Numerics of Backward SDEs

Emmanuel.Gobet@imag.fr

Laboratoire Jean Kuntzmann

Grenoble University

Agenda

- Lecture 1: short overview of theory of BSDEs and applications
- Lecture 2: different approaches for solving BSDEs, pros and cons.
  Picard iterations. Discrete time BSDE.
- Lecture 3: rates of convergence of time discretization
- Lectures 4 and 5: empirical regression methods and robust algorithms

1 Short overview of theory of BSDEs and applications

[Ref: Pardoux, Peng '90 ; Ma, Yong '99 ; El Karoui, Peng, Quenez '97 . . . ; El Karoui, Hamadene, Matoussi '08 for a recent review]

1.1 The simplest case

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a filtered probability space supporting a q-dimensional Brownian motion $(W_t)_{t\geq 0}$ (with $(\mathcal{F}_t)_t$= the $\mathbb{P}$-augmentation of $(\mathcal{F}_t^W)_t$).

Theorem. If ξ is a scalar r.v. in $\mathbb{L}_2(\mathcal{F}_T)$, then $\mathbf{Y_t} = \mathbb{E}(\xi|\mathcal{F}_\mathbf{t})$ is a $\mathbb{L}_2$-martingale which can be represented as a stochastic integral w.r.t. W of the (unique) adapted process $(Z_t)_t$ (with $\mathbb{E}(\int_0^T |Z_t|^2 dt) < \infty$):

$$\xi = \mathbb{E}(\xi) + \int_0^T Z_s dW_s,$$

$$\begin{aligned} Y_t &= \mathbb{E}(\xi|\mathcal{F}_t) \\ &= \mathbb{E}(\xi) + \int_0^t Z_s dW_s, \quad \textbf{(forward representation)} \\ &= \xi - \int_t^T Z_s dW_s. \quad \textbf{(backward representation)} \end{aligned}$$

$Y_t = \xi - \int_t^T Z_s dW_s$ (cont'd)

With the BSDE formalism, it writes

$$-\mathbf{dY_t} = -\mathbf{Z_t dW_t},$$
$$\mathbf{Y_T} = \xi.$$

Stochastic target problem: the target (**terminal condition**) ξ is usually given, as well the **terminal time** T (here assumed to be deterministic).

⚠ **Constraint: Y_t has to be $\mathcal{F}_t$-adapted** (taking $Y_t = \xi$ and $Z \equiv 0$ is not admissible).

$\rightsquigarrow$ The Z-process plays the role of a control, making Y adapted.

More general BSDEs

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

where $(y, z) \mapsto f(t, \omega, y, z)$ is the so-called **driver** or **generator** (possibly random).

1.2 Existence and uniquess in $\mathbb{L}_2$ for Lipschitz drivers

Assumptions:

- $f : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^q \mapsto \mathbb{R}$ is $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable ($\mathcal{P}$=set of $\mathcal{F}_t$-progressively measurable scalar processes on $\Omega \times [0,T]$).

  In practice, $\mathbf{f(t,\omega,y,z) = f(t, X_t(\omega), y, z)}$ where X solves a forward SDE and $f(t,x,y,z)$ is continuous.
- Lipschitz driver: $\mathbf{|f(t,\omega,y_1,z_1) - f(t,\omega,y_2,z_2)| \leq C_f(|y_1-y_2| + |z_1-z_2|)}$, uniformly in (t,ω).
- Bound on the driver: $\mathbb{E}(\int_0^T f^2(t,0,0)dt) < \infty$.

Notations:

1. $\mathbb{H}^2_{\beta,T}$ = set of $\mathbb{R}$ (or $\mathbb{R}^q$)-valued $\mathcal{F}$-adapted processes U such that $\mathbb{E}(\int_0^T e^{\beta t}|U_t|^2 dt) < \infty$.
2. $\mathbb{S}^2_{\beta,T}$ = set of scalar $\mathcal{F}$-adapted continuous processes Y such that $\mathbb{E}(\sup_{t\in[0,T]} e^{\beta t}|Y_t|^2) < \infty$.

Existence and uniqueness

Theorem. Under the previous notations and for any square-integrable terminal condition ξ, there is a unique solution (Y, Z) in $\mathbb{S}^2_{0,T} \times \mathbb{H}^2_{0,T}$ to the BSDE:

$$\begin{aligned} -\mathbf{dY_t} &= \mathbf{f(t, Y_t, Z_t)dt - Z_t dW_t}, \\ \mathbf{Y_T} &= \xi, \end{aligned}$$

or equivalently $\mathbf{Y_t} = \xi + \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{f(s, Y_s, Z_s)ds} - \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{Z_s dW_s}$.

Proof by Picard's fixed point theorem

Used two ingredients:

1. the solution (Y, Z) is the fixed point of a contracting mapping in the Hilbert space $\mathbb{H}^2_{\beta,T} \times \mathbb{H}^2_{\beta,T}$ (for some β depending on C_f),
2. a priori estimates.

A candidate for the mapping

Consider two processes $(y, z) \in \mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T}$ and set $(h_s = f(s, y_s, z_s))_s \in \mathbb{H}^2_{0,T}$.

Define

$$M_t = \mathbb{E}\Big(\xi + \int_0^T h_s ds | \mathcal{F}_t\Big).$$

One checks the following:

- M is a $\mathbb{L}_2$-martingale and for some $Z \in \mathbb{H}^2_{0,T}$

$$M_t = M_0 + \int_0^t Z_s dW_s = \xi + \int_0^T h_s ds - \int_t^T Z_s dW_s.$$

- By setting $Y_t := M_t - \int_0^t h_s ds$, one has $Y \in \mathbb{H}^2_{0,T}$.
- It defines a mapping $\Theta : (y, z) \in \mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T} \mapsto (Y, Z) \in \mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T}$.
- Backward representation:

$$Y_t = \xi + \int_t^T h_s ds - \int_t^T Z_s dW_s.$$

A priori estimates in $\mathbb{H}^2_{\beta,T}$ for β large enough

Take $(Y_1, Z_1) = \Theta(y_1, z_1)$ and $(Y_2, Z_2) = \Theta(y_2, z_2)$. Then, Ito's formula applied to $e^{\beta s}|Y_{1,s} - Y_{2,s}|^2$ gives

$$\begin{aligned} 0 =& \mathbb{E}(e^{\beta t}|Y_{1,t} - Y_{2,t}|^2 + \int_t^T \beta e^{\beta s}|Y_{1,s} - Y_{2,s}|^2 ds + \int_t^T e^{\beta s}|Z_{1,s} - Z_{2,s}|^2 ds) \\ &+ \underbrace{\mathbb{E}(\int_t^T 2e^{\beta s}(Y_{1,s} - Y_{2,s})(f(s, y_{1,s}, z_{1,s}) - f(s, y_{2,s}, z_{2,s}))ds)}_{\geq -\mathbb{E}\left(\int_t^T e^{\beta s}\left[2\frac{C_f^2}{\epsilon}|Y_{1,s}-Y_{2,s}|^2+\epsilon|y_{1,s}-y_{2,s}|^2+\epsilon|z_{1,s}-z_{2,s}|^2\right]ds\right)} \end{aligned}$$

using the inequality $2ab \leq \frac{a^2}{\epsilon} + \epsilon b^2$ for any $\epsilon > 0$. Taking $\epsilon = \frac{1}{2}$, $\beta = 1 + 4C_f^2$ and $t = 0$ gives

$$\mathbb{E}(\int_0^T e^{\beta s}[|Y_{1,s} - Y_{2,s}|^2 + |Z_{1,s} - Z_{2,s}|^2]ds) \leq \frac{1}{2}\mathbb{E}(\int_0^T e^{\beta s}\left[|y_{1,s} - y_{2,s}|^2 + |z_{1,s} - z_{2,s}|^2\right]ds),$$

that is

$$\|(Y_1 - Y_2, Z_1 - Z_2)\|^2_{\mathbb{H}^2_{\beta,T}} \leq \frac{1}{2}\|(y_1 - y_2, z_1 - z_2)\|^2_{\mathbb{H}^2_{\beta,T}} \quad !!$$

Application to effectively construct a solution

Construction of sequence of processes $(Y_k, Z_k)_k$ converging to (Y, Z) in $\mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T}$ sequence:

1. Initialization: $Y^0 \equiv 0$, $Z^0 \equiv 0$.
2. Iteration: set $(Y_{k+1}, Z_{k+1}) = \Theta(Y_k, Z_k)$, that is

$$Y_{k+1,t} = \xi + \int_t^T f(s, Y_{k,s}, Z_{k,s})ds - \int_t^T Z_{k+1,s} dW_s.$$

- Due to contraction propert of Φ, the convergence is geometric.
- At each step k, Y_k is given by conditional expectations.
  Could be computed...
- Z_k comes from Brownian martingale representation theorem.
  How to compute it?

$\rightsquigarrow$ ***Long is the road to a practical algorithm...***

1.3 Linear BSDE (see [EPQ97])

Consider the solution to $Y_T = \xi \in \mathbb{L}_2$ and

$$-\mathbf{dY_t} = [\varphi_\mathbf{t} + \mathbf{Y_t}\alpha_\mathbf{t} + \mathbf{Z_t}\gamma_\mathbf{t}]\mathbf{dt} - \mathbf{Z_t dW_t}$$

(with bounded (α, γ) and $\varphi \in \mathbb{H}^2_{0,T}$). The unique solution (Y, Z) in $\mathbb{S}^2_{0,T} \times \mathbb{H}^2_{0,T}$ is such that

$$\mathbf{Y_t} = \mathbb{E}[\xi \mathbf{\Gamma^t_T} + \int_\mathbf{t}^\mathbf{T} \mathbf{\Gamma^t_s} \varphi_\mathbf{s} \mathbf{ds} | \mathcal{F}_\mathbf{t}]$$

where $(\Gamma^t_s)_{s \geq t}$ solves the linear SDE

$$d\Gamma^t_s = \Gamma^t_s(\alpha_s ds + \gamma_s \cdot dW_s), \quad \Gamma^t_t = 1,$$

or equivalently $\Gamma^s_t = \exp(\int_t^s (\alpha_r - \frac{1}{2}|\gamma_r|^2)dr + \int_t^s \gamma_r \cdot dW_r)$.

Proof. Existence and uniqueness: clear.

Representation: check that $Y_t\Gamma^0_t + \int_0^t \Gamma^0_s \varphi_s ds$ is a uniformly integrable martingale.

Corollary: If $\varphi \geq 0$ and $\xi \geq 0$, then $Y \geq 0$.

Application to comparison of BSDEs

Theorem: consider (ξ_1, f_1) and (ξ_2, f_2) two standard parameters of BSDE and denote by (Y_1, Z_1) and (Y_2, Z_2) the two solutions in $\mathbb{S}^2_{0,T} \times \mathbb{H}^2_{0,T}$. Assume that

1. $\Delta\xi = \xi_1 - \xi_2 \geq 0$,
2. $\Delta f(t) = f_1(t, Y_{2,t}, Z_{2,t}) - f_2(t, Y_{2,t}, Z_{2,t}) \geq 0$ (one compares drivers along the second solution).

Then a.s for any t, we have $\mathbf{Y_{1,t} \geq Y_{2,t}}$.

Corollary: if $\xi \geq 0$ and $f(t, 0, 0) \geq 0$, then $Y_t \geq 0$ (generalization of the LBSDE case).

Remark: the comparison is strict (i.e. $Y_{1,0} = Y_{2,0}$ implies $\Delta\xi = 0$ and $\Delta f(t) \equiv 0$ a.s.).

Proof of comparison theorem

The BSDE difference $(\Delta Y, \Delta Z) = (Y_1 - Y_2, Z_1 - Z_2)$ is the unique solution of

$$
\begin{aligned}
\Delta Y_t =&\Delta\xi,\\
-d\Delta Y_t =&(f_1(t, Y_{1,t}, Z_{1,t}) - f_2(t, Y_{2,t}, Z_{2,t}))dt - \Delta Z_t dW_t\\
=&(f_1(t, Y_{1,t}, Z_{1,t}) - f_1(t, Y_{2,t}, Z_{1,t}))dt + (f_1(t, Y_{2,t}, Z_{1,t}) - f_1(t, Y_{2,t}, Z_{2,t}))dt\\
&+ \Delta f(t)dt - \Delta Z_t dW_t\\
=&[\alpha_t \Delta Y_t + \Delta Z_t \gamma_t + \Delta f(t)]dt - \Delta Z_t dW_t.
\end{aligned}
$$

$\rightsquigarrow$ This is a LBSDE ☺

Since $\Delta\xi \geq 0$ and $\Delta f(t) \geq 0$, we deduce $\Delta Y_t \geq 0$. □

Remark: we have only used the Lipschitz property of the driver f_1.

1.4 Different generalizations

Non brownian filtration

Assume now that $(\mathcal{F}_t)$ is a right-continous complete filtration.

