Least-square regression Monte Carlo for approximating BSDEs and semilinear PDEs

Plamen Turkedjiev

BP International Plc.

20th July 2017

Forward-Backward Stochastic Differential Equations (FBSDEs)

Definitions and relations in continuous time

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

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

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

Feymann-Kac relation (Pardoux-Peng-92): $(Y_t, Z_t) = (Y(t, X_t), Z(t, X_t))$ where $(Y(t,x), Z(t,x))$ deterministic and solve $Y(t,x) = u(t,x)$ and $Z(t,x) = \nabla u(t,x)\sigma(t,x)$ for

$$\partial_t u(t,x) + \mathscr{L}(t,x)u(t,x) = f(t,x,u,\nabla_x u\sigma), \quad u(T,x) = \Phi(x),$$

$$\mathscr{L}(t,x)g(x) = \langle b(t,x), \nabla_x g(x)\rangle + \tfrac{1}{2} trace(\sigma\sigma^\top(t,x)Hess(g)(x)).$$

CE Z loss

First steps to discrete time approximation

Goals of numerical method

(1) approximate the stochastic process $\tilde{X} \approx X$;

(2) compute approximations of $Y(t,x)$ and $Z(t,x)$ minimizing the loss function

$$l(\phi,\psi) := \mathbb{E}[\sup_{0\le t\le T} |\phi(t,\tilde{X}_t) - Y(t,X_t)|^2] + \mathbb{E}[\int_0^T |\psi(t,\tilde{X}_t) - Z(t,X_t)|^2 dt];$$

(3) tune the approximation algorithm to minimize the computational cost.

In this talk, we are not concerned with approximating X; we drop the notation $\tilde{X}$ hereafter.

✗ The loss function is not tractable and we must make an approximation.

Finite time grid approximation

Let $\pi = \{0 = t_0 < \ldots < t_n = T\}$ and define the loss function

$$l_\pi(\phi, \psi) := \max_{t \in \pi} \mathbb{E}[|\phi(t, X_t) - Y(t, X_t)|^2] + \sum_i \mathbb{E}[\int_{t_i}^{t_{i+1}} |\psi(t, X_t) - Z(s, X_s)|^2 ds]$$

- Clearly $l_\pi(\cdot)$ is an approximation of $l(\cdot)$.
- The choice of π will affect the efficiency of the approximation.
- The regularity and boundedness of Φ, f, b, and σ will influence the efficiency of the approximation.

Conditional expectation formulation

By taking conditional expectations in ‣ BSDE:

$$\begin{aligned} Y_t &= \mathbb{E}\left[\Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds \middle| \mathscr{F}_t\right] \quad a.s. \\ &= \arg\inf_{\Psi_t \in \mathscr{A}(t)} \mathbb{E}\left[\left|\Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \Psi_t\right|^2\right] \end{aligned}$$

where $\mathscr{A}(t) = \mathbf{L}_2(\mathscr{F}_t; \mathbb{R})$. Markov property: replace $\mathscr{A}(t)$ by

$$\mathscr{A}_t = \{\psi : \mathbb{R}^d \to \mathbb{R} \mid \mathbb{E}[|\psi(X_t)|^2] < \infty\},$$

$$Y_t = \arg\inf_{\psi(t,\cdot) \in \mathscr{A}_t} \mathbb{E}\left[\left|\Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \psi(t, X_t)\right|^2\right]$$

‣ D1

Reformulation of the Y-part of the loss

Orthogonality of conditional expectation:

$$\mathbb{E}[|\psi(t,X_t)-\Phi(X_T)-\int_t^T f(s,X_s,Y_s,Z_s)ds|^2]$$

$$=\mathbb{E}[|\psi(t,X_t)-Y(t,X_t)|^2]+\mathbb{E}[|Y(t,X_t)-\Phi(X_T)-\int_t^T f(s,X_s,Y_s,Z_s)ds|^2]$$

The Y part of the loss function becomes

$$l_{\pi,y}(t,\psi)=\mathbb{E}[|\psi(t,X_t)-\Phi(X_T)-\int_t^T f(s,X_s,Y_s,Z_s)ds|^2].$$

