagneux. Discrete time approximation for continuously and discretely reflected BSDE's. *Stochastic Processes and their Applications*, 118:2269–2293, 2008.

[BCFK02] V. Bally, M.E. Caballero, B. Fernandez, and N. El Karoui. Reflected BSDE's, PDE's and variational inequalities. *INRIA Report 4455*, 2002.

[BCZ05] V. Bally, L. Caramellino, and A. Zanette. Pricing american options by monte carlo methods using a malliavin calculus approach. *Monte Carlo Methods and Applications*, 11:97–133, 2005.

[BE08] B. Bouchard and R. Elie. Discrete-time approximation of decoupled Forward-Backward SDE with jumps. *Stochastic Processes and their Applications*, 118:53–75, 2008.

[BE09] P. Barrieu and N. El Karoui. Pricing, hedging and optimally designing derivatives via minimization of risk measures. In Rene Carmona, editor, *Volume on Indifference Pricing*. Princeton University Press, 2009.

[Ber95] Y.Z. Bergman. Option pricing with differential interest rates. *Rev. of Financial Studies*, 8(2):475–500, 1995.

[BP03] V. Bally and G. Pagès. Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Process. Appl.*, 106(1):1–40, 2003.

[BT96] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104-1:43–60, 1996.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[Che97] D. Chevance. Numerical methods for backward stochastic differential equations. In *Numerical methods in finance*, Publ. Newton Inst., pages 232–244. Cambridge Univ. Press, Cambridge, 1997.

[CK93] J. Cvitanić and I. Karatzas. Hedging contingent claims with constrained portfolios. *Ann. Appl. Probab.*, 3(3):652–681, 1993.

[CM96] J. Cvitanic and J. Ma. Hedging options for a large investor and Forward-Backward SDEs. *The Annals of Applied Probability*, 6(2):370–398, 1996.

[DE92] D. Duffie and L. Epstein. Stochastic differential utility. *Econometrica*, 60(2):353–394, 1992.

[EH97] N. El Karoui and S.J. Huang. A general result of existence and uniqueness of backward stochastic differential equations. In *Backward stochastic differential equations (Paris, 1995–1996)*, volume 364 of *Pitman Res. Notes Math. Ser.*, pages 27–36. Longman, Harlow, 1997.

[EKP+97] N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M.C. Quenez. Reflected solutions of backward SDE's and related obstacle problems for PDE's. *Ann. Probab.*, 25(2):702–737, 1997.

[EPQ97] N. El Karoui, S.G. Peng, and M.C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

[EQ95] N. El Karoui and M.C. Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control Optim.*, 33(1):29–66, 1995.

[EQ97] N. El Karoui and M.C. Quenez. Non-linear pricing theory and BSDEs. In *Financial mathematics (Bressanone 1996)*, pages 191–246. Lectures Notes in Maths 1656, 1997.

[ER01] N. El Karoui and R. Rouge. Pricing via utility maximization and entropy. *Math. Finance*, 10(2), 2001.

[FLL+99] E. Fournié, J.M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[FLLL01] E. Fournié, J.M. Lasry, J. Lebuchoux, and P.L. Lions. Applications of Malliavin calculus to Monte Carlo methods in finance, II. *Finance and Stochastics*, 5(2):201–236, 2001.

[GH07] S. Geiss and M. Hujo. Interpolation and approximation in $L_2(\gamma)$. *Journal of Approximation Theory*, 144:213–232, 2007.

[Gil08] M.B. Giles. Multilevel Monte Carlo path simulation. *Operation Research*, 2008.

[GKKW02] L. Gyorfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression.* Springer Series in Statistics, 2002.

[GL96] G. Golub and C.F. Van Loan. *Matrix computations.3rd ed.* Baltimore, MD: The Johns Hopkins Univ. Press. xxvii, 694 p. , 1996.

[GL00] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.

[GL06] E. Gobet and J. P. Lemor. Numerical simulation of BSDEs using empirical regression methods: theory and practice. In *Proceedings of the Fifth Colloquium on BSDEs (29th May - 1st June 2005, Shangai) - Available on http://hal.archives-ouvertes.fr/hal-00291199/fr/*, 2006.

[GL07] E. Gobet and C. Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic Processes and their Applications*, 117(7):803–829, 2007.

[GL09] E. Gobet and C. Labart. Solving BSDE with adaptive controle variate. *Submitted to SIAM Numerical Analysis*, 2009.

[GLW05] E. Gobet, J.P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3):2172–2202, 2005.

[GM09] E. Gobet and A. Makhlouf. L2-time regularity of BSDEs with irregular terminal functions. *In revision for Stochastic Processes and their Applications, available at http://hal.archives-ouvertes.fr/hal-00291768/fr/*, 2009.

[GT01] E. Gobet and E. Temam. Discrete time hedging errors for options with irregular pay-offs. *Finance and Stochastics*, 5(3):357–367, 2001.

[har92] *Applied nonparametric regression.* Cambridge University Press, 1992.

[HMPY08] Y. Hu, J. Ma, S. Peng, and S. Yao. Representation theorems for quadratic F-consistent nonlinear expectations. *Stochastic Processes and Their Applications*, 118(9):1518–1551, 2008.

[JK95] E. Jouini and H. Kallal. Arbitrage in securities markets with short-sales constraints. *Math. Finance*, 5(3):197–232, 1995.

[KH01] A. Kohatsu-Higa. Weak approximations. A Malliavin calculus approach. *Math. Comp.*, 70(233):135–172, 2001.

[Kor95] R. Korn. Contingent claim valuation in a market with different interest rates. *Mathematical Methods of Operations Research*, 42:255–274, 1995.

[LGW06] J.P. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.

[LS01] F. Longstaff and E.S. Schwartz. Valuing american options by simulation: A simple least squares approach. *The Review of Financial Studies*, 14:113–147, 2001.

[MPMT02] J. Ma, P. Protter, J. San Martín, and S. Torres. Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.*, 12(1):302–316, 2002.

[MZ02a] J. Ma and J. Zhang. Path regularity for solutions of backward stochastic differential equations. *Probab. Theory Related Fields*, 122(2):163–190, 2002.

[MZ02b] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *Ann. Appl. Probab.*, 12(4):1390–1418, 2002.

[Pen03] S. Peng. Nonlinear expectations, noninear evaluations and risk measures. In *Lectures presented in CIME EMS Bressanone, Italy July 6 13, 2003*, 2003.

[Sek06] J. Sekine. On exponential hedging and related quadratic backward stochastic differential equations. *Applied mathematics and optimization*, 54(2):131–158, 2006.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *Ann. Appl. Probab.*, 14(1):459–488, 2004.

[Zha05] J. Zhang. Representation of solutions to BSDEs associated with a degenerate FSDE. *Ann. Appl. Probab.*, 15(3):1798–1831, 2005.