Adaptive Importance Sampling Monte Carlo Least-Squares regression algorithm for BSDEs

Plamen Turkedjiev
King's College London

With Emmanuel Gobet

Outline

- Markov BSDE setting
- BSDE algorithm without importance sampling
- Importance sampling scheme
- New error estimates for least-squares regression
- Error estimates for BSDE algorithms
- Key technique: randomized initialization of simulations
- Numerical examples

Setting: Markov Backward Stochastic Differential Equations

BSDE

(X, Y, Z) are $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q$-valued processes

$$X_t = X_0 + \int_0^t b(s, X_s)dt + \int_0^t \sigma(s, X_s)dW_s,$$

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

Markov property: $(Y_t, Z_t) = (Y(t, X_t), Z(t, X_t))$ where $(Y(t, x), Z(t, x))$ deterministic.

Everything is “nice”: b, σ, g Lipschitz in x; f Lipschitz in (x, y, z).
σ uniformly elliptic. g and f bounded.

For error estimates: f does not depend on z.

Numerical method without importance sampling

Backwards-in-time recursion

$$Y(t, X_t) = \mathbb{E}[g(X_T) + \int_t^T f(s, X_s, Y(s, X_s))ds | X_t].$$

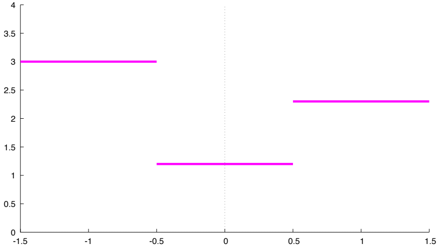
Use $Y^M(T, x) = g(x)$ and

$$Y^M(t_i, x) = \arg \inf_{\phi \in \mathscr{F}_Y} \frac{1}{M} \sum_{m=1}^{M} |g(X_N^{(m)}) + \sum_{j=i}^{N-1} f(Y^M(t_{j+1}, X_{j+1}^{(m)}))\Delta_j - \phi(X_{t_i}^{(m)})|^2.$$

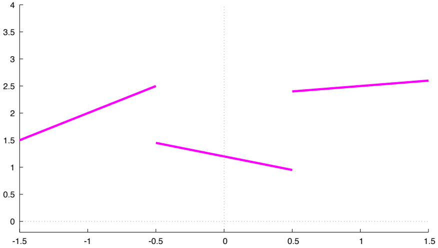
where $\mathscr{F}_Y \subset \mathbf{L}_{2,i} := \{\phi : \mathbb{R}^d \to \mathbb{R} \ : \ \mathbb{E}[|\phi(X_{t_i})|^2] < \infty\}$.

Choice of basis functions

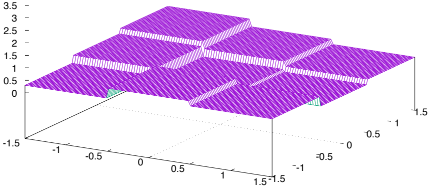
Piecewise constant approximation, 3 basis functions



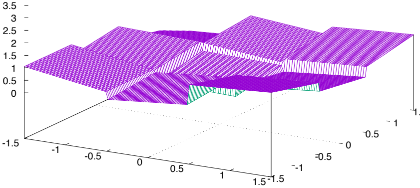
Piecewise affine approximation, 6 basis functions



Piecewise constant approximation, 9 basis functions



Piecewise affine approximation, 27 basis functions



Approximation of Z

We use the Malliavin weight's scheme [Gobet-T. 16]

$$Z(t_i, x) \approx \arg \inf_{\phi \in \mathscr{F}_Z} \frac{1}{M} \sum_{m=1}^{M} |g(X_T^{(m)}) H_{t_i,T} + \sum_{j=i+1}^{N-1} f(Y^M(t_{j+1}, X_{t_{j+1}})) H_{t_i,t_j} \Delta_j - \phi(X_{t_i}^{(m)})|^2.$$

$H_{t,s}$ are $\mathscr{F}_s$-measrable random variables satisfying

$$\mathbb{E}_t[H_{t,s}] = 0 \quad \text{and} \quad \mathbb{E}_t[|H_{t,s}|^2] \leq C_{Mall}(t-s)^{-1}.$$

Explicit form given for examples.

[Gobet-T. 16] E. Gobet and P. Turkedjiev.
Approximation of backward stochastic differential equations using Malliavin weights and least-squares regression.
Bernoulli, 22(1):530–562, 2016.