Z part of the loss

The optimal discrete Z is also a conditional expectation, BSDE:

$$
\begin{aligned}
Z_\pi(t_i, x) &:= \arg\inf_{\phi \in \mathscr{A}_t} \mathbb{E}[\int_{t_i}^{t_{i+1}} |\phi(X_{t_i}) - Z(s, X_s)|^2 ds] \\
&= \frac{1}{t_{i+1} - t_i} \mathbb{E}[\int_{t_i}^{t_{i+1}} Z_s ds | X_{t_i} = x] \\
&= \mathbb{E}\left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left(\Phi(X_T) - \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds \right) \middle| X_{t_i} = x \right]
\end{aligned}
$$

D1

Z part of the loss

As before, we use orthogonality property of the conditional expectation

$$\begin{aligned}
&\mathbb{E}[|\phi(t_i, X_{t_i}) - Z_\pi(t_i, X_{t_i})|^2] \\
+&\mathbb{E}[|Z_\pi(t_i, X_{t_i}) - \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}\left(\Phi(X_T) + \int_{t_i}^T f(s, X_s, Y_s, Z_s)ds\right)|^2] \\
&= \mathbb{E}[|\phi(t_i, X_{t_i}) - \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}\left(\Phi(X_T) + \int_{t_i}^T f(s, X_s, Y_s, Z_s)ds\right)|^2] \\
&=: l_{\pi,z}(t_i, \phi).
\end{aligned}$$

The discrete loss is approximated by

$$l_\pi(\psi, \phi) \approx \max_{t_i \in \pi} l_{\pi,y}(t_i, \psi) + \sum_{t_i \in \pi} l_{\pi,z}(t_i, \phi)(t_{i+1} - t_i)$$

✗ The loss function is still not tractable because of the integral.

Equivalent continuous time representations

One-step vs. multistep approximation

From the tower law,

$$\begin{aligned} Y(t_i, x) &= \mathbb{E}\left[\Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s)ds \middle| X_{t_i} = x\right] \\ &= \mathbb{E}\left[Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s)ds \middle| X_{t_i} = x\right]. \end{aligned}$$

Likewise,

$$\begin{aligned} Z_\pi(t_i, x) &= \mathbb{E}\left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}\left(\Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s)ds\right) \middle| X_{t_i} = x\right] \\ &= \mathbb{E}\left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}\left(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s)ds\right) \middle| X_{t_i} = x\right]. \end{aligned}$$

D2

Decomposition into a system

Define $(\hat{y}, \hat{z})$ and $(\tilde{Y}, \tilde{Z})$ solving respectively

$$\hat{y}_t = \Phi(X_T) - \int_t^T \hat{z}_s dW_s,$$
$$\tilde{Y}_t = \int_t^T f(s, X_s, \hat{y}_s + \tilde{Y}_s, \hat{z}_s + \tilde{Z}_s) ds - \int_t^T \tilde{Z}_s dW_s.$$

Observe that $Y_t = \hat{y}_t + \tilde{Y}_t$ and $Z_t = \hat{z}_t + \tilde{Z}_t$.

✓ The representation is beneficial:

- The functions $\hat{y}(t, X_t) = \hat{y}_t$, $\hat{z}(t, X_t) = \hat{z}_t$ come from linear equation.
- The functions $\tilde{Y}(t, X_t) = \tilde{Y}_t$, $\tilde{Z}(t, X_t) = \tilde{Z}_t$ are generally *smoother* than their $Y(t, x)$, $Z(t, x)$ counterparts.

▸ ML

Adding zero

From the conditional expectation

$$Y(t_i, x) = \mathbb{E}\left[\Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s)ds \middle| X_{t_i} = x\right]$$
$$= \mathbb{E}\left[\underbrace{\Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s)ds - \int_{t_i}^{T} Z_s dW_s}_{= Y(t_i, x)} \middle| X_{t_i} = x\right]$$

✓ In other words, the integrand has conditional variance zero. More to come...

