Approximation of BSDEs using least-squares regression and Malliavin weights

Plamen Turkedjiev (turkedji@math.hu-berlin.de)

3rd July, 2012

Joint work with Prof. Emmanuel Gobet (École Polytechnique)

FBSDEs

- $T > 0$,
- W q-dimensional Brownian motion,
- $(\Omega, \mathscr{F}, \mathbb{P})$ filtered probability space with usual conditions, but filtration may be larger than that generated by W,
- $\xi \in \mathbf{L}^2(\mathscr{F}_T)$,

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s - (L_T - L_t)$$

where X is d-dimensional, $(t, x, y, z) \mapsto f(t, x, y, z)$ is Borel measurable. Typically,

- X is a jump-diffusion driven by W and a Poisson random measure, L is a martingale orthogonal to W, and $\xi = \Phi(X_T)$
- X is a diffusion driven by W, $L \equiv 0$, and $\xi = \Phi(X_{t_1}, \dots, X_T)$ or $\xi = \Phi(X_T, \int_0^T X_t dt)$

Local Lipschitz condition and Quadratic BSDEs

We consider time-local Lipschitz continuous driver:

$$|f(t,x,y,z) - f(t,x',y',z')| \le L_f(|x-x'| + \frac{|y-y'| + |z-z'|}{(T-t)^{(1-\theta)/2}})$$

MOTIVATION:
Assume X is a diffusion ($L \equiv 0$) and driver satisfies quadratic growth condition

$$\begin{aligned} |f(t,x,y,z)| &\le c(1+|y|+|z|^2) \\ |f(t,x,y,z) - f(t,x,y',z')| &\le c(1+|z|+|z'|)(|y-y'|+|z-z'|) \end{aligned}$$

and $x \mapsto \Phi(x)$ is Hölder continuous and bounded.
Then $|Z_t| \le L_f(T-t)^{-(1-\theta)/2}$ holds $\mathbb{P} \times dt$-a.e. for constants L_f and θ independent of t.

Locally Lipschitz can replace quadratic in this special problem!

Driver with exploding bound and variance reduction

For all $t > 0$ and $x \in \mathbb{R}^d$, there exists $\alpha \in (0, 1]$ and $C_f > 0$ such that

$$|f(t, x, 0, 0)| \leq \frac{C_f}{(T - t)^{1-\alpha}}$$

MOTIVATION:

$\xi = \Phi(X_T)$ and $x \mapsto \Phi(x)$ is α-Hölder continuous and bounded.
X is a diffusion process ($L \equiv 0$) so that $v_t(x) = \mathbb{E}[\Phi(X_T)|X_t = x]$ is smooth.
$f(t, x, y, z)$ uniformly Lipschitz continuous and unif. bounded at $(y, z) = (0, 0)$.

$(v_t(X_t), \nabla v_t(X_t)\sigma(t, X_t))$ solves BSDE with data $(\xi, 0)$. Suppose we can solve this BSDE! $|\nabla v_t(x)| \leq C(T - t)^{\alpha-1}$ standard from PDE theory.

$(Y_t - v_t(X_t), Z_t - \nabla v_t(X_t)\sigma(t, X_t))$ solves a BSDE with data $(0, f^0)$, where
$f^0(t, x, y, z) = f(t, x, y + v_t(x), z + \nabla v_t(x)\sigma(t, x))$.
This BSDE may be better behaved for simulation purposes.

Key property: discretizability of FBSDEs

Time-grid: $\pi = (0 = t_0 < \ldots < t_N = T)$.
Paritcularly important grid: for $\beta \in (0,1]$, π^β for which $t_i^\beta := T - T(1 - \frac{1}{N})^{1/\beta}$.

Theorem

If $\alpha = 1$, let $\beta = 1$; else let $\beta < \alpha$. Under the given assumptions, there exists a positive constant C, independent of N, such that

$$\max_{0 \le i \le N-1} \sup_{t_i^\beta \le t < t_{i+1}^\beta} \mathbb{E}|Y_t - Y_{t_i^\beta}|^2 + \sum_{i=0}^{N-1} \int_{t_i^\beta}^{t_{i+1}^\beta} \mathbb{E}|Z_t - Z_{t_i^\beta}|^2 dt \quad \le \quad CN^{-1}$$

We say that $O(N^{1/2})$ is the optimal rate of convergence for a discrete-time approximation of the BSDE.