Look for solution (Y, Z, L) to

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t dW_t - dL_t,$$
$$Y_T = \xi,$$

where Y is RCCL process, Z is predictable and L is a RCCL martingale, orthogonal to W.

Theorem. For square integrable terminal conditions and Lipschitz drivers, there is a unique solution (Y, Z, L) in $\mathbb{S}^2_{0,T} \times \mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T}$.

[REF: El Karoui-Peng-Quenez '97 , or Barles-Buckdahn-Pardoux '97 for drivers depending on L . . .]

Extension to BSDE in $\mathbb{L}_p$, $p > 1$

[REF: Briand-Delyon-Hu-Pardoux-Stroica '03]

One can replace

1. $\xi \in \mathbb{L}_2$ by $\xi \in \mathbb{L}_p$.
2. $\mathbb{E}(\int_0^T |f(t,0,0)|^2 ds) < \infty$ by $\int_0^T |f(t,0,0)| ds \in \mathbb{L}_p$.

Then, existence and uniqueness of solutions in $\mathbb{L}_p$-spaces.

Monotonic drivers

[REF: Darling, Pardoux '97 ...]

Assume that

1. $(y_1 - y_2)\big[f(t, y_1, z) - f(t, y_2, z)\big] \le \mu |y_1 - y_2|^2$ ($\mu \in \mathbb{R}$) (f is not necessarly Lipschitz in y).
2. $y \mapsto f(t, y, z)$ is continuous + a growth condition.
3. f is Lipschitz w.r.t. z.

Then, existence and uniqueness in L_2 (and L_p in [BDH$^+$03]).

Continuous drivers

[REF: Lepeltier, San Martin '97]

Assume that

1. f has a linear growth in y and z: $|f(t,y,z)| \le \alpha_t + k|y| + k|z|$ with $\alpha \in \mathbb{H}^2_{0,T}$
2. $(y,z) \mapsto f(t,y,z)$ is continuous

Then, **existence of a minimal solution** $(\underline{Y}, \underline{Z})$ and **a maximal solution** $(\overline{Y}, \overline{Z})$, i.e. for any other solution (Y,Z), one has

$$\underline{Y} \le Y \le \overline{Y} \qquad a.s.$$

Sketch of proof for the minimal solution :

1. Inf-convolution approximation: for $n \geq k$ define

$$\mathbf{f_n(t,y,z)} = \inf_{\mathbf{(y',z')\in\mathbb{Q}^{q+1}}}\{\mathbf{f(t,y',z')+n|(y-y',z-z')|}\}.$$

2. f_n is a standard Lipschitz driver with Lipschitz constant equal to n: denote by (Y_n, Z_n) the associated solution.

3. $(f_n)_n$ is increasing $\Longrightarrow Y_n \leq Y_{n+1} \Longrightarrow \dots (Y_n, Z_n)_n$ has a limit in $\mathbb{H}^2_{0,T} \times \mathbb{H}^2_{0,T}$. Denote by $(\underline{Y}, \underline{Z})$ this limit.

4. $f_n(t, y_n, z_n) \to f(t, y, z)$ for any $(y_n, z_n) \to (y, z) \Longrightarrow \dots$ the previous limit solves the BSDE.

5. Clearly $f_n(t, y, z) \leq f(t, y, z) \Longrightarrow Y_n \leq Y$ for any solution (Y, Z) $\Longrightarrow (\underline{Y}, \underline{Z})$ is the minimal solution.

Quadratic BSDE

[REF: Kobylanski '00, Lepeltier, San Martin '98]

Assume that

1. f has a linear growth in y and quadratic in z: $|f(t,y,z)| \le k(1+|y|+|z|^2)$,
2. $(y,z) \mapsto f(t,y,z)$ is continuous,
3. the terminal condition ξ is bounded.

Then, **existence of a maximal solution** $(\overline{Y}, \overline{Z})$ with a bounded $\overline{Y}$.

Extension to ξ with exponential growth condition, see [BH06].

Simple example of quadratic BSDE

$$y_t = \mathbb{E}(\exp(2\xi)|\mathcal{F}_t)) = \exp(2\xi) - \int_t^T z_s y_s dW_s,$$

$$Y_t = \frac{1}{2}\log(y_t) = \xi + \int_t^T \frac{z_s^2}{4} ds - \int_t^T \frac{z_s}{2} dW_s.$$

$\Longrightarrow (Y, \frac{z}{2})$ solves a BSDE with driver $f(t,y,z) = z^2$.

1.5 Connection with PDEs: formal link

Assume that $f(t,\omega,x,y) = f(t,X_t,y,z)$ and $\xi = g(X_T)$ where X is a forward SDE:

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t.$$

Take a smooth (?) solution u to

$$\begin{aligned}
&\partial_t u(t,x) + \sum_i b_i(t,x)\partial_{x_i} u(t,x) \\
&\qquad + \frac{1}{2}\sum_{i,j}[\sigma\sigma^*]_{i,j}(t,x)\partial^2_{x_i,x_j} u(t,x) + f(t,x,u(t,x),\nabla u\sigma(t,x)) = 0 \\
&u(T,x) = g(x).
\end{aligned}$$

Then by Ito's formula $\mathbf{Y_t = u(t,X_t)}$ and "$\mathbf{Z_t = \nabla u\sigma(t,X_t)}$" solves the BSDE with driver f and terminal condition $\xi = g(X_T)$.

⚠ In general, solutions in viscosity sense and not in classical sense (unless an ellipticity condition is fulfilled).

[Ref: [PP92], [Par98] . . .]

1.6 Reflected BSDEs and optimal stopping [EKP+97]

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

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

Assumptions:

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

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

Applications to optimal stopping problems

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

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

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

Methods of construction of a solution

1. **Picard iteration + Snell envelops.**
   So far, 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 Lipschitz 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 in finance: 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

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

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 $\mathbf{1}_{[0,2^{-n}]} \le \varphi^n \le \mathbf{1}_{[0,2^{-(n-1)}]}$.
- 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.

1.7 An application of BSDE: Pricing/Hedging of European style contingent claims

[Ref: El Karoui, Peng, Quenez '97 ; El Karoui, Quenez '97 ; Peng '03; El Karoui-Hamadène-Matoussi '08]

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

Usual assumptions:

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

   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).

1.7.1 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**, 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.**

1.7.2 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 $\xi(\mathbf{S_t} : \mathbf{0} \le \mathbf{t} \le \mathbf{T}) = \xi$ (a path-dependent functional of S).

? Possible to replicate of the option: $Y_T = \xi$?

? Link with the risk-neutral valuation rule?

Positive answer with BSDE

Theorem. If $\xi(S_t : 0 \leq t \leq T) \in \mathbb{L}_2(\mathbb{P})$, then there is a solution $(Y, Z) \in \mathbb{H}_2^2$ to the LBSDE and thus to the hedging problem.

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

Proof.

- Apply standard BSDE results for existence and uniqueness.
- The hedging strategy is given by $\phi_{\mathbf{t}} = \mathbf{Z}_{\mathbf{t}} \sigma_{\mathbf{t}}^{-\mathbf{1}}$.

Finally, this is a LBSDE, which has an explicit representation

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

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 $\mathbb{P}$!

1.7.3 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

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

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 $\rightsquigarrow$

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

  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

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

  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

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

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

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 $\mathbf{f}(\mathbf{t},\mathbf{y},\mathbf{z}) = -\mathbf{r}_\mathbf{t}\mathbf{y} - \mathbf{z}\theta_\mathbf{t}^\mathbf{l} + [\mathbf{z}\sigma_\mathbf{t}^{-\mathbf{1}}]^- \sigma_\mathbf{t}(\theta_\mathbf{t}^\mathbf{l} - \theta_\mathbf{t}^\mathbf{s})$.

1.7.4 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:

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

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: $\mathbf{dY_t = r_t Y_t dt} + \phi_t^1 \sigma_t^1 \theta_t \mathbf{dt} + \phi_t^1 \sigma_t^1 \mathbf{dW_t}$.

⚠ 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

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

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

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

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

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 $dY_t = r_t Y_t dt + Z_t \theta_t^1 dt + Z_t dW_t$ where $\boldsymbol{\theta_t^1 = [\sigma_t^1]^*[\sigma_t^1 \sigma_t^{1,*}]^{-1} \sigma_t^1 \theta_t}$.

Define $[\mathbf{Z_t^1}]^* := \text{Proj}^{\perp}_{\text{Range}([\sigma_t^1]^*)}(\mathbf{Z_t^*}) = [\sigma_t^1]^*[\phi_t^1]^*$ and $\mathbf{Z_t^2 := Z_t - Z_t^1}$.

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

It follows

- $Z_t \theta_t^1 = Z_t^1 \theta_t^1 + Z_t^2 \theta_t^1 = \phi_t^1 \sigma_t^1 \theta_t^1 + \underbrace{Z_t^2 [\sigma_t^1]^* [\sigma_t^1 \sigma_t^{1,*}]^{-1} \sigma_t^1 \theta_t}_{=0} = \phi_t^1 \sigma_t^1 \theta_t^1$,
- $Z_t dW_t = \phi_t^1 \sigma_t^1 dW_t + \underbrace{Z_t^2 dW_t}_{=:dM_t}$.

Thus, $\mathbf{dY_t = r_t Y_t dt + \phi_t^1 \sigma_t^1 \theta_t^1 dt + \phi_t^1 \sigma_t^1 dW_t + dM_t}$.

In addition, $\mathbf{< \int_0^{\cdot} \sigma_s^1 dW_s, M >_t = \int_0^t \sigma_s^1 [Z_s^2]^* ds = 0}$

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

Uniqueness is proved similarly.

1.8 An application of BSDE: dynamically consistent evaluation [(Peng '03)]

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

- $Y^{-g_\mu+k}_{s,t}(X) \le \mathcal{E}_{s,t}(X) \le Y^{g_\mu+k}_{s,t}(X)$ for all $X \in \mathbb{L}^2(\mathcal{F}_t)$,
- $\mathcal{E}_{s,t}(X) - \mathcal{E}_{s,t}(X') \le Y^{g_\mu}_{s,t}(X - X')$ 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... Interesting applications for risk measures.

See **[Artzner, Delbaen, Eber, Heath'99; Barrieu, El Karoui '09...]**

Other connections and applications

- 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.
- American options **[El Karoui, Kapoudjian, Pardoux, Peng, Quenez '97]**
- Switching problems **[Hamadene, Jeanblanc '07...]**.
- 2BSDE **[Cheridito, Soner, Touzi and Victoir '07 ...]**
- ...

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 independence (mixing of independent simulations)...

2.1 Quick overview of different probabilistic numerical approaches

1. **Picard iteration**. The non-linear problem is approximated by a sequence of linar problems: $Y_{k+1,t} = \xi + \int_t^T f(s, Y_{k,s}, Z_{k,s})ds - \int_t^T Z_{k+1,s} dW_s$.

   List of issues:

   - how to compute **conditional expectations**?
   - how to compute the predictable process in the Predictable Representation Theorem (like a gradient)?
   - impact of the approximated forward component simulation on the BSDE approximation?
   - convergence of the processes versus convergence of the value functions?
   - choice of the norms, to handle Picard iterations and value function approximation, in a closed form?
   - ...

   Related works: **[Labart PhD thesis ’07, Bender-Denk’07 , G.-Labart ’10]**

2. **Dynamic programming equation.**

- Split the interval $[0, T]$ into N sub-intervals of same size (why so?)
- On each small interval, the time variable can be made constant, the process $(Y_t, Z_t)_{0\leq t\leq T}$ becomes approximatively piecewise constant.

  $\rightsquigarrow$ **Discrete time BSDE** $(Y^N_{t_k}, Z^N_{t_k})_{0\leq k\leq N}$.

  Error analysis of the time discretization? **[Bally '97, Chevance '97 , Zhang '04. . .]**
- Leads to solve a system of N iterated conditional expectations.
  - **What method?**
    - binomial tree method or random walks techniques [BDM01] . . .
    - quantization methods [Che97, BP03] . . .
    - Malliavin calculus [BT04] . . .
    - non parametric regression [GLW05]. . .
    - cubature formulas [CM10].
  - **Which accuracy?** error propagations along the N iterated steps (in a more important way compared to Picard iterations).

2.2 Intricate mixing of weak and strong approximations

REMINDERS

Strong approximation. $(X_t^N)_{0\leq t\leq T}$ is a strong approximation of $(X_t)_{0\leq t\leq T}$ if $\sup_{t\leq T}\|X_t^N - X_t\|_{\mathbb{L}_p} \to 0$ (or $\|\sup_{t\leq T}|X_t^N - X_t|\|_{\mathbb{L}_p} \to 0$) as N goes to ∞.

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

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

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

Euler scheme: the simplest scheme to use. Define $t_k = k\frac{T}{N} = kh$.