Adding zero

From the conditional expectation

$$Z_\pi(t_i, x) = \mathbb{E}\left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left(\Phi(X_T) + \int_{t_i}^T f(s, X_s, Y_s, Z_s) ds \right) \middle| X_{t_i} = x \right]$$

$$= \mathbb{E}\Bigg[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}$$

$$\times \left(\underbrace{\Phi(X_T) + \int_{t_i}^T f(s, X_s, Y_s, Z_s) ds - Y(t_i, x) - \int_{t_{i+1}}^T Z_s dW_s} \right) \Bigg| X_{t_i} = x \Bigg]$$

$$= Y(t_{i+1}, X_{t_{i+1}}) - Y(t_i, x) + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s) ds$$

✓ The integrand has low conditional variance zero. More to come...

▸ D3

Malliavin representation (Hu-Nualart-Song-11)

Rather than computing $Z_\pi(t,x)$, directly use the representation

$$\begin{aligned} Z(t,x) &= \mathbb{E}\left[D_t\Phi(X_T) + \int_{t_i}^T \nabla_x f(s,X_s,Y_s,Z_s)D_tX_s ds \middle| X_t = x\right] \\ &\quad + \mathbb{E}\left[\int_t^T \partial_y f(s,X_s,Y_s,Z_s)D_tY_s ds \middle| X_t = x\right] \\ &\quad + \mathbb{E}\left[\int_t^T \nabla_z f(s,X_s,Y_s,Z_s)D_tZ_s ds \middle| X_t = x\right] \\ &= \mathbb{E}\left[\Gamma(t,T)D_t\Phi(X_T) + \int_{t_i}^T \Gamma(t,s)\nabla_x f(s,X_s,Y_s,Z_s)D_tX_s ds \middle| X_t = x\right] \end{aligned}$$

with $D_tX_\tau = \nabla_x X_\tau (\nabla_x X_t)^{-1}\sigma(t,X_t)$ and

$$\Gamma(t,s) = \exp\left(\int_t^s \nabla_z f_\tau dW_\tau + \int_t^s (\partial_y f_\tau - \frac{1}{2}|\nabla_z f_\tau|^2 d\tau\right)$$

✗ Valid under restricted conditions.

Malliavin integration by parts

(Ma-Zhang-02)(T.-15) Rather than computing $Z_\pi(t,x)$, directly use the representation

$$Z(t,x) = \mathbb{E}\left[\Phi(X_T)M(t,T) + \int_{t_i}^{T} f(s,X_s,Y_s,Z_s)M(t,s)ds \middle| X_t = x\right]$$

for random variables

$$M(t,s) := \frac{1}{s-t}\int_t^s \sigma^{-1}(\tau,X_\tau)D_t X_\tau dW_\tau)^\top.$$

✗ Valid under restricted conditions.

✓ Sometimes $M(t,s)$ is available in closed form. E.g. for $X_t = W_t$ or geometric Brownian motion, $M(t,s) = \frac{W_s - W_t}{s-t}$.

▸ D3

Continuous time approximations

Truncation

Let $\Phi_M(x) = \Phi(\mathscr{T}_{1,M}(x))$, $f_M(t,x,y,z) = f(t,x,\mathscr{T}_{2,M}(y),\mathscr{T}_{3,M}(z))$ and define

$$Y_M(t) = \Phi_M(X_T) + \int_t^T f_M\big(t, X_s, Y_M(s), Z_M(s)\big)ds - \int_t^T Z_M(s)dW_s.$$

✓ Processes $(Y_M, Z_M) \approx (Y, Z)$ have better stability conditions, i.e. a priori estimates, comparison theorems.

$\rightarrow$ Important to approximate case of super-linear f (Chassagneux-Richou-16) (Lionnet-dos Reis-Szpruch-15).

Discrete time approximation

Discretizing the integral

Define $\Delta_i = t_{i+1} - t_i$, $\Delta W_j = W_{t_{j+1}} - W_{t_j}$, $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\mathscr{F}_{t_i}]$.

$$Y_i = \Phi(X_T) + \sum_{j\geq i} \mathbb{E}_j[f(t_j, X_{t_j}, Y_{j+1}, Z_j)]\Delta_j - \sum_{j\geq i} Z_j \Delta W_j - \sum_{j\geq i} \Delta L_j$$

where L_j *discrete time BSDE*. Kunita-Watanabe: $\exists!(Y, Z, L)$ s.t. $\{W_i L_i \ : \ i = 0, \ldots, n\}$ is a martingale w.r.t. discrete filtration and

$$Y_i = \mathbb{E}_i[\Phi(X_T) + \sum_{j\geq i} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j],$$

$$Z_i = \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i}(\Phi(X_T) + \sum_{j\geq i+1} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j)].$$

Discrete time analogue of ▸ Y ▸ Z

✓ Markov property: $Y_i = y_i(X_{t_i})$ and $Z_i = z_i(X_{t_i})$.