Algorithm 1: Multistep dynamical programming

Let $\Delta_i = t_{i+1} - t_i$ and $\Delta W_i := W_{t_{i+1}} - W_{t_i}$.
Recurssively build approximation of the solution, starting at $i = N - 1$:

$$\begin{cases} \Delta_i Z_i & = \mathbb{E}_i[\Delta W_i^\top(\xi + \sum_{k=i+1}^{N-1} f(t_k, X_k, Y_{k+1}, Z_k)\Delta_k)], \\ Y_i & = \mathbb{E}_i[\xi + \sum_{k=i}^{N-1} f(t_k, X_k, Y_{k+1}, Z_k)\Delta_k], \\ Y_N & = \xi. \end{cases}$$

Consistency conditions for the time-grid:

$$\sup_{k<N} \frac{\Delta_k}{(T - t_k)^{1-\theta}} \to 0 \text{ as } N \to \infty, \quad \limsup_{N\to\infty} \sup_{k<N-1} \frac{\Delta_k}{\Delta_{k+1}} \leq \infty.$$

Theorem

For N sufficiently large, there exists a positive constant C independent of the time-grid such that

$$\max_{0\leq k\leq N-1} \mathbb{E}|Y_i - Y_{t_i^\beta}|^2 + \sum_{i=0}^{N-1} \mathbb{E}|Z_i - Z_{t_i^\beta}|^2 \Delta_i \leq CN^{-1}$$

Assumptions and properties

Markov structure Let $\xi = \Phi(X_N)$ and X be a Markov chain.This ensures $(Y_i, Z_i) = (y_i(X_i), z_i(X_i))$ for measurable (*unknown*) functions y_i and z_i.

Almost sure bounds Let $x \mapsto \Phi(x)$ be bounded. This ensures that $\exists C_y > 0$ such that,$\forall k$, $|Y_k| \leq C_y$ and $|Z_k| \leq \frac{C_y}{\sqrt{\Delta_k}}$ $\mathbb{P}$-almost surely.

Basis functions For each $0 \leq l \leq q$ and $0 \leq k \leq N-1$, take a finite number of functions $p_{l,k}(\cdot) = (p^i_{l,k})_{1 \leq i \leq K}$ such that $p_{l,k} : \mathbb{R}^d \to \mathbb{R}$ is deterministic and $\mathbb{E}[|p_{l,k}(X_k)|^2] < \infty$. Form basis of finite dimensional subspaces of $L_2(\mathscr{F}_{t_k})$.

Simulations Take M independent simulations of the Brownian increments ΔW and the explanatory Markov chain X. Denote these simulations by $(X^m_k)_{1 \leq m \leq M}$ and $(\Delta W^m_k)_{1 \leq m \leq M}$ respectively. Let $p^m_{l,k} := p_{l,k}(X^m_k)$.

Definition For $R > 0$, the truncated Brownian increment is defined by $[\Delta W_i]_R = -R\sqrt{\Delta_i} \vee \Delta W_i \wedge R\sqrt{\Delta_i}$.

Emprical regression algorithm

Set $y_N^{R,M}(\cdot) = \Phi(\cdot)$. Then, for $i < N$, compute coefficients

$$\alpha_{l,i}^M = \text{arg}\min_\alpha \frac{1}{M}\sum_{m=1}^M \Big| \frac{[\Delta W_{l,i}]_R}{\Delta_i}\Big(\Phi(X_N^m) + \sum_{k=i+1}^{N-1} f_i(y_{k+1}^{R,M}(X_{k+1}^m), z_k^{R,M}(X_k^m))\Delta_k\Big) - \alpha \cdot p_{l,k}^m\Big|^2$$

$$\alpha_{0,k}^M = \text{arg}\min_\alpha \frac{1}{M}\sum_{m=1}^M |\Phi(X_N^m) + \sum_{k=i+1}^{N-1} f_i(y_{k+1}^{R,M}(X_{k+1}^m), z_k^{R,M}(X_k^m))\Delta_k - \alpha \cdot p_{0,k}^m|^2.$$

The coefficients are not independent of one another!