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

Converges at rate $\frac{1}{2}$ for strong approximation and 1 for weak approximation.

Milshtein scheme (under restriction on σ): rate 1 for both strong and weak appr.

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

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

Formal error analysis

$$\mathbb{E}|Y_t^{N,M,K} - Y_t| \leq \mathbb{E}|u^{N,M,K}(t, X_t^N) - u(t, X_t^N)| + \mathbb{E}|u(t, X_t^N) - u(t, X_t)|$$
$$\leq |u^{N,M,K}(t,.) - u(t,.)|_{\mathbb{L}_\infty} + \|\nabla u\|_{\mathbb{L}_\infty} \mathbb{E}|X_t^N - X_t|.$$

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

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

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

2.3 Resolution by Picard iteration [G' and Labart '10]

Applied to **Markovian BSDEs**:

1. Forward component: $dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad 0 \le t \le T.$
2. Backward component: $-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t$ and $Y_T = \Phi(X_T)$.

Assumptions:

- f is bounded Lipschitz
- $\Phi \in C_b^{2+\alpha}$
- $b, \sigma \in C^{1,3}$
- uniform ellipticity

This numerical approach combines two ingredients:

1. **Picard iterations**: approximation of BSDEs by a sequence of linear BSDEs
2. **Iterative control variates** to efficiently solve linear PDEs/BSDEs [G' and Maire '05]

First ingredient: Picard iteration

BSDE = limit of a sequence of linear BSDE

We start with $\hat{Y}^0 = 0, \hat{Z}^0 = 0$.

We iteratively define $(\hat{Y}^{k+1}, \hat{Z}^{k+1})$ from $(\hat{Y}^k, \hat{Z}^k)$ by

$$\begin{cases} -d\hat{Y}_t^{k+1} &= f(t, X_t, \hat{Y}_t^k, \hat{Z}_t^k)dt - \hat{Z}_t^{k+1} dW_t, \\ \hat{Y}_T^{k+1} &= \Phi(X_T). \end{cases}$$

Representations as expectations:

$$\mathbf{\hat{Y}_t^k = u_k(t, X_t) = \mathbb{E}\Big(\Phi(X_T) + \int_t^T f(s, X_s, \hat{Y}_s^{k-1}, \hat{Z}_s^{k-1}) ds | X_t\Big),}$$

$$\mathbf{\hat{Z}_t^k = \nabla u_k(t, X_t)\sigma(t, X_t).}$$

Then, the sequence $(\hat{Y}^k, \hat{Z}^k)_k$ converges to (Y, Z)

- at a geometric rate
- in a suitable L_2 norm.

Equivalently: by writing $\hat{Y}^k_t = u_k(t, X_t)$ and $\hat{Z}^k_t = \nabla u_k(t, X_t)\sigma(t, X_t)$, one has

$$\partial_t \mathbf{u_k} + \mathcal{L}\mathbf{u_k} + \mathbf{f}(.,.,\mathbf{u_{k-1}}, \nabla \mathbf{u_{k-1}}\sigma) = \mathbf{0} \quad \text{and} \quad \mathbf{u_k}(\mathbf{T}, .) = \mathbf{\Phi}(.).$$

It means that the sequence of solutions of linear PDEs $(u_k, \nabla u_k)_k$ converges (in a L_2 norm) to $(u, \nabla u)$, solution of the semi-linear PDE

$$\partial_t \mathbf{u} + \mathcal{L}\mathbf{u} + \mathbf{f}(.,.,\mathbf{u}, \nabla \mathbf{u}\sigma) = \mathbf{0} \quad \text{and} \quad \mathbf{u}(\mathbf{T}, .) = \mathbf{\Phi}(.).$$

Remarks.

- Symmetric role of the variables (t, x) ($\neq$ from the dynamic programming equation)
- Geometric convergences. Which norms?
  - ? Norms on the processes versus norms on the value functions?

Second ingredient: adaptative control variates [G' and Maire '05]

Purpose: Monte Carlo resolution of linear PDEs of type

$$\partial_t u + \mathcal{L}u + f = 0 \quad \text{and} \quad u(T,.) = \Phi(.),$$

using an **efficient scheme which computes a global solution**.

Probabilistic solution: $u(t,x) = \mathbb{E}_{t,x}[\Phi(X_T) + \int_t^T f(X_s)ds] = \mathbb{E}(\mathbf{\Psi}(\mathbf{\Phi}, \mathbf{f}, \mathbf{X}^{\mathbf{t},\mathbf{x}}))$.

Principle: compute a sequence of solution $(u_k)_k$ by writing

$$u_{k+1} = u_k + \text{ Monte-Carlo evaluations of the error } (u - u_k).$$

Probabilistic representation of the correction term $c_k = u - u_k$:

$$c_k(t,x) = u(t,x) - u_k(t,x) = \mathbb{E}(\Psi(\Phi - u_k, f + \partial_t u_k + \mathcal{L}u_k, X^{t,x})).$$

⚠ This approach is different from the usual martingale control variates.

Numerical algorithm:

- take n **points** $(t_i, x_i)_{1\leq i\leq n} \subset \Re^d$: simulated stochastic processes (Euler scheme with N time steps) will start from these points.
- Evaluate $c_k(t_i, x_i)$ using M independent simulations
$$c_k^M(t_i, x_i) = \frac{1}{M}\sum_{m=1}^{M} \Psi(\Phi - u_k, f + \partial_t u_k + \mathcal{L}^N u_k, X^{t_i,x_i,N,m}).$$
- To construct the global solution $c_k^M(\bullet)$ based on the values $[c_k^M(t_i, x_i)]_{1\leq i\leq n}$, we use a linear approximation operator: $\boxed{\mathcal{P}c_k^M(\bullet) = \sum_{i=1}^n c_k^M(t_i, x_i)\omega_i(\bullet)}$ for some weight functions ω_i. **Examples**: interpolation, projection, Kernel-based estimator...

To sum up, we get $\boxed{\mathbf{u_{k+1}} = \mathcal{P}(\mathbf{u_k} + \mathbf{c_k^M}).}$

Main estimate: for some $\rho < 1$ (depending on M and N),

$$\|u - u^{k+1}\|_2^2 \leq \rho\|u - u^k\|_2^2 + C\|u - \mathcal{P}u\|_2^2(\frac{1}{N} + \frac{1}{M}).$$

Convergence at a geometric rate. No need to take N and M large.

Remarks

- Quick convergence up the approximation error given by the operator $\mathcal{P}$.
- Provides a smooth solution in t and x.
- Ready for massive parallel computing.
- Requires that $\mathcal{P}$ transforms pointwise evaluations into global C^2 functions (to compute $\mathcal{L}u_k$).
- In practice, the computations of $\mathcal{L}u_k$ may be quite time demanding, especially if $\mathcal{P}$ is non local operator.

Application to BSDEs

Iteration k. Suppose that a function u_k of class C^2 is built.

Correction term $c_k = u - u_k$:

$$c_k(t,x) = \mathbb{E}_{t,x}\Big[\Phi(X_T) - u_k(T, X_T^N) \\ + \int_t^T [f(s, X_s, u(s,X_s), \nabla u\sigma(s,X_s)) + (\partial_t + \mathcal{L}^N)u_k(s, X_s^N)]ds\Big]$$

At the points (t_i, x_i), it is pratically computed as

$$\frac{1}{M}\sum_{m=1}^{M}\Big[\Phi(X_T^{m,N}) - u_k(T, X_T^{m,N}) \\ + \int_t^T [f(s, X_s^{m,N}, u_k(s, X_s^{m,N}), \nabla u_k\sigma(s, X_s^{m,N})) + (\partial_t + \mathcal{L}^N)u_k(s, X_s^N)]ds\Big]$$

with independent simulated Euler schemes starting from (t_i, x_i).

Then, we take $u_{k+1} = \mathcal{P}^k(u_k + c_k^M)$.

Two issues to handle for the mathematical analysis

1. Choice of the grid $(t_i, x_i)_{1\leq i\leq n}$? different from one iteration to another?

2. Choice of operator $\mathcal{P}$? assumptions on $\mathcal{P}$?

3. Choice of the norm to measure errors?

Needs for a non asymptotic analysis.

Choice of the norm

Take $\mu > 0$ and $\beta > 0$. Define:

$$\|u\|^2_{H^\mu_{\beta,X}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \mathbb{E}|u(s, X^x_s)|^2 dx ds.$$

Equivalent to $\int_0^T e^{\beta s} \mathbb{E}|u(s, X^\mu_s)|^2 ds$ where $(X^\mu_s)_s$ stands for $(X_s)_s$ with a random initial value with a density proportional to $e^{-\mu|x|}$.

Motivations: norm equivalence results

1. $\|u\|^2_{H^\mu_{\beta,X}} \sim \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \mathbb{E}|u(s, x)|^2 dx ds$ (similar to **[Bally and Matoussi '01]**).
2. $\|u\|^2_{H^\mu_{\beta,X^N}} \leq_c \|u\|^2_{H^\mu_{\beta,X}}$;

   In the following, we simply write $\|u\|^2_{\mu,\beta}$.
3. **[Bensoussan-Lions '84]**: if u solves $\partial_t u + \mathcal{L}u + f = 0$ with $u(T, .) = 0$, then

$$\|u\|^2_{\mu,\beta} + \|\nabla u\|^2_{\mu,\beta} + \|D^2 u\|^2_{\mu,\beta} + \|\partial_t u\|^2_{\mu,\beta} \leq c\|f\|^2_{\mu,\beta}.$$

Main assumptions of the operator $\mathcal{P}$

1. $\mathcal{P}$ approximates well a function and its spatial derivatives: for any smooth function H (with $c(H) = \|H\|_\infty + \|\nabla H\|_\infty + \|\nabla H\|_{1/2,t} < \infty$):

$$\|H - \mathcal{P}H\|^2_{\mu,\beta} + \|\nabla H - \nabla(\mathcal{P}H)\|^2_{\mu,\beta} \leq \epsilon_2(\mathcal{P})\ c^2(H)$$
$$+ \epsilon_1(\mathcal{P})\ (\|H\|^2_{\mu,\beta} + \|\nabla H\|^2_{\mu,\beta} + \|D^2 H\|^2_{\mu,\beta} + \|\partial_t H\|^2_{\mu,\beta})$$

with $\epsilon_1(\mathcal{P}), \epsilon_2(\mathcal{P}) \to 0$.

2. For any random function $(t,x) \mapsto H(t,x)$ with $\mathbb{E}(H(t,x)) = 0$, one has

$$\mathbb{E}\|\mathcal{P}H\|^2_{\mu,\beta} + \mathbb{E}\|\nabla(\mathcal{P}H)\|^2_{\mu,\beta} \leq c_3(\mathcal{P})\mathbb{E}\|H\|^2_{\mu,\beta}.$$

General theorem: global error estimates

To simplify the exposure, we neglect the Euler scheme error ($N = +\infty$).

Set $Y_t^k = u_k(t, X_t)$ and $Z_t = \nabla u_k(t, X_t)\sigma(t, X_t)$.

Theorem. Define the quadratic error $\mathcal{E}_k = \|Y - Y^k\|_{\mu,\beta}^2 + \|Z - Z^k\|_{\mu,\beta}^2$. Then

$$\mathcal{E}_\mathbf{k} \leq \rho \mathcal{E}_\mathbf{k-1} + \eta$$

where

$$\rho = \underbrace{\frac{4(1+T)L_f^2}{\beta}}_{\text{Picard}} + C\Big(\underbrace{L_f^2\ \epsilon_1(\mathcal{P})}_{\text{operator } \mathcal{P}} + \underbrace{\frac{c_3(\mathcal{P})}{M}}_{\text{M.C.}} \Big), \qquad \eta = C\Big(\epsilon_1(\mathcal{P}) + \epsilon_2(\mathcal{P})\Big) c_{1,2}^2(u).$$

Corollary. For β, M and for $\mathcal{P}$-parameters large enough, we have $\rho < 1$ and

$$\limsup_{k\to\infty} \mathcal{E}_k \leq C \frac{\big(\epsilon_1(\mathcal{P}) + \epsilon_2(\mathcal{P})\big) c_{1,2}^2(u)}{1-\rho}.$$

Remark. The lim sup result holds also for $\beta = 0$.

Example of grids and operator $\mathcal{P}$: kernel estimator

Grid: at each iteration k, take a new grid of n points $(T_i^k, X_i^k)_{1\leq i\leq n}$, that are i.i.d. and uniformly distributed on $[0,T]\times[-a,a]^d$ (for a large enough).

Operator: defined by

$$\mathcal{P}_k v(t,x) = \sum_{i=1}^{n} \omega_i(t,x) v(T_i^k, X_i^k)$$

where

- the local weight $\omega_i(t,x)$ is proportional to $K_t(\frac{t-T_i^k}{h_t})K_x(\frac{x-X_i^k}{h_x})$
- $K_t(.)$ and $K_x(.)$ are two C^2- kernel functions, with compact support
- bandwith h_x and h_t.