Discretizing the integral

The loss function is approximated by

$$l(\psi,\phi) \approx \max_{t_i\in\pi} \tilde{l}_{\pi,y}(t_i,\psi) + \sum_{t_i\in\pi} \Delta_i \tilde{l}_{\pi,z}(t_i,\phi)$$

where

$$\tilde{l}_{\pi,y}(t,\psi) = \mathbb{E}[|\psi(t,X_t) - \Phi(X_T) - \sum_{j\geq i} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j|^2]$$

$$\tilde{l}_{\pi,z}(t,\phi) = \mathbb{E}[|\psi(t,X_t) - \frac{\Delta W_i}{\Delta_i}(\Phi(X_T) + \sum_{j\geq i+1} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j)|^2]$$

✗ $(Y_i, Z_i)_{t_i\in\pi}$ is still not tractable because conditional expectations generally not available analytically. Loss function is still not tractable!

Other formulations

$Y_n = \Phi(X_T)$ and

$$Y_i = \mathbb{E}_i[Y_{i+1} + f(t_i, X_{t_i}, Y_{i+1}, Z_i)\Delta_i], \quad Z_i = \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i} Y_{i+1}].$$

Discrete time analogue of continuous time equations.

Likewise,

$$Y_i = \mathbb{E}_i[\Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j - \sum_{j \geq i} Z_j \Delta W_j],$$

$$Z_i = \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i}(\Phi(X_T) + \sum_{j \geq i+1} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j - Y_i - \sum_{j \geq i} Z_j \Delta W_j)]$$

is the discrete time analogue of continuous time equations.

PE T ML USES

Convergence result

(A1) $\Phi(\cdot)$ is θ_Φ-Hölder continuous;

(A2) $L_f, C_f \in [0,\infty)$ and $\theta_L, \theta_C \in [0,1)$ s.t.
$|f(t,x,0,0)| \le C_f(T-t)^{\theta_C - 1}$, and

$$|f(t,x_1,y_1,z_1)-f(t,x_2,y_2,z_2)| \le L_f \frac{|x_1-x_2|+|y_1-y_2|+|z_1-z_2|}{(T-t_i)^{(1-\theta_L)/2}};$$

(A3) $b(t,x)$ and $\sigma(t,x)$ twice differentiable in x, $\frac{1}{2}$-Hölder in t, bounded and bounded partial derivatives, and $\exists \eta > 0 \; s.t. \; x^\top \sigma\sigma^\top x > \eta |x|^2$.

(A4) For $\beta \in (0,1]$, $t_i = T - T(1-i/N)^\beta$.

For $\beta < \gamma \wedge \theta_\Phi \wedge \theta_L$, let $\gamma = \theta_C \wedge (2\theta_C \wedge \theta_\Phi + \theta_L)$,
(Gobet-Makhlouf-10)(T.-15) show

$$\inf l(\psi,\phi) \le O(n^{-1}) \mathbf{1}_{\theta_\Phi+\gamma+\beta\ge 1} + O(n^{-\gamma}) \mathbf{1}_{\theta_\Phi+\gamma+\beta<1}.$$

Convergence result

(A1) $\Phi(\cdot)$ is Lipschitz continuous;
(A2) $f(t,x,y,z)$ is Lipschitz continuous in (x,y,z) with linear growth, $\frac{1}{2}$-Hölder continuous in t;
(A3) $b(t,x)$ and $\sigma(t,x)$ are Lipschitz continuous with linear growth in x and $\frac{1}{2}$-Hölder in t.