Adaptive importance sampling scheme

Continuous time heuristics

General BSDE: $\mathscr{Y}_t = \xi + \int_t^T f(s, \mathscr{Y}_s, \mathscr{Z}_s)ds - \int_t^T \mathscr{Z}_s dW_s$.

$$\mathscr{Y}_t = \mathbb{E}_{\mathbb{P}}[\xi + \int_t^T f(s, \mathscr{Y}_s, \mathscr{Z}_s)ds \mid \mathscr{F}_t].$$

Change of probability measure: for predictable process h s.t. $\int hdW$ is BMO,

$$W_t^{(h)} := W_t - \int_0^t h_r^\top dr,$$

$$L_t^h := e^{-\int_0^t h_r dW_r + \frac{1}{2}\int_0^t |h_r|^2 dr} = e^{-\int_0^t h_r dW_r^{(h)} - \frac{1}{2}\int_0^t |h_r|^2 dr},$$

$$L_{t,s}^h := \frac{L_s^h}{L_t^h} \quad \text{for } 0 \le t \le s \le T.$$

Girsanov's theorem: $\mathbb{Q}^h|_{\mathscr{F}_t} = [L_t^h]^{-1}\mathbb{P}|_{\mathscr{F}_t}$, $W^{(h)}$ is $\mathbb{Q}$-Brownian motion, and

$$\mathscr{Y}_t = \mathbb{E}_{\mathbb{Q}^h}\left[\xi L_{t,T}^h + \int_t^T f(s, \mathscr{Y}_s, \mathscr{Z}_s) L_{t,s}^h ds \;\mid \mathscr{F}_t\right].$$

Continuous time heuristics

Let $S(t,h) := \xi L^h_{t,T} + \int_t^T f(s, \mathscr{Y}_s, \mathscr{Z}_s) L^h_{t,s} ds$.

Aim: choose h s.t. $\sigma_t^2(h) := \mathbb{V}\mathrm{ar}_{\mathbb{Q}}(S(t,h)|\mathscr{F}_t)$ is minimal.

Ito formula on $\mathscr{Y}_\cdot L^h_\cdot$ gives:

$$\begin{aligned} S(t,h) &= (L^h_t)^{-1}\left(\mathscr{Y}_T L^h_T + \int_t^T f(s, \mathscr{Y}_s, \mathscr{Z}_s) L^h_s ds\right) \\ &= \mathscr{Y}_t + (L^h_t)^{-1} \int_t^T L^h_s (\mathscr{Z}_s - \mathscr{Y}_s h_s) dW^{(h)}_s. \end{aligned}$$

therefore $h_s = \mathscr{Z}_s / \mathscr{Y}_s$ implies that $\sigma_t^2(h) = 0$!

Need $\mathscr{Y} > 0$ a.s., but this is w.l.o.g. under assumptions of boundedness, Lipschitz continuity, etc.

Numerical scheme

Let

$$X_{i+1} = X_i + b(t_i, X_i)\Delta_i + \sigma(t_i, X_i)\Delta W_i,$$
$$\bar{X}^M_{i+1} = \bar{X}^M_i + \left\{ b(t_i, \bar{X}^M_i) + \frac{\sigma(t_i, \bar{X}_i) Z^M(t_i, \bar{X}^M_i)}{Y^M(t_i, \bar{X}^M_i)} \right\} \Delta_i + \sigma(t_i, \bar{X}^M_i)\Delta W_i.$$

with

$$Y^M(t_i, x) = \arg \inf_{\phi \in \mathscr{F}_Y} \frac{1}{M} \sum_{m=1}^{M} |g(\bar{X}^{M,m}_N) L^m_{i+1,N} + \sum_{j=i}^{N-1} f(Y(t_{j+1}, \bar{X}^{M,m}_{j+1})) L^m_{i+1,j+1} - \phi(\bar{X}^{M,m}_i)|^2,$$

$$Z^M(t_i, x) \approx \arg \inf_{\phi \in \mathscr{F}_Z} \frac{1}{M} \sum_{m=1}^{M} |g(X^{(m)}_N) H_{i,N} + \sum_{j=i+1}^{N-1} f(Y(t_{j+1}, X^{(m)}_{j+1})) H_{i,j} - \phi(X^{(m)}_i)|^2.$$

How to simulate $\bar{X}^M_i$ when $Y^m(t_j, \cdot)$ unknown for $j \leq i$?