Set

$$y_i^{R,M}(x) = -C_y \vee \alpha_{0,i}^M \cdot p_{0,i}(x) \wedge C_y,$$
$$z_{l,i}^{R,M}(x) = -\frac{C_y}{\sqrt{\Delta_i}} \vee \alpha_{l,i}^M \cdot p_{0,i}(x) \wedge \frac{C_y}{\sqrt{\Delta_i}}.$$

Key ingredient: concentration of measure inequalities

Needed, amongst other things, to deal with the lack of independence between regression coefficients. The following example comes from [Györfi et al. 2002, Theorem 11.2]. Benefit: the estimates are *distribution-free*.

Theorem

Let $\mathscr{F} \subset \{f : \mathbb{R}^d \to [-B, B]\}$ and $(Z_i)_{1 \le i \le n}$ be i.i.d. Then, for all $\varepsilon > 0$.

$$\mathbb{P}(\exists f \in \mathscr{F} : (\mathbb{E}[|f(Z)|^2])^{1/2} - 2(\frac{1}{n}\sum_{i=1}^{n} |f(Z_i)|^2)^{1/2} > \varepsilon)$$
$$\le \mathbb{E}[\mathscr{N}_2(\frac{\sqrt{2}}{24}\varepsilon, \mathscr{F}, Z_{1:n})] \exp\left(-\frac{n\varepsilon^2}{288B^2}\right)$$

Proposition

If $\mathscr{F}$ is in a K-dimensional vector space,

$$\mathscr{N}_2(\varepsilon, \mathscr{F}, z_{1:n}) \le 3\left(\frac{2eB^2}{\varepsilon^2}\log\left(\frac{3eB^2}{\varepsilon^2}\right)\right)^K$$

Error estimates

Norm: For function Ψ, define $\|\Psi\|_{k,M}^2 := \frac{1}{M}\sum_{m=1}^{M} |\Psi(X_k^m)|^2$.

Theorem

For N sufficiently large, there exists a possitive constant C independent of the time-grid, M and the basis functions such that

$$\sum_{k=0}^{N-1} \Big\{ \mathbb{E}[\|y_k - y_k^{R,M}\|_{k,M}^2] + \mathbb{E}[\|z_k - z_k^{R,M}\|_{k,M}^2] \Big\} \Delta_k$$

$$\leq C \sum_{k=0}^{N-1} \left\{ \min_{\alpha} \mathbb{E}|y_k^R(X_k) - \alpha \cdot p_{0,k}|^2 \Delta_k + \sum_{l=1}^{q} \min_{\alpha} \mathbb{E}|z_{l,k}^R(X_k) - \alpha \cdot p_{l,k}|^2 \Delta_k \right\}$$

$$+ C \sum_{k=0}^{N-1} \left\{ \frac{KN\Delta_k}{M} + \frac{KNR^2}{M} \right\} + CN^{-\theta_{conv}}$$

$$+ CNKR^2 3^{CN} \sum_{k=0}^{N-1} \exp\left(- \frac{CMR^{-2}\Delta_k}{KN^{1+\theta_{conv}}} \right) \prod_{i=k}^{N-1} \left(\frac{CKR^2}{N^{-\theta_{conv}} \Delta_i \Delta_k} \right)^{CK} + \dots$$

Complexity analysis

- Aim: reduce error to $O(N^{-2\theta_{conv}})$.
- Assume $y_i \in C_b^{\kappa+1+\eta}$, $z_i \in C_b^{\kappa+\eta}$.
- Local polynomials on disjoint hypercubes, degree $\kappa+1$ for Y and κ for Z.
  Bias approximation: $O(N^{-2\theta_{conv}})$ if $\delta_z = cN^{-\frac{\theta_{conv}}{\kappa+\eta}}$.
  $\Rightarrow K = cN^{d\frac{\theta_{conv}}{\kappa+\eta}}$ up to log terms.
- *Large deviation* terms: $M = cKN^{2+2\theta_{conv}} = cN^{2+2\theta_{conv}+2d\frac{\theta_{conv}}{\kappa+\eta}}$ up to log terms.
- Computational work $\mathscr{C} = cMN = cN^{3+2\theta_{conv}+2d\frac{\theta_{conv}}{\kappa+\eta}}$ up to log terms.
  $\Rightarrow N^{-2\theta_{conv}} \leq c\mathscr{C}^{\frac{-1}{2(1+\frac{3}{2\theta_{conv}}+\frac{d}{\kappa+\eta})}}$.
- ODP scheme with $\theta_{conv} = 1/2$: $N^{-1} \leq c\mathscr{C}^{\frac{-1}{2(4+\frac{2d}{\kappa+1+\eta})}}$.
  $\Rightarrow$ if $\kappa+\eta > 1$, MDP has better performance.