(Zhang-04) shows $\inf l(\psi,\phi) \leq O(n^{-1})$; (Gobet-Labart-07) show additionally under $\Phi \in C_1(\mathbb{R}^d : \mathbb{R})$ that $\inf l(\psi,\phi) \leq O(n^{-2})$.

Two alternatives

Conditioning inside the driver (Pagès-Sagna-17):

$$Y_i = \Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, \mathbb{E}_j[Y_{j+1}], Z_j)\Delta_j - \sum_{j \geq i} Z_j \Delta W_j - \sum_{j \geq i} \Delta L_j$$

Implicit version:

$$Y_i = \Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, Y_j, Z_j)\Delta_j - \sum_{j \geq i} Z_j \Delta W_j - \sum_{j \geq i} \Delta L_j$$

✓ There are many references for implicit numerical scheme, (Chassagneux-Richou-16) prove that it tends to be more stable than the explicit version (with modification on ΔW terms).

Picard scheme for One-step/multistep implicit schemes

One-step scheme from (Gobet-Lemor-Warin-05):

$$
\begin{aligned}
Y_{q+1,i} &= \mathbb{E}_i[Y_{q,i+1}] + f(t_i, X_{t_i}, Y_{q,i}, Z_{q,i})\Delta_i \\
Z_{q+1,i} &= \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i} Y_{q+1,i}].
\end{aligned}
$$

Multistep scheme of (Bender-Denk-07)

$$
\begin{aligned}
Y_{q+1,i} &= \mathbb{E}_i[\Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, Y_{q,j}, Z_{q,j})\Delta_j] \\
Z_{q+1,i} &= \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i}(\Phi(X_T) + \sum_{j \geq i+1} f(t_j, X_{t_j}, Y_{q,j}, Z_{q,j})\Delta_j)].
\end{aligned}
$$

High order discretization of the integral

(Chassagneux-Crisan-14) Let $(Y_n, Z_n) = (\Phi(X_T), \nabla_x \Phi(X_T)\sigma(T, X_T))$.

For $j = 1, \dots, q$, and for $i < n$: set $(Y_{i,q}, Z_{i,q}) = (Y_{i+1}, Z_{i+1})$ and

$$Y_{i,j} = \mathbb{E}_{i,j}[Y_{i+1} + c_j \Delta_i \sum_{k=j}^{q} a_{j,k} f(t_k, X_{t_k}, Y_{i,k}, Z_{i,k})]$$

$$Z_{i,j} = \mathbb{E}_{i,j}[H_{i,j} Y_{i+1} + \Delta_i \sum_{k=j+1}^{q} A_{j,k} H_{i,k} f(t_k, X_{t_k}, Y_{i,k}, Z_{i,k})]$$

Set $(Y_i, Z_i) = (Y_{i,0}, Z_{i,0})$.

Given sufficient smoothness and Hörmander condition, optimal four stage explicit scheme loss is $\inf l(\psi, \phi) \leq O(n^{-6})$.

Given sufficient smoothness and Hörmander condition, optimal three stage implicit scheme loss is $\inf l(\psi, \phi) \leq O(n^{-6})$.

Discrete time Malliavin weights scheme

(T.-15)(Gobet-T.-15) Recalling Malliavin representation of Z, discrete approximation of the integral and Malliavin weight terms (first order approximation):

$$Y_i = \mathbb{E}_i\left[\Phi(X_T) + \sum_{j\geq i} f(t_j, X_{t_j}, Y_{j+1}, Z_j)\Delta_j\right],$$

$$Z_i = \mathbb{E}_i\left[\Phi(X_T)M_{i,n} + \sum_{j\geq i+1} f(t_j, X_{t_j}, Y_{j+1}, Z_j)M_{i,j}\Delta_j\right].$$

New loss function for Z:

$$\hat{l}_{\pi,z}(t,\phi) = \mathbb{E}[|\phi(t,X_t) - \Phi(X_T)M_{i,n} - \sum_{j\geq i+1} f(t_j, X_{t_j}, Y_{j+1}, Z_j)M_{i,j}\Delta_j|^2].$$

PE T

Convergence result

(A1) $\Phi(\cdot)$ is θ_Φ-Hölder continuous;