Error estimates for least-squares regression

Setting

$Y \in \mathbb{R}$ and $X \in \mathbb{R}^d$ correlated RVs, $\{(Y_i, X_i) \;:\; 1 \le i \le n\}$ i.i.d. copies.

Problem: estimate $m(x) = \mathbb{E}[Y|X = x]$, assuming $|m|_\infty \le L$

Let $\mathscr{F} :=$ span$\{\mathbf{1}_{H_k}, \ldots, \mathbf{1}_{H_K}\}$, $H_k \subset \mathbb{R}^d$ are disjoint.

Approximation:

$$m_n(x) = \mathscr{T}_L\left(\psi^\star(x)\right), \quad \psi^\star(x) = \arg\inf_{\phi\in\mathscr{F}} \frac{1}{n}\sum_{i=1}^n |Y_i - \phi(X_i)|^2.$$

$\mathscr{T}_L$ truncation onto $[-L, L]$.

Theorem ([Gobet, T.])

For each $k \in \{1, \ldots, K\}$, *define* $\text{osc}_k^{(m)} := \sup_{x,y \in H_k} |m(x) - m(y)|$.

Define also the upper bound $\sigma^2 := \sup_{x \in \mathbb{R}^d} \mathbb{V}\text{ar}(Y \mid X = x)$. *Then*

$$\mathbb{E}[\int (m(x) - m_n(x))^2 \mathbb{P}(X \in dx)] \\ \leq C \sum_{k=1}^{K} [\text{osc}_k^{(m)}]^2 \nu(H_k) + CK\frac{\sigma^2}{n} + CL^2 \nu(D^c)$$

where $\nu(dx) = \mathbb{P}(X \in dx)$, $D := \cup_{k=1}^{K} H_k$.

In below reference, σ^2 is replaced by $L^2 \log(n)$; valid for general basis.

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

Error analysis for the BSDE scheme without importance sampling

Error estimates

For Y approximation space $\mathscr{F}_Y := \text{span}\{\mathbf{1}_{H_k}, \ldots, \mathbf{1}_{H_K}\}$.

For Z general approximation space $\mathscr{F}_Z$.

f does not depend on z.

$$\mathbb{E}[\|y_i(\cdot) - y_i^M(\cdot)\|_{\rho_i}^2]^{1/2} \leq C\left((\mathscr{E}^{Y,i})^{1/2} + \sum_{k=i+1}^{N-1} (\mathscr{E}^{Y,k})^{1/2}\Delta_k\right),$$

$$\mathbb{E}[\|z_i(\cdot) - z_i^M(\cdot)\|_{\rho_i}^2]^{1/2} \leq C\left((\mathscr{E}^{Z,i})^{1/2} + \sum_{k=i+1}^{N-1} \frac{(\mathscr{E}^{Y,k+1})^{1/2}}{\sqrt{t_k - t_i}}\Delta_k\right)$$

where $\rho_i(dx) = \mathbb{P}(X_i \in dx)$.

If f depends on z, $\mathscr{E}^{Z,k}$ terms appear in the sums.

Error terms

$$\mathscr{E}^{Y,i} := \sum_{k=1}^{K_Y} [\mathrm{osc}_k^{(y_i)}]^2 p_{k,i} + \sigma_{Y,i,M}^2 \frac{K_{Y,i}}{M} + C_y^2 \mathbb{P}(X_i \notin D^{(Y)}),$$

$$\mathscr{E}^{Z,i} := \inf_{\varphi \in \mathscr{F}_Z} \|z_i(\cdot) - \varphi(\cdot)\|_{\rho_i}^2 + C_z^2 \frac{q K_Z \log(M)}{M} + \frac{C_{Mall}^2}{T - t_i} \frac{K_{Z,i}}{M}.$$

where $p_{k,i} = \mathbb{P}(X_i \in H_k)$.

Without importance sampling, $\sigma_{Y,i,M}^2 \leq C$ is the best we can do.

Key technique: random initialization

Markov approach for sampling

Solution of minimization problem at time t_i doesn't depend on distribution of X_i!

How to simulate $\bar{X}$ from time 0, which depends on unknowns Y and Z?

Simulate the particles of X and $\bar{X}$ starting at t_i using arbitrary distribution for the initial position X_i; usual dynamics afterwards, e.g. for $j > i$:

$$X_j = X_{j-1} + B(t_{j-1}, X_{j-1})h + \sigma(t_{j-1}, X_{j-1})\Delta W_{j-1}.$$

Talk of Emmanuel Gobet: stratified sampling.