☹ Kernel estimators are known to be not the most efficient in practice for high dimensional problems.

😐 But it satisfies the assumptions on the operator $\mathcal{P}$.

Derivations of the global error estimates

Theorem.

$$\mathcal{E}_k \leq \rho \mathcal{E}_{k-1} + \eta$$

where

$$\rho = \underbrace{\frac{4(1+T)L_f^2}{\beta}}_{\text{Picard}} + C\left(\underbrace{h_x^2 + h_t^2}_{\text{bias}^2} + \underbrace{\frac{T(2a)^d}{nh_t h_x^d}}_{\text{var.}}(1 + \underbrace{M^{-1}h_x^{-2}}_{\text{M.C.}})\right),$$

$$\eta = C\left(\underbrace{h_x^2 + h_t}_{\text{bias}^2} + \underbrace{\frac{T(2a)^d}{nh_t h_x^{d+2}}}_{\text{var.}} + \underbrace{h_x^{-1}e^{-\mu a}a^{d-1} + h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a}}_{\text{bounded grid}}\right).$$

These non asymptotic estimates enable to balance optimally the parameters h_x, h_t, n.

Few numerical experiments

Call option in Black-Scholes model

One-dimensional SDE:

$$dX_t = \left(b_0 - \frac{\sigma^2}{2}\right) dt + \sigma dW_t, \quad X_0 = x.$$

Driver: $f(t, x, y, z) = -ry - \theta z$ with $\theta = \frac{\mu_0 - r}{\sigma}$.

Terminal condition: $\Phi(x) = (e^x - K)^+$.

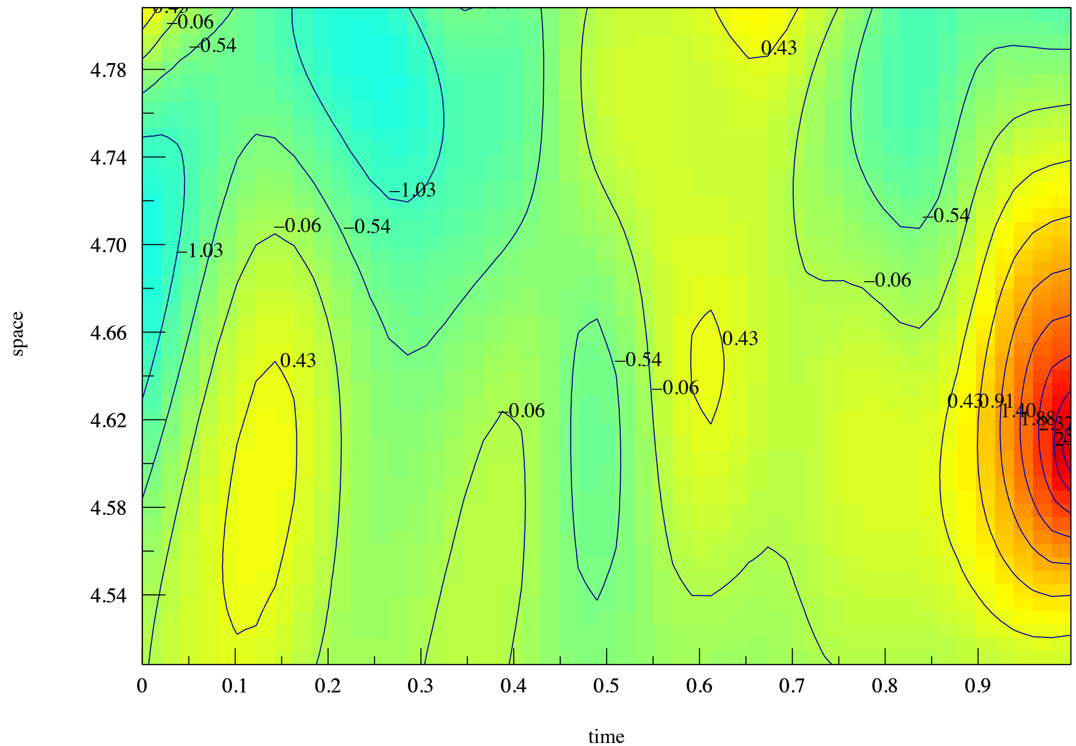
Parameters:

b_0	σ	r	T	K
0.1	0.2	0.02	1	100

n	N	M	h_x	h_t	$2a$	β	μ
2500	100	100	0.1	0.1	1.2	0	1

$\mathrm{Err}(Y^k - Y)^2$ and $\mathrm{Err}(Z^k - Z)^2$ w.r.t. $k \Longrightarrow$

	$E(Y^k - Y)$	$E(Z^k - Z)$
k=1	0.0743476	0.0265350
k=2	0.0014802	0.0104687
k=3	0.0010029	0.0082452
k=4	0.0008865	0.0076881
k=5	0.0008373	0.0075321



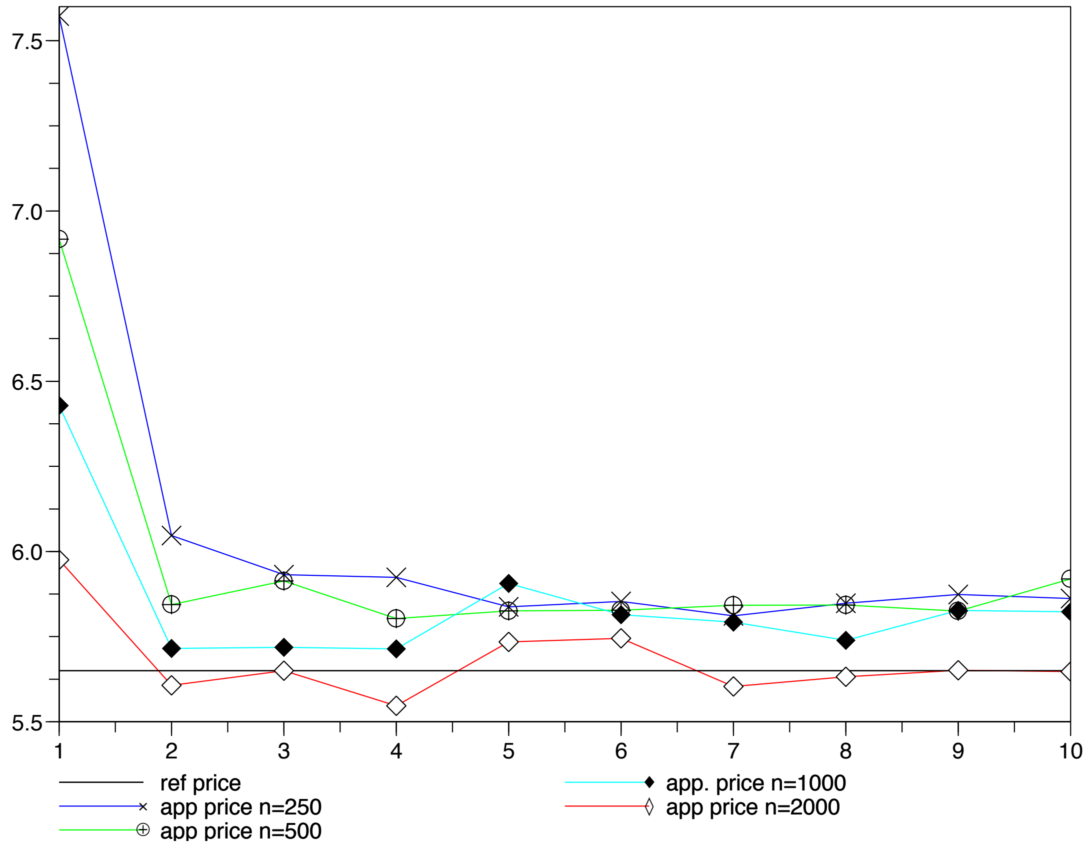
$\Longleftarrow$ Level sets for the price error at iteration 10

Three-dimensional example: basket call

Terminal condition: $\Phi(x) = (\frac{1}{3}(x_1 + x_2 + x_3) - K)^+$.

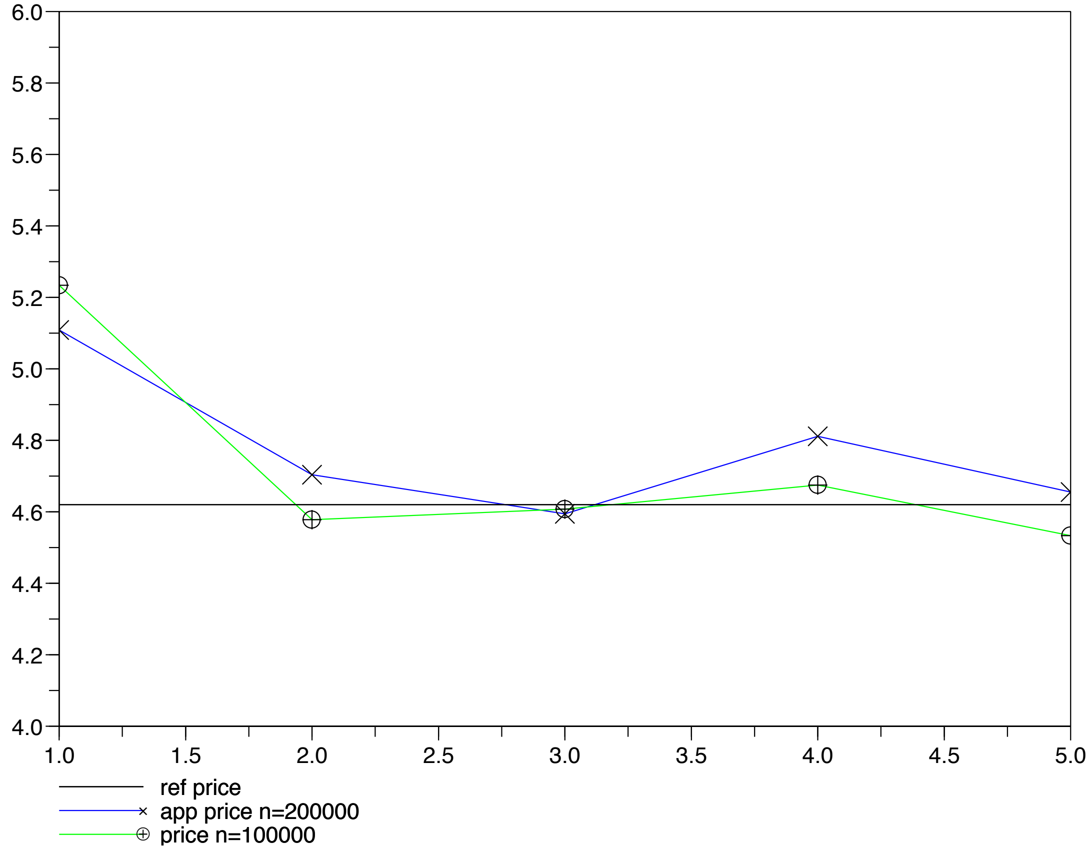
Algorithm parameters: $h_t = n^{-1/3}$, $h_x = 2an^{-1/3}$, $N = M = 100$, ...

Numerical price at time 0 w.r.t. iteration:



Similar Basket Call but in dimension 5

Numerical price w.r.t. iteration:



Pros and cons of the algorithm

Solution provided:

- Provides a global solution using Monte Carlo simulations (no system to invert) and without the inaccuracy of Monte Carlo methods.
- Provides a solution smooth w.r.t. space variables AND time variables.
- ⚠ Final accuracy depends heavily on $\mathcal{P}$.

Computational cost:

- Parallel computing
- Geometric convergence: not many iterations are needed.
- ⚠ Non local operators $\mathcal{P}$ may lead to larger computational costs.

Pros and cons of the algorithm (Cont'd)

Convergence:

- Convergence in weighted Sobolev norms (even if the solution is assumed to be $C^{1,2}$). So far, no pointwise convergence.
- Norms handle both the errors on the processes and the value functions.
- No confidence intervals.
- As usual, the kernel estimator perfomance depends on the dimension and the right bandwith is delicate to choose.
- Better choice of $\mathcal{P}$? Work In Progress (Wang's PhD thesis).

2.4 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

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

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

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

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

Implicit scheme

More closely related to the idea of discrete BSDE.

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

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.

2.4.1 Time discretization error

Define the measure of the quadratic 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 **[G', Lemor '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? $\gamma = ?$

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

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

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

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

The easy part: using the Euler scheme

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

2.4.3 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 '09]** 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 $f_{X_T}(.)$, then for any $p \geq 1$

$$\sup_{\mathbf{a}\in\mathbb{R}} \mathbb{E}|\mathbf{1}_{\mathbf{X_T^N}<\mathbf{a}} - \mathbf{1}_{\mathbf{X_T}<\mathbf{a}}| \leq \mathbf{3}\left(|\mathbf{f_{X_T}}|_{\mathbb{L}_\infty} \|\mathbf{X_T^N} - \mathbf{X_T}\|_{\mathbb{L}_\mathbf{p}}\right)^{\frac{\mathbf{p}}{\mathbf{p+1}}}.$$

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]**.