(A2) $L_f, C_f \in [0,\infty)$ and $\theta_L, \theta_C \in [0,1)$ s.t.
$|f(t,x,0,0)| \le C_f(T-t)^{\theta_C - 1}$, and

$$|f(t,x_1,y_1,z_1) - f(t,x_2,y_2,z_2)| \le L_f \frac{|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|}{(T-t_i)^{(1-\theta_L)/2}};$$

(A3) $b(t,x)$ and $\sigma(t,x)$ twice differentiable in x, $\frac{1}{2}$-Hölder in t, bounded and bounded partial derivatives, and $\exists \eta > 0 \; s.t. \; x^\top \sigma\sigma^\top x > \eta |x|^2$.

(A4) For $\beta \in (0,1]$, $t_i = T - T(1-i/N)^\beta$.

For $\beta < \gamma \wedge \theta_\Phi \wedge \theta_L$, let $\gamma = \theta_C \wedge (2\theta_C \wedge \theta_\Phi + \theta_L)$, (T.-15) shows

$$\inf l(\psi,\phi) \le O(n^{-1})\mathbf{1}_{\theta_\Phi+\gamma+\beta\ge 1} + O(n^{-\gamma})\mathbf{1}_{\theta_\Phi+\gamma+\beta<1}.$$

Least-squares regression

General setup

Let $S_{t_i,T}(\omega) = S_i(\omega(t_i), \ldots \omega(t_n))$ be a random fuctional and define

$$l_\pi(t_i, \phi) := \mathbb{E}[|\phi(X_{t_i}) - S_{t_i,T}(X)|^2]$$

Then $\arg\inf_{\phi \in \mathscr{A}_{t_i}} l_\pi(t_i, \phi)(x) = \phi^\star(x) := \mathbb{E}[S_{t_i,T}(X) | X_t = x]$.

✗ Generally can't compute $\mathbb{E}[\cdot]$ for a search policy.

General setup

Estimating the measure by empirical measure,

$$l_\pi(t,\phi) \approx l_{\pi,M}(t,\phi) := \frac{1}{M}\sum_{m=1}^{M} |\phi(X_{t_i}^{(m)}) - S_{t_i,T}(m, X^{(m)})|^2$$
$$\Rightarrow \arg\inf_{\phi\in\mathscr{A}_{t_i}} l_\pi(t,\phi) \approx \arg\inf_{\phi\in\mathscr{A}_{t_i}} l_{\pi,M}(t,\phi)$$

✗ $\mathscr{A}_{t_i}$ infinite dimensional, not suitable for a search policy.

General setup

Two stage approximation:

- Choosing finite dimensional hypothesis space $\mathscr{K} \subset \mathscr{A}_{t_i}$,

$$\arg\inf_{\mathscr{A}_{t_i}} l_\pi(t_i, \phi) \approx \arg\inf_{\mathscr{K}} l_\pi(t_i, \phi);$$

  approximation error

$$\mathbb{E}[|\phi^*(X_{t_i}) - \phi^*_{\mathscr{K}}(X_{t_i})|^2] = \inf_{\mathscr{K}} \mathbb{E}[|\phi^*(X_{t_i}) - \phi(X_{t_i})|^2].$$

  because $\forall \phi \in \mathscr{K}$

$$\mathbb{E}[|S_{t_i,T}(X) - \phi(X_{t_i})|^2] = \mathbb{E}[|S_{t_i,T}(X) - \phi^*(X_{t_i})|^2] + \mathbb{E}[|\phi^*(X_{t_i}) - \phi(X_{t_i})|^2].$$

☠ The choice of hypothesis space is crucial: good space "is close to" the solution.

General setup

- Approximate the probability measure with the empirical measure

$$\arg\inf_{\mathscr{A}_{t_i}} l_\pi(t_i, \phi) \approx \arg\inf_{\mathscr{K}} l_{\pi,M}(t_i, \phi)$$

where $l_{\pi,M}(t_i, \phi) := \frac{1}{M}\sum_{m=1}^{M} |S_{t_i,T}(m, X^{(m)}) - \phi(X_{t_i}^{(m)})|^2$.