Sufficient condition for error estimates

For **every** i, $X_i^{(i)}$ sampled from density p satisfying Uniform Sub-Exponential Sandwiching (USES) property

$$\forall \lambda \in [0,\Lambda], x \in \mathbb{R}^d, \quad \frac{p(x)}{C(\Lambda)} \leq \int_{\mathbb{R}^d} p(x + z\sqrt{\lambda}) \frac{e^{-\frac{|z|^2}{2}}}{(2\pi)^{d/2}} dz \leq C(\Lambda) p(x),$$

$\exists C_p > 0$ such that, for all $\phi : \mathbb{R}^d \to \mathbb{R}$ square integrable and $j \geq i$,

$$\frac{\mathbb{E}[|\phi(X_i)|^2]}{C_p} \leq \mathbb{E}[|\phi(X_j)|^2] \leq C_p \mathbb{E}[|\phi(X_i)|^2].$$

Suitable densities: Laplace, logistic, twisted exponential, Parato type,... [Gobet, T.][Gobet, T. et al].

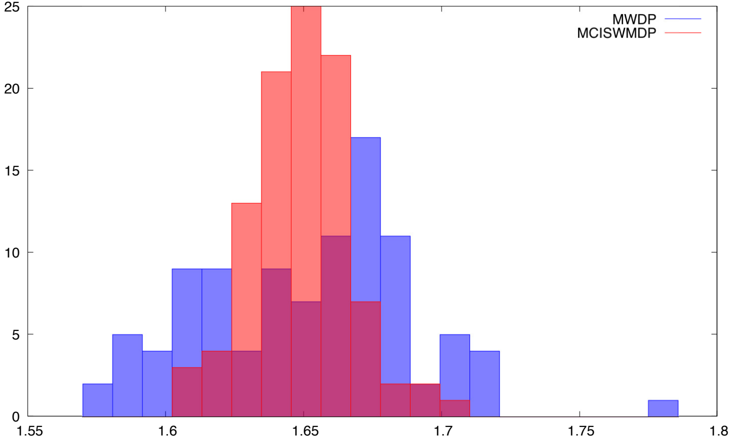
[Gobet, T. et al] E. Gobet, J. Lopez-Salas, P. Turkedjiev and C. Vasquez.
Stratified regression Monte-Carlo scheme for semilinear PDEs and BSDEs with large scale parallelization on GPUs.
In revision for SIAM Journal of Scientific Computing, preprint hal-01186000, 2015.

Numerical examples

Example 1

Let $\gamma > 0$ and $\lambda > 0$, let $g(x) := 1 + \gamma + \sin(\lambda \underline{\mathbf{1}}_q^\top x)$ and

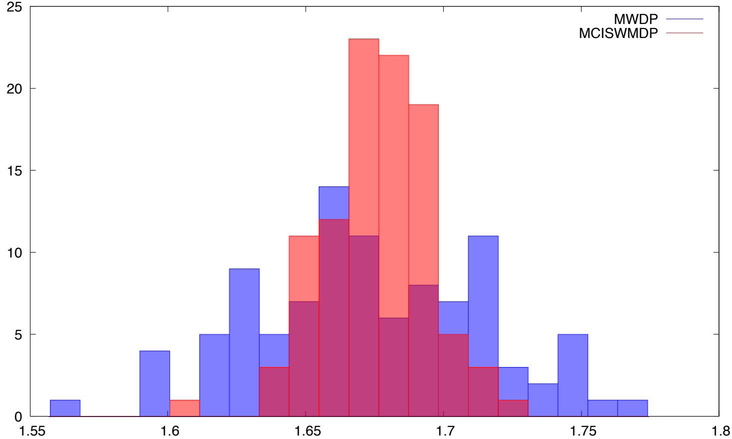
$$f(t,x,y) := 1 \wedge \left(y - \gamma - 1 - \frac{\sin(\lambda \underline{\mathbf{1}}_q^\top x)}{\exp(\lambda^2 q (T-t)/2)} \right)^2 .$$



Example 2

Add z to the driver:

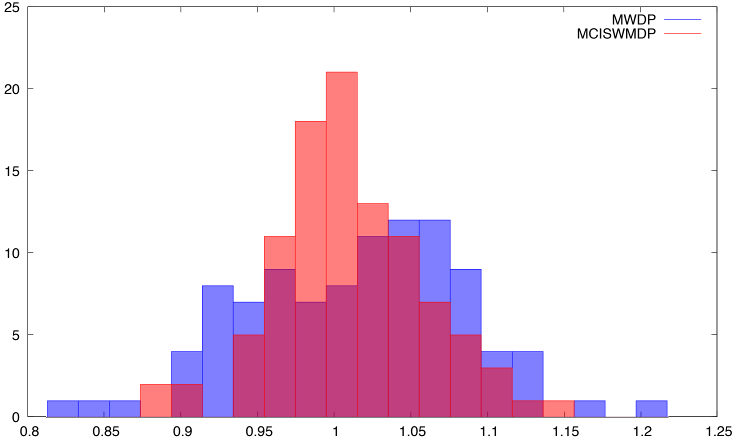
$$f(t,x,y) := 1 \wedge \left(y - \gamma - 1 - \frac{\sin(\lambda \underline{\mathbf{1}}_q^\top x)}{\exp(\lambda^2 q(T-t)/2)} \right)^2 + q \wedge \left| z - \frac{\lambda \cos(\lambda \underline{\mathbf{1}}_q^\top x)}{\exp(\lambda^2 q(T-t)/2)} \right|^2;$$



Example 3

Let $\omega(t,x) = \exp(t + \underline{\mathbf{1}}_q^\top x)$, $g(x) = \gamma + \omega(T,\lambda x)(1+\omega(T,\lambda x))^{-1}$ and

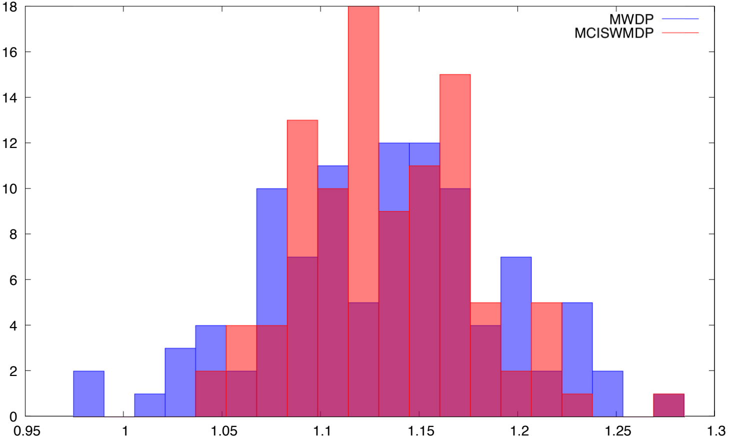
$$f(t,x,y) = \frac{\omega(t,\lambda x)}{(1+\omega(t,\lambda x))^2}\left(q\lambda^2(y-\gamma) - 1 - \frac{q}{2}\lambda^2\right).$$



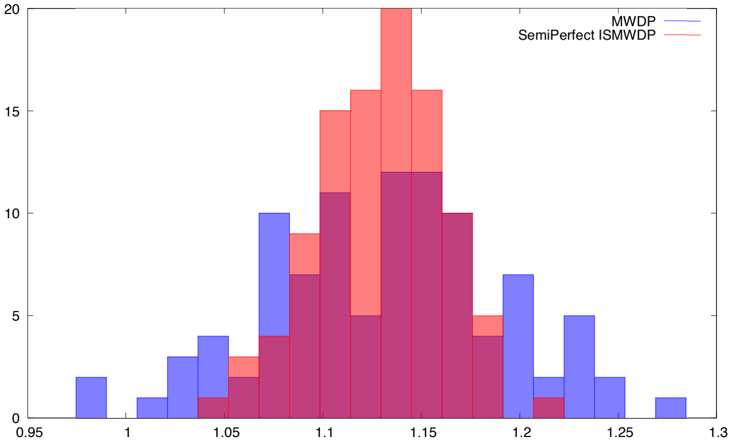
Example 4

Add z to the driver:

$$f(t,x,y,z) = \frac{z}{\lambda}\left(q\lambda^2(y-\gamma) - 1 - \frac{q}{2}\lambda^2\right).$$



Example 4



References

[Gobet, T.] E. Gobet and P. Turkedjiev.
Importance sampling for variance reduction in linear regression algorithms for backward stochastic differential equations.
Available on https://hal.archives-ouvertes.fr/hal-01169119.

Thank You!