2.4.4 The L_2-regularity of Z

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

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

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

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

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 $r \leq t$

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

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

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]**.

2.5 The case of irregular terminal function $\Phi(X_T)$ [G., Makhlouf '10, Geiss-Geiss-G. '10]

$\rightsquigarrow$ **New ideas about fractional smoothness**

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

2.5.1 The index α to measure the regularity

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

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

and define

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

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

The index α is also called **fractional regularity** (notion introduced by **[Geiss-Geiss '04]** ...).

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

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 $\mathbb{D}_{1,2}$

[Geiss, Geiss '04; Geiss, Hujo '07]

Interpolations between two Banach spaces E_0 and E_1 ($E_1 \subset E_0$) **[Bergh, Löfström '76]**.

- Define the K-functional by

$$K(\Phi, \lambda; E_0, E_1) = \inf\{\|\Phi^0\|_{E_0} + \lambda\|\Phi^1\|_{E_1} \text{ such that } \Phi = \Phi^0 + \Phi^1\}$$

  for $\Phi \in E_0$.

- For $\alpha \in (0, 1)$ and $p \in [1, \infty]$, the **interpolation space** $(E_0, E_1)_{\alpha,p}$ is the set of elements $\Phi \in E_0$ such that

$$|\Phi|_{(E_0,E_1)_{\alpha,p}} := \|\lambda^{-\alpha} K(\Phi, \lambda; E_0, E_1)\|_{\mathbb{L}_p((0,+\infty),\frac{d\lambda}{\lambda})} < \infty.$$

In the following, we mainly consider the case $p = \infty$ for which

$$|\Phi|_{(E_0,E_1)_{\alpha,\infty}} := \sup_{\lambda\in]0,1]} \lambda^{-\alpha} K(\Phi, \lambda; E_0, E_1) < \infty.$$

Specification in the case of scalar BM [Nualart '06]

Write $\gamma_1(dx) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx$ for the one-dimensional Gaussian measure.

A function $\Phi : \mathbb{R} \mapsto \mathbb{R}$ s.t. $\Phi \in \mathbb{L}_2(\gamma_1)$ can be decomposed through its Hermite/chaos decomposition:

$$\Phi = \sum_{k\geq 0} a_k H_k \qquad \text{with} \qquad H_k(x) = \frac{(-1)^k}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k}(e^{-\frac{x^2}{2}}).$$

Define $E_0 = \mathbb{L}_2(\gamma_1) = \{\Phi : \ s.t.\ \|\Phi\|^2_{E_0} := \|\Phi\|^2_{\mathbb{L}_2(\gamma_1)} = \sum_{k\geq 0} a_k^2 < \infty\}$,

$$E_1 = \mathbb{D}_{1,2}(\gamma_1) = \{\Phi : \|\Phi\|^2_{E_1} := \|\Phi\|^2_{\mathbb{L}_2(\gamma_1)} + \|\Phi'\|^2_{\mathbb{L}_2(\gamma_1)} = \sum_{k\geq 0}(1+k)a_k^2 < \infty\}.$$

Computations of the K-functional

We decompose $\Phi = \sum_k a_k H_k$ into $\Phi^0 + \Phi^1 = \sum_k b_k H_k + \sum_k (a_k - b_k) H_k$. Then

$$\|\Phi^0\|_{\mathbb{L}_2} + \lambda \|\Phi^1\|_{\mathbb{D}_{1,2}} = \Big(\sum_k b_k^2\Big)^{1/2} + \lambda\Big(\sum_k (1+k)(a_k - b_k)^2\Big)^{1/2}$$

$$\sim_{\sqrt{2}} \left(\sum_k (b_k^2 + \lambda^2(1+k)(a_k - b_k)^2)\right)^{1/2},$$

$$\inf_{\Phi=\Phi^0+\Phi^1} \|\Phi^0\|_{\mathbb{L}_2} + \lambda \|\Phi^1\|_{\mathbb{D}_{1,2}} \sim_{\sqrt{2}} \left(\sum_k a_k^2 \frac{\lambda^2(1+k)}{1+\lambda^2(1+k)}\right)^{1/2}.$$

Thus, $\boldsymbol{\Phi} \in (\mathbb{L}_\mathbf{2}, \mathbb{D}_\mathbf{1,2})_{\alpha,\infty}$ iif $\sup_{\lambda \in]\mathbf{0},\mathbf{1}]} \lambda^{-\mathbf{2}\alpha} \sum_\mathbf{k} \mathbf{a}_\mathbf{k}^\mathbf{2} \frac{\lambda^\mathbf{2}(\mathbf{1}+\mathbf{k})}{\mathbf{1}+\lambda^\mathbf{2}(\mathbf{1}+\mathbf{k})} < \infty.$

Characterisation of $\Phi(W_1) \in \mathbb{L}_{2,\alpha}$ in terms of the $(a_k)_k$

Using that the time-space Hermite polynomial $(\mathbf{t^{k/2} H_k(\frac{W_t}{\sqrt{t}})})_\mathbf{t}$ defines a martingale, we get that

$$M_t := \mathbb{E}(\Phi(W_1)|\mathcal{F}_t) = \mathbb{E}\Big(\sum_k a_k H_k(W_1)|\mathcal{F}_t\Big) = \sum_k a_k t^{k/2} H_k\Big(\frac{W_t}{\sqrt{t}}\Big).$$

Thus,

$$\begin{aligned}\mathbb{E}(\Phi(W_1) - \mathbb{E}(\Phi(W_1)|\mathcal{F}_t))^2 &= \mathbb{E}(M_1 - M_t)^2 \\ &= \mathbb{E}(M_1^2) - \mathbb{E}(M_t^2) \\ &= \sum_k a_k^2 - \sum_k a_k^2 t^k.\end{aligned}$$

Then

$$\mathbf{\Phi(W_1)} \in \mathbb{L}_{\mathbf{2},\alpha} \quad \text{iif} \quad \sup_{\mathbf{t}\in[\mathbf{0},\mathbf{1}[} \frac{\sum_\mathbf{k} \mathbf{a_k^2(1 - t^k)}}{\mathbf{(1 - t)}^\alpha} < \infty.$$

Corollary. There exist functions Φ such that $\Phi(W_1) \notin \bigcup_{\alpha\in]0,1]} \mathbb{L}_{2,\alpha}$.

Equivalent characterisations

Theorem (see [GH07]). For any $\alpha \in (0,1)$, one has

$$\Phi(W_1) \in \mathbb{L}_{2,\alpha} \Longleftrightarrow \Phi \in (\mathbb{L}_2, \mathbb{D}_{1,2})_{\alpha,\infty}.$$

Remark. However, the $\mathbb{L}_{2,\alpha}$-characterisation leads to more tractable computations on practical examples.

Proof of $\Longleftarrow$. One has to prove $\Phi(W_1) \in \mathbb{L}_{2,\alpha}$, i.e.
$\sup_{t\in[0,1[} (1-t)^{-\alpha} \sum_k a_k^2(1-t^k) < \infty$, or equivalently to
$$\sup_{t\in[0,1[} (1-t)^{1-\alpha} \sum_k a_k^2 k t^{k-1} < \infty.$$

Define n_t such that $1 - \frac{1}{n_t} \leq t \leq 1 - \frac{1}{n_t+1}$: then, one can check that $kt^{k-1} \leq \frac{c}{1-t}$ for $k \geq n_t$. It implies

$$
\begin{aligned}
\sup_{t\in[0,1[}(1-t)^{1-\alpha}\sum_k a_k^2 k t^{k-1} &= \sup_{t\in[0,1[}(1-t)^{1-\alpha}\Big(\sum_{k=0}^{n_t} a_k^2 k t^{k-1} + \sum_{k>n_t} a_k^2 k t^{k-1}\Big)\\
&\leq \sup_{t\in[0,1[}(1-t)^{1-\alpha}\Big(\sum_{k=0}^{n_t} a_k^2 k + \sum_{k>n_t} a_k^2 \frac{c}{(1-t)}\Big)\\
&\leq_c \sup_{t\in[0,1[}(1-t)^{-\alpha}\Big(\sum_{k\geq 0} a_k^2 \min((1+k)(1-t),1)\Big)\\
&\overset{1-t=\lambda^2}{\sim_c} \sup_{\lambda\in]0,1]}\lambda^{-2\alpha}\Big(\sum_{k\geq 0} a_k^2\frac{\lambda^2(k+1)}{1+\lambda^2(k+1)}\Big)<\infty
\end{aligned}
$$

since $\Phi \in (\mathbb{L}_2, \mathbb{D}_{1,2})_{\alpha,\infty}$. □

Proof of $\Longrightarrow$. See [GH07].

2.5.2 Equivalent estimates on u and its derivatives [GM10]

Now assume X is a general SDE, under uniform ellipticity.

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

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

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

Extra equivalence results

Theorem. Let $\alpha \in (0,1]$. Consider a function Φ bounded (or expontentially bounded).

Then the three following assertions are equivalent:

i) $\int_0^T (T-t)^{-1-\alpha}\mathbb{E}\left|\Phi(X_T) - \mathbb{E}(\Phi(X_T)|\mathcal{F}_t)\right|^2 dt < \infty$.

ii) $\int_0^T (T-t)^{-\alpha}\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 dt < \infty$.

iii) $\int_0^T (T-t)^{1-\alpha}\mathbb{E}\left|D^2 u(t,X_t)\right|^2 dt < \infty$.

2.5.3 Application to the L_2-regularity of Z-components

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

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

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

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 $\pi = \{t_k, k = 0 \dots N\}$,

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

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

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

with $\beta < \alpha$ yields an error as N^{-1} for the $\mathbb{L}_2$-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. We establish that the usual representation of Z^0 using Malliavin derivatives holds. But the situation is not so standard because in general $\int_0^T \mathbb{E}|a_r^0|^2 dr = \infty$ for $\Phi(X_T) \in \mathbb{L}_{2,\alpha}$.

However we can prove $\boldsymbol{\int_0^T |a_r^0|_{\mathbb{L}_2} dr < \infty}$, which allows the use of results on BSDEs in $\mathbb{L}_p$, from **[Briand, Delyon, Hu, Pardoux, Stoica '03]**.

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

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

$\Longrightarrow$

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

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

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

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

Corollary for numerical computations. Regarding the problem of approximating accurately the Z component, it is better to solve first the BSDE (y,z) (simple problem) and then solve the BSDE difference $(Y-y, Z-z)$.

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

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

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

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

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

with an index $\beta < \alpha$.

2.6 Extra asymptotic results for smooth data

Error expansion w.r.t. the number of time steps [G., Labart '07a]

Consider uniform time grids. 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. Inspired by the Malliavin calculus approach of **[Kohatsu-Higa '01]** .

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.

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

$\Longrightarrow$ **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 '97]** when f does not depend on z.

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.

  To compute the full 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.
  - Compute explicitely the transition operator from X_{t_i} to $X_{t_{i+1}}$ and then compute the integral of g w.r.t. $\mathbb{P}_{X_{t_{i+1}}|X_{t_i}}(dx)$?
  - Simulate the transition?

- How many regression functions to compute?

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

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

- In which points $X \in \mathbb{R}^d$? 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: $\mathcal{X}^{\mathbf{j}} = \{\mathbf{x}^{\mathbf{j}}_{\mathbf{m}} : \mathbf{1} \le \mathbf{m} \le \mathbf{M_j}\} = \operatorname{argmin} \mathbb{E}(\min_{\mathbf{l}} |\mathbf{X}_{\mathbf{t_j}} - \mathbf{x}^{\mathbf{j}}_{\mathbf{l}}|^{\mathbf{2}})$.

- Existence of stochastic algorithm to compute these points (Kohonen algorithm).
- Quite slow. Better to compute them off-line.
- Grid already known in the case of Gaussian r.v. for various dimensions and various number of points. **[see Gilles Pages website]**.
- Suitable for $\mathbb{L}_2$-approximations (and Lipschitz functions).
- 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, at the end.
- ☺ many (stochastic) computations are made off-line.
- ☹ require the pre-computations of quantified grids of weights.

First applied to BSDEs by **[Chevance '97]**. For RBSDEs (with f independent of z), see **[Bally, Pages '03]** . Rates of convergence avalaible.

3.2 Representation of conditional expectations using Malliavin calculus

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

⚠ Requires the extra knowledge about the joint distribution of (Signal, Response).

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

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

Corollary. $\mathbb{E}(G|F=x)$ can be empirically evaluated using the sample $(F_i)_i$ far from x!!

- Actually, we look for $H(F, g(G)) = g(G)\tilde{H}(F, G)$.
  Representation with factorization not so immediate to obtain (possible for SDE).