Let $\{p_1(x), \ldots, p_K(x)\}$ be a basis for $\mathscr{K}$, $X := [p_k(X_{t_i}^{(m)})]_{m,k}$, and $y = [S_{t_i,T}(m, X^{(m)})]_m$:

$$\inf_{\mathscr{K}} l_{\pi,M}(t_i, \phi) = \inf_{\beta \in \mathbb{R}^K} \frac{1}{M} |X\beta - y|_2^2$$

✓ The right-hand side is a least-squares problem (least-squares regression): finally a tractable algorithm!

Error estimation

Assume $\phi^\star(x) := \mathbb{E}[S_{t_i,T}(X)|X_{t_i} = x]$ is bounded by L.

Define $\phi^\star_{\mathscr{K},M}(x) = \mathscr{T}_L(p(x)^\top \beta^\star_M)$.

$$
\begin{aligned}
&\mathbb{E}[|\phi^\star(X_{t_i}) - \phi^\star_{\mathscr{K},M}(X_{t_i})|^2] \\
&= \mathbb{E}[|\phi^\star(X_{t_i}) - \phi^\star_{\mathscr{K},M}(X_{t_i})|^2 - \frac{2}{M}|\phi^\star(X^{(\cdot)}_{t_i}) - \phi^\star_{\mathscr{K},M}(X^{(\cdot)}_{t_i})|_2^2] \\
&\quad + \mathbb{E}[\frac{2}{M}|\phi^\star(X^{(\cdot)}_{t_i}) - \phi^\star_{\mathscr{K},M}(X^{(\cdot)}_{t_i})|_2^2] \\
&\leq \mathbb{E}[\sup_{\phi\in\mathscr{K}}\left(\mathbb{E}[|\phi^\star(X_{t_i}) - \mathscr{T}_L(\phi(X_{t_i}))|^2] - \frac{2}{M}|\phi^\star(X^{(\cdot)}_{t_i}) - \mathscr{T}_L(\phi(X^{(\cdot)}_{t_i}))|_2^2\right)_+] \\
&\quad + \mathbb{E}[\frac{2}{M}|\phi^\star(X^{(\cdot)}_{t_i}) - \phi^\star_{\mathscr{K},M}(X^{(\cdot)}_{t_i})|_2^2]
\end{aligned}
$$

Concentration of measure

Very conservative upper bound (Gobet-T-15):

$$\mathbb{E}[\sup_{\phi\in\mathscr{K}}\left(\mathbb{E}[|\phi^\star(X_{t_i})-\mathscr{T}_L(\phi(X_{t_i}))|^2]-\frac{2}{M}|\phi^\star(X_{t_i}^{(\cdot)})-\mathscr{T}_L(\phi(X_{t_i}^{(\cdot)}))|_2^2\right)_+] \leq \frac{2028(K+1)\log(3M)L^2}{M}.$$

✓ Converges as $M\to\infty$.

✗ Low variance of $S_{t_i,T}(X)$ doesn't appear to improve estimates.

☠ Tricky, conservative estimation using Vapnik-Chervonenkis dimension.

Empirical measure part

$$\mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - \phi^\star_{\mathscr{K},M}(X_{t_i}^{(\cdot)})|_2^2]$$

$$\leq \mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - p(X_{t_i}^{(\cdot)})^\top \beta^\star_M|_2^2]$$

$$= \underbrace{\mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - p(X_{t_i}^{(\cdot)})^\top \hat{\beta}^\star_M|_2^2]}_{\leq M \inf_{\phi \in \mathscr{K}} \mathbb{E}[|\phi^\star(X_{t_i}) - \phi(X_{t_i})|^2]} + \mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top (\beta^\star_M - \hat{\beta}^\star_M)|_2^2]$$

where $\hat{\beta}^\star_M := \arg\inf_{\beta \in \mathbb{R}^K} |\phi^\star(X_{t_i}^{(m)}) - p(X_{t_i}^{(m)})^\top \beta|_2^2$.