BSDE numerics: Implicit vs explicit

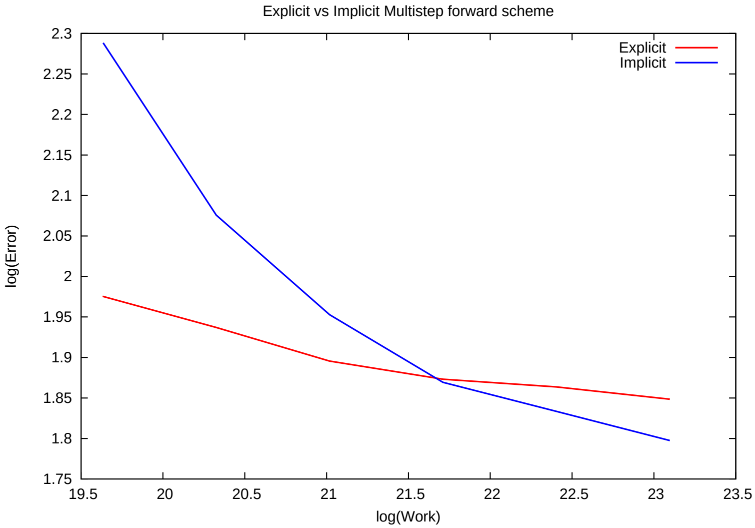
$q = 3$; $f(z) = 1.5|z^{(1)}|$; $\Phi(x) = (x^{(3)} - 100)^+$; $X_t^{(j)} = 100e^{(\sigma W_t)^{(j)}}$ with

$$\sigma = \begin{pmatrix} \sigma_1\sqrt{1-\rho^2} & 0 & \sigma_1\rho \\ 0 & \sigma_2\sqrt{1-\rho^2} & \sigma_2\rho \\ 0 & 0 & \sigma_3 \end{pmatrix}, \quad \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \rho \end{pmatrix} = \begin{pmatrix} 0.01 \\ 0.05 \\ 0.03 \\ 0.1 \end{pmatrix}$$

$N = 16$; Basis $\prod_{i=0}^{q} g_i(\ln(x_i))$ for g_i Hermite polynomials with $\sum_i deg(g_i) \le 3$.
Explicit solution:

$$Y_t = BlackScholesCall(t, X_t; \sigma_3, 100),$$
$$Z_t = (0, 0, BlackScholesHedge(t, X_t; \sigma_3, 100)).$$

BSDE numerics



Representation theorem due to Ma/Zhang

X is a diffusion: $dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$.
Also need gradient process and it's inverse:

$$d\nabla X_t = b_x(t, X_t)\nabla X_t dt + \sigma(t, X_t)\nabla X_t dW_t,$$
$$d\nabla X_t^{-1} = (-b_x(t, X_t) - \sigma_x(t, X_t)^2)\nabla X_t^{-1}dt + \sigma_x(t, X_t)\nabla X_t^{-1}dW_t.$$

Representation theorem due to Ma/Zhang for Z:

$$Z_t = \mathbb{E}_t[\xi H_T^t + \int_t^T f(r, X_r, Y_r, Z_r)H_r^t dr]$$

where $(r-t)H_r^t = (\int_r^t [\sigma^{-1}(s, X_s)\nabla X_s \nabla X_t^{-1}\sigma(t, X_t)]^\top dW_s)^\top$.

H_r^t are the Malliavin weights; the representation formula is derived by means of Malliavin's calculus, but remains true in the Lipschitz case, even though the BSDE is not Mallivin differentiable.

Algorithm 2: Malliavin weights

Let $(t_k - t_i)H_k^i = (\sum_{j=i}^{k-1} \sigma^{-1}(t_j, X_j)\nabla X_j \nabla X_i^{-1}\sigma(t_i, X_i)]^\top \Delta W_j)^\top$.
Recursively build the approximation starting at $i = N-1$:

$$\begin{cases} Z_i & = \mathbb{E}_i[\xi H_N^i + \sum_{k=i+1}^{N-1} f(t_k, X_k, Y_{k+1}, Z_k)H_k^i \Delta_k], \\ Y_i & = \mathbb{E}_i[\xi + \sum_{k=i}^{N-1} f(t_k, X_k, Y_{k+1}, Z_k)\Delta_k], \\ Y_N & = \xi. \end{cases}$$