- The H are obtained using Malliavin calculus, or a direct integration by parts when densities are known.

  For instance, if $F = W_{t_k}$ and $G = W_{t_{k+1}}$ in dimension 1, then one can take

$$\tilde{H}(F, G) = \frac{W_{t_k}}{t_k} - \frac{W_{t_{k+1}} - W_{t_k}}{t_{k+1} - t_k}.$$

- In practice, large variances (because $\sup_k(t_{k+1} - t_k) \to 0$) $\rightsquigarrow$ needs for variance reduction techniques (see **[Bouchard, Ekeland, Touzi '04]**).

- For non trivial dynamics (general SDE), the computational time needed to simulate H may be very large (Skorohod integrals to evaluate).

- Using the Riesz tranform **[Malliavin, Thalmaier '06]**, we only need one integration by parts, but the weights do not belong to $\mathbb{L}_2$!

  Localization techniques developed by **[Kohatsu-Higa and Yasuda '09]**.

- In any case, this approach requires a non degeneracy condition (ellipticity).

- For BSDEs, available rates of convergence w.r.t. N and M **[Bouchard, Touzi '04]** using N **independent set of simulated paths**.

- What happens if we use one set of paths?

- Efficiency compared to Quantization approach?

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 (or signal). Y is the response.

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 '10]** using weighted Sobolev space estimates).

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

3.3.2 Projection on a set of functions

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

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

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

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 (one can only simulate X).

- 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 large instabilities (ill-conditioned matrix) to solve this least-squares problem **[Golub, Van Loan '96]**. Recommended to use SVD.

- 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))? $\leadsto$ **Non asymptotic results...**

3.3.3 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 $m(.)$?

$$\oplus_{k\geq 0}\mathcal{P}_k(X) = \mathbb{L}_2(X)?$$

This is implicitely assumed in Longstaff-Schwartz algorithm for American options [LS01].

But **this is false in general**.

Counter-exemple (see Feller's book)

Take $X = \exp(W_1)$. Then $\sin(2\pi \log(X))$ is in $\mathbb{L}_2$ but is orthogonal to any polynomials!!

$$\mathrm{Proj}^{\perp}_{\mathcal{P}_{\mathbf{k}}(\mathbf{X})}[\sin(\mathbf{2}\pi \log(\mathbf{X}))] = \mathbf{0}, \qquad \forall \mathbf{k} \geq \mathbf{0}.$$

► In fact, the expected property is related to the moment problem:

is a r.v. characterized by its polynomial moments?

A sufficient condion: if for some $a > 0$ one has $\mathbb{E}(e^{a|X|}) < \infty$, the polynomials are dense in $\mathbb{L}_2$-functions.

In the good cases, convergence rates? some results by **[Guo]**, in the context of spectral methods for PDEs. But available for very smooth $m(.)$ (too smooth for BSDEs frameworks).

3.3.4 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 $\mu = \frac{1}{M}\sum_{m=1}^{M} \delta_{X^m}(.)$.

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

- 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.

  Theoretically, 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 $(X_1, Y_1), \cdots, (X_M, Y_M)$.
- $\sigma^2 = \sup_x \text{Var}(Y|X = x) < \infty$
- $F_M = \text{Span}(f_1, ...f_{K_M})$ a linear vector space of dimension K_M, which may depend on the data!

Notations: $|f|^2_M = \frac{1}{M}\sum_{i=1}^M f^2(X_i)$. Write μ^M for the empirical measure associated to $(X_1, \cdots, X_M)$.

$$\tilde{m}_M(.) = \arg\min_{f \in F_M} \frac{1}{M}\sum_{i=1}^M |f(X_i) - Y_i|^2.$$

Theorem. $\mathbb{L}_2(\mu^M)$-error: $\mathbb{E}(|\tilde{\mathbf{m}}_{\mathbf{M}} - \mathbf{m}|^2_{\mathbf{M}} | \mathbf{X_1}, \cdots, \mathbf{X_M}) \le \sigma^2 \frac{\mathbf{K_M}}{\mathbf{M}} + \min_{\mathbf{f} \in \mathbf{F_M}} |\mathbf{f} - \mathbf{m}|^2_{\mathbf{M}}$.

Proof

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

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

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

$$\mathbf{\tilde{m}_M}(.) = \sum_{\mathbf{j}} \alpha_{\mathbf{j}}\mathbf{f_j}(.) \quad \text{with} \quad \alpha_{\mathbf{j}} = \frac{\mathbf{1}}{\mathbf{M}}\sum_{\mathbf{i}} \mathbf{f_j}(\mathbf{X_i})\mathbf{Y_i}.$$

Lemma. Denote $\mathbb{E}^*(.) = \mathbb{E}(.|X_1, \cdots, X_M)$. Then $\mathbb{E}^*(\tilde{m}_M(.))$ is the least-squares solution of $\arg\min_{f\in F_M} \frac{1}{M}\sum_{i=1}^M |f(X_i) - m(X_i)|^2 = \arg\min_{f\in F_M} |f - m|_M^2$.

Proof.

- The above least-squares solution is given by $\sum_j \alpha_j^* f_j(.)$ with $\alpha_j^* = \frac{1}{M}\sum_i f_j(X_i)m(X_i)$.
- As a conditional expectation, $\mathbb{E}^*(\tilde{m}_M(.)) = \sum_j \mathbb{E}^*(\alpha_j)f_j(.)$.

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

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

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

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

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

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

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

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

$$\Longrightarrow \mathbb{E}^*|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 \le \sigma^2\sum_j \frac{1}{M^2}\sum_i f_j^2(X_i) = \sigma^2\frac{K_M}{M}.$$

Corollary. If in addition the linear space $F_M = \text{Span}(f_1, ... f_{K_M})$ does not depend on the data $(X_i)_{1 \le i \le M}$, then

$$\mathbb{E}(|\tilde{m}_M - m|_M^2) \le \underbrace{\sigma^2 \frac{K_M}{M}}_{\text{variance term}} + \underbrace{\min_{f \in F_M} |f - m|^2_{\mathbb{L}_2(\mu)}}_{\text{bias term}}.$$

Proof. Follows from $\mathbb{E}(\min_{f \in F_M} |f - m|_M^2) \le \min_{f \in F_M} \mathbb{E}(|f - m|_M^2) = \min_{f \in F_M} |f - m|^2_{\mathbb{L}_2(\mu)}$.

Next step: estimates on $\mathbb{L}_2(\mu)$ instead of $\mathbb{L}_2(\mu^M)$

i.e. replace an empirical mean by its true mean, to get estimates under the true law.

How far is the empirical mean $\frac{1}{M}\sum_{i=1}^M f(X_i)$ from its true mean $\mathbb{E}(f(X))$, whatever the function $f(.) = |\tilde{m}_M(.) - m(.)|^2$ is?

$\rightsquigarrow$ Related to techniques from **uniform law of large numbers**. **[Van Der Vaart, Wellner '96; Gyorfi, Kohler, Krzyzak, Walk '02; . . .]**.

Uniform law of large numbers

Consider $(Z_1, \cdots, Z_M)$ a i.i.d. sample of size M.

For $\mathcal{F} \subset \{f : \mathbb{R}^d \mapsto [0, B]\}$, one would need to quantifty

$$\mathbb{P}\big(\exists f \in \mathcal{F} : |\frac{1}{M}\sum_{i=1}^{M} f(Z_i) - \mathbb{E}f(Z)| > \epsilon\big)$$

as a function of ϵ and M?

Application: it enables to replace an empirical mean by its expectation, uniformly in the class of functions $\mathcal{F}$, up to error ϵ with high probability (explicitely quantified).

Other application: by Borel-Cantelli lemma, may lead to uniform laws of large numbers:

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

as $M \to \infty$.

Basic computations

▶ **If the cardinality of $\mathcal{F}$ is finite.** Then

$$\mathbb{P}(\exists \mathbf{f} \in \mathcal{F} : |\frac{\mathbf{1}}{\mathbf{M}}\sum_{\mathbf{i=1}}^{\mathbf{M}} \mathbf{f}(\mathbf{Z_i}) - \mathbb{E}\mathbf{f}(\mathbf{Z})| > \epsilon) \leq |\mathcal{F}| \sup_{f\in\mathcal{F}} \mathbb{P}\Big(|\frac{1}{M}\sum_{i=1}^{M} f(Z_i) - \mathbb{E}f(Z)| > \epsilon\Big)$$

$$\leq \mathbf{2}|\mathcal{F}| \exp\big(-\frac{\mathbf{2M}\epsilon^{\mathbf{2}}}{\mathbf{B^2}}\big)$$

by Hoeffding inequality (remind that $f(.) \in [0, B]$).

▶ **If the cardinality of $\mathcal{F}$ is infinite.** Suppose that $\mathcal{F}$ can be finitely ϵ-covered w.r.t. $\|.\|_{\mathbb{L}_\infty}$: there exists a finite set $\mathcal{F}_{\epsilon,\infty} = \{f_j : 1 \leq j \leq \mathcal{N}_\infty(\epsilon, \mathcal{F})\} \subset \mathcal{F}$ such that for any $f \in \mathcal{F}$, there is $f_j \in \mathcal{F}_{\epsilon,\infty}$ s.t. $|f - f_j|_{\mathbb{L}_\infty} \leq \epsilon$.

Simple example: $\mathcal{F} := \{f = \sum_{k=1}^{K} \alpha_k \mathbf{1}_{\mathcal{C}_k}$ with $\alpha_k \in [0, B]\}$.

Then

$$\mathbb{P}(\exists \mathbf{f} \in \mathcal{F} : |\frac{\mathbf{1}}{\mathbf{M}}\sum_{\mathbf{i=1}}^{\mathbf{M}} \mathbf{f}(\mathbf{Z_i}) - \mathbb{E}\mathbf{f}(\mathbf{Z})| > \epsilon) \leq \mathbf{2}|\mathcal{F}_{\frac{\epsilon}{\mathbf{3}},\infty}| \exp\big(-\frac{\mathbf{2M}(\epsilon/\mathbf{3})^{\mathbf{2}}}{\mathbf{B^2}}\big).$$

ϵ-cover of $\mathcal{F}$ w.r.t. $\mathbb{L}_p$-norms

Definition. For a class of functions $\mathcal{F}$ and a given empirical measure μ^M associated to M points $Z_{1:M} = (Z_1, \cdots, Z_M)$, we define a **$\epsilon$-cover of $\mathcal{F}$** w.r.t. $\mathbb{L}_1(\mu^M)$ by a **collection $(f_1, \cdots, f_M)$ in $\mathcal{F}$** such that for any $f \in \mathcal{F}$, there is a $j \in \{1, \cdots, N\}$ s.t. $|f - f_j|_{\mathbb{L}_1(\mu^M)} < \epsilon$.

Set $\mathcal{N}_1(\epsilon, \mathcal{F}, \mathbf{Z_{1:M}})$=the smallest size N of ϵ-cover of $\mathcal{F}$ w.r.t. $\mathbb{L}_1(\mu_M)$.

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

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

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

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

Remark. ☺ These estimates are distribution-free.

Applications to the $L_2(\mu)$-estimates of the regression errors

Theorem. Assume $\sigma^2 = \sup_{x\in\mathbb{R}^d} \mathrm{Var}(Y|X=x) < \infty$ and $m(.) = \mathbb{E}(Y|X=.) \in \mathbb{L}_\infty$.

For a K_M-dimensional linear vector space $\mathcal{F}_M$, define

$$\tilde{m}_M(.) = \arg\min_{f\in F_M} \frac{1}{M}\sum_{i=1}^{M} |f(X_i) - Y_i|^2,$$

$$m_M(.) = -\|m\|_{\mathbb{L}_\infty} \vee \tilde{m}_M(.) \wedge \|m\|_{\mathbb{L}_\infty}.$$

Then, for any $\delta > 0$, one has

$$\mathbb{E}\Big(|\mathbf{m_M}(.) - \mathbf{m}(.)|^{\mathbf{2}}_{\mathbb{L}_{\mathbf{2}}(\mu)}\Big) \le \mathbf{c}_\delta \max(\sigma^{\mathbf{2}}, \|\mathbf{m}\|^{\mathbf{2}}_{\mathbb{L}_\infty}) \frac{(\mathbf{1} + \log(\mathbf{M}))}{\mathbf{M}} \mathbf{K_M} + (\mathbf{1}+\delta) \min_{\mathbf{f}\in\mathbf{F_M}} |\mathbf{f} - \mathbf{m}|^{\mathbf{2}}_{\mathbb{L}_{\mathbf{2}}(\mu)},$$

where c_δ is an (explicit) universal constant such that $c_\delta \to \infty$ as $\delta \to 0$.

Remarks.

- These estimates are distribution-free (provided an uniform bound on the conditional variance).
- The regression function $m(.)$ has to be bounded.

3.5 Extensions to dynamic programmation equations

Extensions ?

1. increasing number N of regression problems,
2. dependent regression problems,
3. unboundedness of the Z-process,
4. ...

References:

1. Bouchaud, Potters, Sestovic: *Hedged Monte Carlo: low variance derivative pricing with objective probabilities*, Physica A, 2001.
2. Egloff: *Monte Carlo algorithms for optimal stopping and statistical learning*, AAP, 2005.

   Discrete time optimal stopping for general Markov chains.

   Non asymptotic estimates w.r.t. K and the number of simulations M.

   But the number of discretization times N is fixed.

3. G., Lemor, Warin:
   (a) *A regression-based Monte Carlo method to solve backward stochastic differential equations*, AAP 2005.

   Brownian BSDEs.

   Non asymptotic estimates w.r.t. K and the number of time steps N, but with $M = \infty$.

   CLT w.r.t. M, for fixed K and N.

   (b) *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Bernoulli 2006.

   Generalized BSDEs.

   Non asymptotic estimates w.r.t. all the parameters 🙂 .

Numerical solution of BSDEs using empirical simulations

[G',Lemor,Warin '06]

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

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

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

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

$\leadsto$ **Small impact of the threshold R. And gives extra 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**:

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

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

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

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

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? **difficult** because dependent regression operators.

Robust error bounds

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

$$\max_{0\le k\le N}\mathbb{E}|Y_{t_k}^{N,R}-y_k^{N,R,M}(X_{t_k}^N)|^2+h\sum_{k=0}^{N-1}\mathbb{E}|Z_{t_k}^{N,R}-z_k^{N,R,M}(X_{t_k}^N)|^2$$

$$\le C\underbrace{\frac{C_\star^2\log(M)}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}\mathbb{E}(K_{l,k}^M)}_{\text{Monte Carlo error}}+Ch$$

$$+C\sum_{k=0}^{N-1}\Big\{\underbrace{\inf_\alpha\mathbb{E}|y_k^{N,R}(X_{t_k}^N)-\alpha\cdot p_{0,k}(X_{t_k}^N)|^2}_{\text{quadratic approximation error on } Y_{t_k}^{N,R}}+\sum_{l=1}^{q}\underbrace{\inf_\alpha\mathbb{E}|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N)-\alpha\cdot p_{l,k}(X_{t_k}^N)|^2}_{\text{quadratic approximation error on } Z_{l,t_k}^{N,R}}\Big\}$$

$+$ exponentially small term.

The exponentially small term is equal

$$
\begin{aligned}
C\frac{C_\star^2}{h}\sum_{k=0}^{N-1}\Big\{&\mathbb{E}\Big(K_{0,k}^M \exp(-\frac{Mh^3}{72C_\star^2 K_{0,k}^M})\exp(CK_{0,k+1}\log\frac{C\ C_\star(K_{0,k}^M)^{\frac{1}{2}}}{h^{\frac{3}{2}}})\Big)\\
&+h\mathbb{E}\Big(K_{l,k}^M \exp(-\frac{Mh^2}{72C_\star^2 R_0^2 K_{l,k}^M})\exp(CK_{0,k+1}\log\frac{C\ C_\star R_0(K_{l,k}^M)^{\frac{1}{2}}}{h})\Big)\\
&+\exp(CK_{0,k}\log\frac{C\ C_\star}{h^{\frac{3}{2}}})\exp(-\frac{Mh^3}{72C_\star^2})\Big\}.
\end{aligned}
$$

Due to dependent regression problems.

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**: removes the two first contributions in the exponentially small term $\rightsquigarrow$ Improved choice of parameters: $\mathcal{C} \sim \epsilon^{-\frac{1}{4+d}}$.

3.6 Numerical results

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)

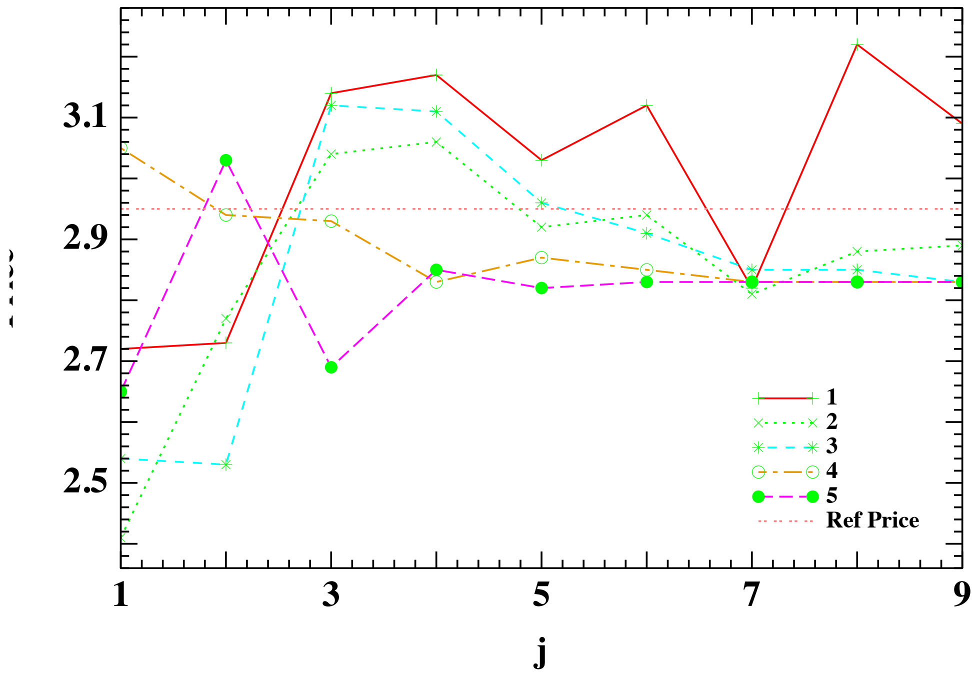
Polynomials of d variables with a maximal degree.

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)

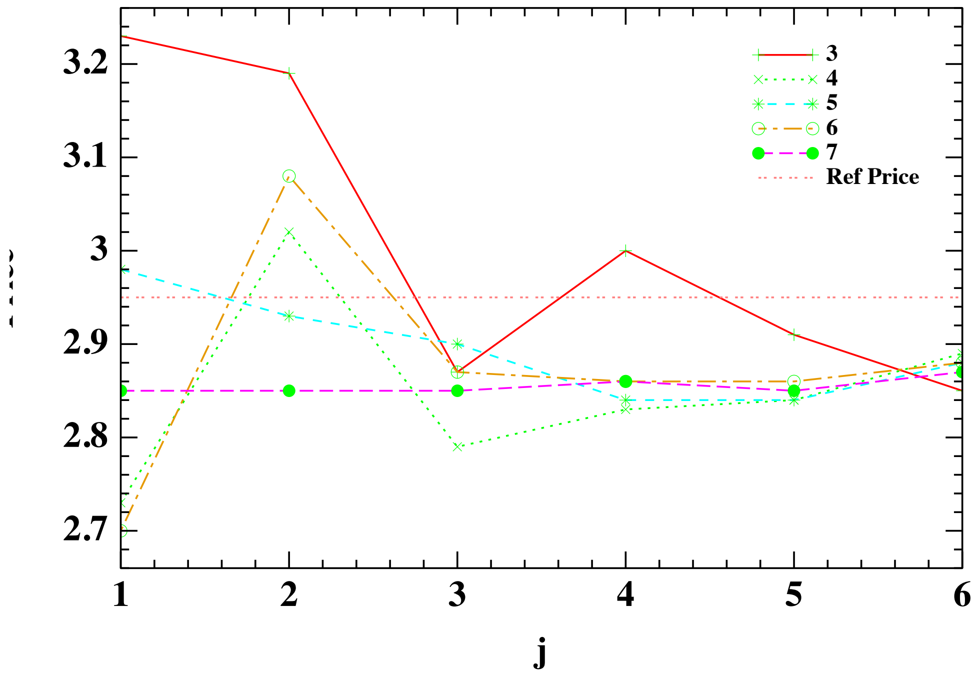
Table 2: Results for the calls combination using **GP**.

Large standard error $\rightsquigarrow$ GP not appropriate

Error convergence $N = \rho^j$, $\delta = h^{\frac{0.2+1}{2}}$ $(\beta = 0.2)$ $M \sim h^{-\alpha}$ $(\alpha^* = 3.4)$

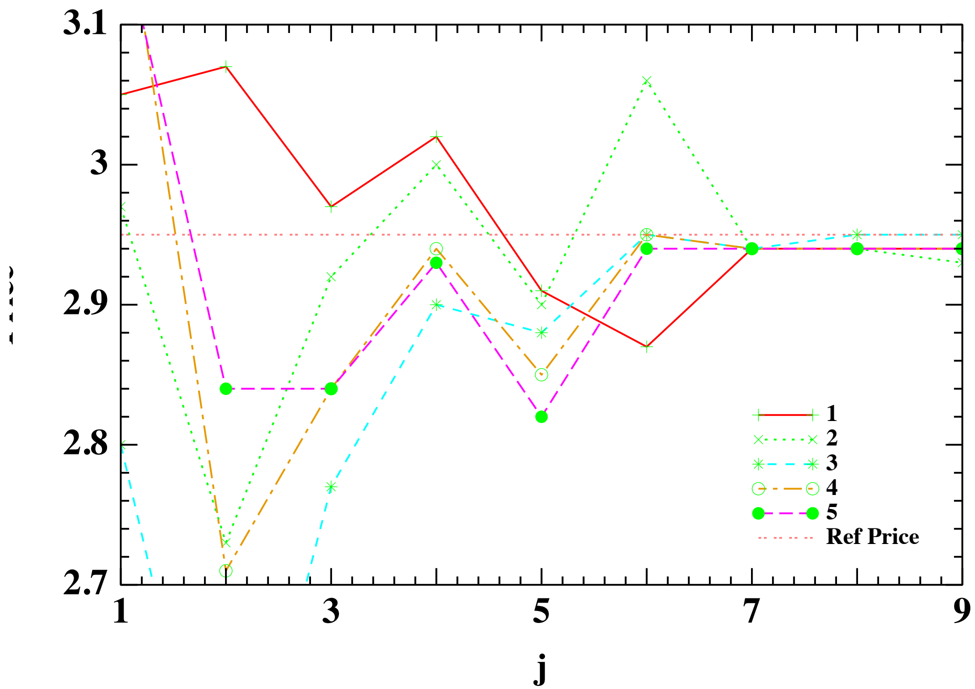


Error convergence $N = \rho^j$, $\delta = h^{\frac{1+1}{2}}$ $(\beta = 1)$ $M \sim h^{-\alpha}$ $(\alpha^* = 5)$



Optimal estimates?

Error convergence $N = \rho^j$, $\delta = h^{\frac{0.6+1}{2}}$ $(\beta = 0.6)$, HC(1,0) $M \sim h^{-\alpha}$ $(\alpha^* = 5.8)$



Ex.2 : Asian option

- Black Scholes model and $\Phi(\mathbf{S}) = (\frac{1}{T}\int_0^T S_t dt - K)_+$.
- Approximation of the integral: $S^N_{t_k} \longrightarrow \left(S^N_{t_k}, \frac{1}{k}\sum_{i=0}^{k-1} S^N_{t_i}(1+\mu\frac{h}{2}+\frac{\sigma}{2}\Delta W_{t_i})\right)^*$ **[Lapeyre and Temam '01]**.
- Problem in dimension 2.
- Parameters:

μ	σ	r	T	S_0	K
0.06	0.2	0.1	1	100	100

- Reference price: **7.04**.
- **HC** with $\delta = 1$, $D = [60, 200]^2$.