Constraint on the time-grid: $\limsup_{N\to\infty} \sup_{i<N} \frac{\Delta_{i+1}}{\Delta_i} < \infty$.
Recall the special time-grid π^β:

Theorem

For sufficiently high N, there exists a positive constant C independent of the time-grid such that

$$\max_{0\le k\le N-1} \mathbb{E}|Y_i - Y_{t_i^\beta}|^2 + \sum_{i=0}^{N-1} \mathbb{E}|Z_i - Z_{t_i^\beta}|^2 \Delta_i \le CN^{-1}$$

Projection estimates

Approximate conditional expectation $\mathbb{E}_i$ by projection on finite subspace of $L_2(\mathscr{F}_{t_i})$:

$$\begin{cases} \hat{Z}_{l,i} & = \operatorname{arg\,min}_{\alpha\cdot p_{l,i}(X_i)} \\ & \quad \mathbb{E}[|\Phi(X_N)H^i_{l,N} + \sum_{k=i+1}^{N-1} f(t_k, X_k, \hat{Y}_{k+1}, \hat{Z}_k)H^i_{l,k}\Delta_k - \alpha\cdot p_{l,i}(X_i)|^2], \\ \hat{Y}_i & = \operatorname{arg\,inf}_{\alpha\cdot p_{0,i}(X_i)} \\ & \quad \mathbb{E}[|\Phi(X_N) + \sum_{k=i}^{N-1} f(t_k, X_k, \hat{Y}_{k+1}, \hat{Z}_k)\Delta_k - \alpha\cdot p_{0,i}(X_i)|^2], \\ \hat{Y}_N & = \Phi(X_N). \end{cases}$$

Projection estimates

Theorem

There exists positive constant C independent of the time-grid such that

$$
\begin{aligned}
\mathbb{E}|Y_i - \hat{Y}_i|^2 \leq 2\mathbb{E}|Y_i - \mathscr{P}_i^Y Y_i|^2 \\
+ C \sum_{k=i}^{N-1} \frac{\{\mathbb{E}|Y_{k+1} - \mathscr{P}_{k+1}^Y Y_{k+1}|^2 + \mathbb{E}|Z_k - \mathscr{P}_k^Z Z_k|^2\}\Delta_k}{(T - t_k)^{1-\theta}} \\
\mathbb{E}|Z_i - \hat{Z}_i|^2 \leq 2\mathbb{E}|Z_i - \mathscr{P}_i^Z Z_i|^2 \\
+ C \sum_{k=i}^{N-1} \frac{\{\mathbb{E}|Y_{k+1} - \mathscr{P}_{k+1}^Y Y_{k+1}|^2 + \mathbb{E}|Z_k - \mathscr{P}_k^Z Z_k|^2\}\Delta_k}{(T - t_k)^{1-\theta}}
\end{aligned}
$$

Almost sure bounds

There exist positive constants C_y and C_z independent of the time-grid such that, $\forall i$,

$$|Y_i| \leq C_y \text{ and } |Z_i| \leq \frac{C_z}{\sqrt{T - t_i}} \quad \mathbb{P}\text{-almost surely}$$

Finally...

Thank You For Your Attention!

References

[GT11] E. Gobet, T.
Approximation for discrete BSDE using least-squares regression.
http://hal.archives-ouvertes.fr/aut/turkedjiev/

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

[CD11] D. Crisan, F. Delarue.
Sharp gradient bounds for solutions of semi-linear PDEs.
http://hal.archive-ouvert.fr/hal-00599543/fr, 2011.

[DG06] F. Delarue, G. Guatteri.
Weak existence and uniqueness for forward-backward SDEs.
Stochastic processes and their applications, 116(12):1712–1742, 2006.

[GGG11] C. Geiss, S. Geiss, and E. Gobet.
Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition.
To appear in Stochastic Processes and their Applications, 2011.

References

[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.

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

[GM10] E. Gobet and A. Makhlouf.
L_2-time regularity of BSDEs with irregular terminal functions.
Stochastic Processes and their Applications, 120:1105–1132, 2010.

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

References

[MZ02] J. Ma, J. Zhang.
Representation theorems for BSDEs.
Annals of Applied Probability, 12(4):1390–1418, 2002.

[Mo10] T. Moseler.
A Picard-type iteration for BSDEs: Convergence and importance sampling.
PhD thesis.