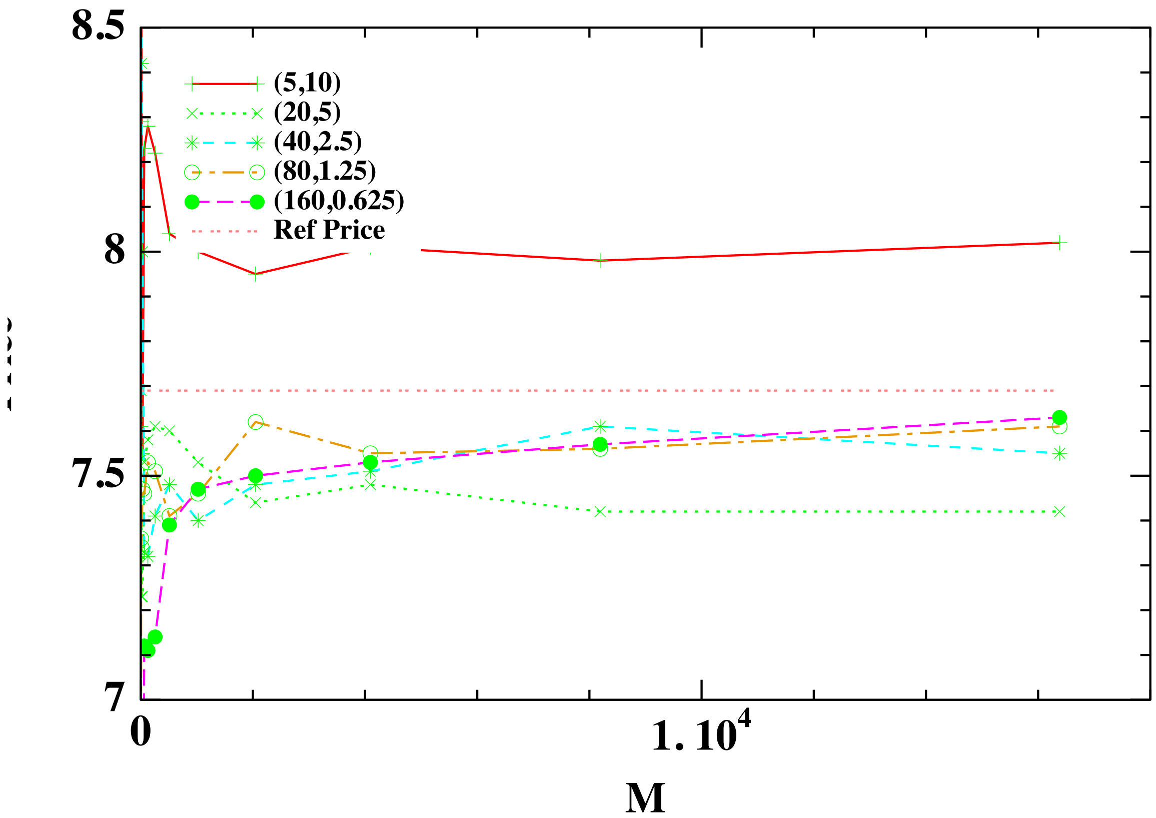
M	2	8	32	128	512	2048	8192	32768
$\overline{Y}^{N,I,M}_{t_0}$	2.26	0.90	4.49	6.68	6.15	6.88	6.99	7.02
$\sigma^{N,I,M}_{t_0}$	4.08	7.80	11.27	4.64	1.11	0.21	0.07	0.02

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

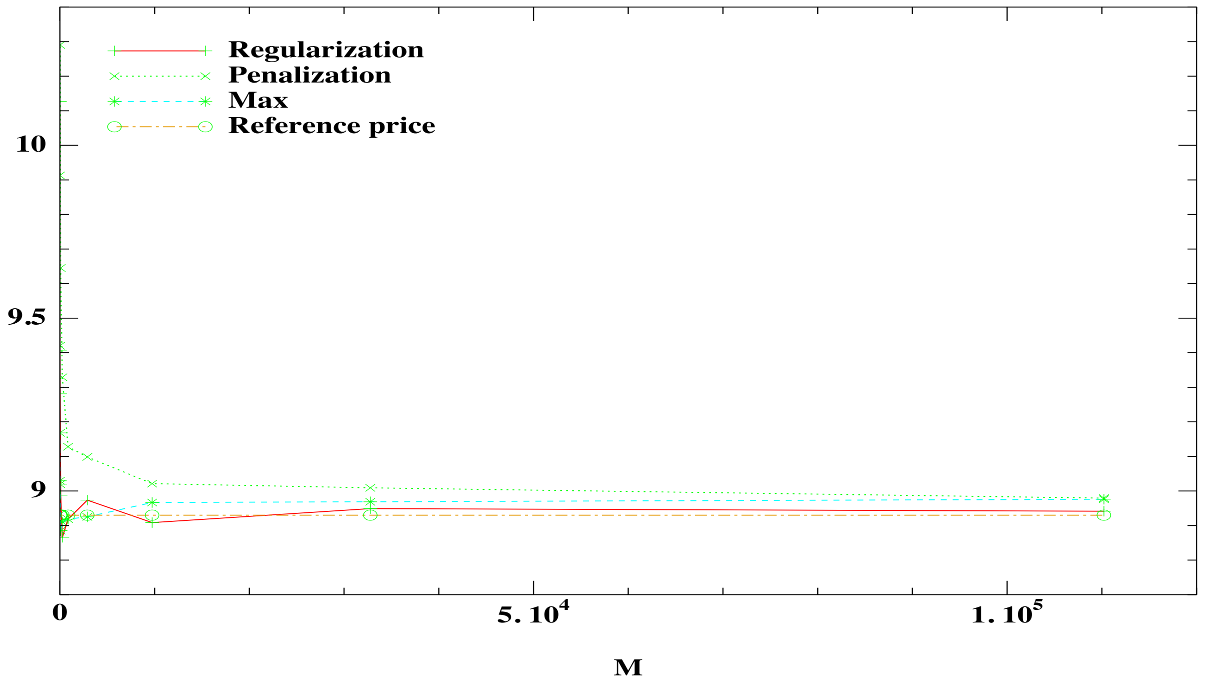
Obtained as a limit of standard BSDEs with driver
$f(s, S_s, Y_s, Z_s) + \rho_\lambda(Y_s - \Phi(s, S_s))(f(s, S_s, \Phi(s, S_s), V_s) + U_s)_-$ etc...
Upper approximation.

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.

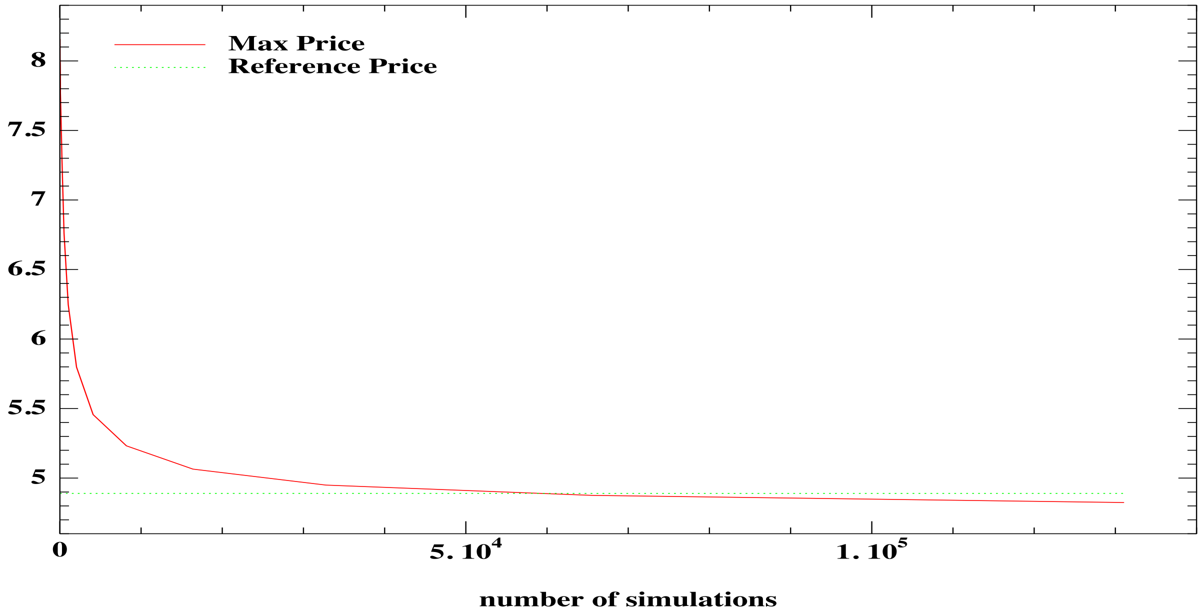
Regularization: $N = 32$, $\delta = 9$, $\lambda = 2$.

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

[ADEH99] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

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

[Bal97] V. Bally. Approximation scheme for solutions of BSDE. In *Backward stochastic differential equations (Paris, 1995–1996)*, volume 364 of *Pitman Res. Notes Math. Ser.*, pages 177–191. Longman, Harlow, 1997.

[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.

[BD07] C. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Processes and their Applications*, 117(12):1793–1823, 2007.

[BDH$^+$03] Ph. Briand, B. Delyon, Y. Hu, E. Pardoux, and L. Stoica. L^p solutions of backward stochastic differential equations. *Stochastic Processes and their Applications*, 108(1):109–129, 2003.

[BDM01] P. Briand, B. Delyon, and J. Mémin. Donsker-type theorem for BSDEs. *Electron. Comm. Probab.*, 6:1–14 (electronic), 2001.

[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.

[BET04] B. Bouchard, I. Ekeland, and N. Touzi. On the Malliavin approach to Monte Carlo approximation of conditional expectations. *Finance and Soochastics*, 8(1):45–71, 2004.

[BH06] P. Briand and Y. Hu. BSDE with quadratic growth and unbounded terminal value. *Probab. Theory Relat. Fields*, 136(4):604–618, 2006.

[BL76] J. Bergh and J. Löfström. *Interpolation spaces. An introduction.* Springer-Verlag, 1976.

[BL84] A. Bensoussan and J.L. Lions. *Impulse control and quasivariational inequalities. μ.* Gauthier-Villars, Montrouge, 1984. Translated from the French by J. M. Cole.

[BM01] V. Bally and A. Matoussi. Weak solutions for SPDEs and Backward Doubly Stochastic Differential Equations. *Journal of Theoretical Probability*, 14:125–164, 2001.

[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.

[CM10] D. Crisan and K. Manolarakis. Solving Backward Stochastic Differential Equations using the Cubature method. *Preprint*, 2010.

[CSTV07] P. Cheridito, H.M. Soner, N. Touzi, and N. Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Commun. Pure Appl. Math.*, 60(7):1081–1110, 2007.

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

[DP97] R.W.R. Darling and E. Pardoux. Backwards SDE with random terminal time and applications to semilinear elliptic PDE. *Ann. Probab.*, 25(3):1135–1159, 1997.

[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.

[EHM08] N. El Karoui, S. Hamadène, and A. Matoussi. Backward stochastic differential equations and applications. In R. Carmona, editor, *Indifference pricing: theory and applications*, chapter 8, pages 267–320. Springer-Verlag, 2008.

[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.

[eMJ07] S. Hamadène et M. Jeanblanc. On the starting and stopping problem. *Mathematics of Operations Research*, 32:182–192, 2007.

[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.

[GG04] C. Geiss and S. Geiss. On approximation of a class of stochastic integrals and interpolation. *Stoch. Stoch. Rep.*, 76(4):339–362, 2004.

[GGG09] C. Geiss, S. Geiss, and E. Gobet. Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition. *In preparation*, 2009.

[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. A note on the rate of convergence of the operator P^k. Technical report, HAL preprint server, http://hal.archives-ouvertes.fr/hal-00373349/fr/, 2009.

[GL10] E. Gobet and C. Labart. Solving BSDE with adaptive controle variate. *SIAM Numerical Analysis*, 48(1):257–277, 2010.

[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.

[GM05] E. Gobet and S. Maire. Sequential control variates for functionals of Markov processes. *SIAM Journal on Numerical Analysis*, 43(3):1256–1275, 2005.

[GM10] E. Gobet and A. Makhlouf. L2-time regularity of BSDEs with irregular terminal functions. *Stochastic Processes and their Applications*, 120:1105–1132, 2010.

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

[Ḧ92] W. Härdle. *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.

[KHY09] A. Kohatsu-Higa and K. Yasuda. Estimating multidimensional density functions using the Malliavin-Thalmaier formula. *SIAM J. Numer. Anal.*, 47:1546–1575, 2009.

[Kob00] M. Kobylanski. Backward stochastic differential equations and partial differential equations with quadratic growth. *The Annals of Probability*, 28(2):558–602, 2000.

[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.

[LS97] J.P. Lepeltier and J. San Martin. Backward stochastic differential equations with continuous coefficient. *Stat. Probab. Lett.*, 32(4):425–430, 1997.

[LS98] J.P. Lepeltier and J. San Martin. Existence for BSDE with superlinear-quadratic coefficient. *Stochastics Stochastics Rep.*, 63(3-4):227–240, 1998.

[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.

[MT06] P. Malliavin and A. Thalmaier. *Stochastic calculus of variations in mathematical finance.* Springer-Verlag, Berlin, 2006.

[MY99] J. Ma and J. Yong. *Forward-Backward Stochastic Differential Equations.* Lecture Notes in Mathematics, 1702, Springer-Verlag, 1999. A course on stochastic processes.

[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.

[Par98] E. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 79–127. Birkhäuser Boston, Boston, MA, 1998.

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

[PP90] E. Pardoux and S.G. Peng. Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.

[PP92] E. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic partial differential equations and their applications, Proc. IFIP Int. Conf., Charlotte/NC (USA) 1991, Lect. Notes Control Inf. Sci. 176*, pages 200–217. Springer, 1992.

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

[VW96] A.W. Van Der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New-York., 1996.

[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.