Statistical error

Normal equations:

$$\beta^* \in \arg\inf_{\beta\in\mathbb{R}^K} |X\beta - y|^2 \iff X^\top X\beta^* = X^\top y.$$

w.l.o.g. basis functions orthonormal in empirical norm, normal equations give

$$\begin{aligned}
&\frac{1}{M}|p(X_{t_i}^{(\cdot)})^\top(\beta_M^\star - \hat{\beta}_M^\star)|_2^2 = |\beta_M^\star - \hat{\beta}_M^\star|_2^2 \\
&= \frac{1}{M^2}\sum_{m_1,m_2=1}^{M}\sum_{k=1}^{K} p_k(X_{t_i}^{(m_1)})p_k(X_{t_i}^{(m_2)}) \\
&\qquad \times (S_{t_i,T}(X^{(m_1)}) - \phi^\star(X_{t_i}^{(m_1)}))(S_{t_i,T}(X^{(m_2)}) - \phi^\star(X_{t_i}^{(m_2)}))
\end{aligned}$$

Statistical error

Taking conditional expectations w.r.t. $\{X_{t_i}^{(m)}\}_m$ and then expectations,

$$\frac{1}{M}\mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top(\beta_M^\star - \hat{\beta}_M^\star)|_2^2] \le K\frac{\sup_x \mathbb{V}(S_{t_i,T}(X)|X_{t_i}=x)K}{M}.$$

✓ Impact of variance is captured in this estimate, where it was not in the concentration of measure.

Special case improvement: piecewise constant basis

(Gobet-T.-16) Let $p_k(x) = \mathbf{1}_{H_k}(x)$, $\{H_k \subset \mathbb{R}^d\}_{k=1,\dots,K}$.
For each $k \in \{1, \dots, K\}$, define $\operatorname{osc}_k^{(m)} := \sup_{x,y \in H_k} |\phi^\star(x) - \phi^\star(y)|$.

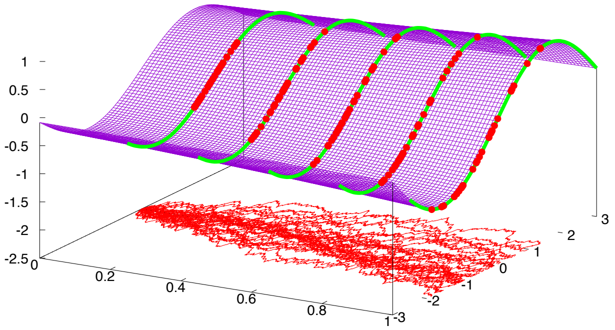
Define also the upper bound $\sigma^2 := \sup_{x \in \mathbb{R}^d} \mathbb{V}(Y \mid X = x)$. Then

$$\mathbb{E}[|\phi^\star(X_{t_i}) - \phi^\star_{\mathscr{K},M}(X_{t_i})|^2]$$
$$\leq C \sum_{k=1}^{K} [\operatorname{osc}_k^{(m)}]^2 \mathbb{P}(X_{t_i} \in H_k) + CK\frac{\sigma^2}{M} + CL^2 \nu(D^c)$$

where $D := \cup_{k=1}^{K} H_k$.

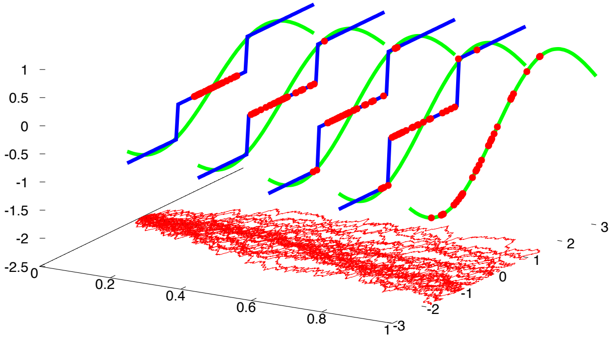
Back to the BSDE approximation

$S_{t_i,T}(X) = \Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, y_{j+1}(X_{t_{j+1}}), z_j(X_{t_j}))\Delta_j$:



Back to the BSDE approximation

$$S^M_{t_i,T}(X) = \Phi(X_T) + \sum_{j\geq i} f(t_j, X_{t_j}, y^M_{j+1}(X_{t_{j+1}}), z^M_j(X_{t_j}))\Delta_j:$$



▸ T ▸ USES

Propagation of error

$$
\begin{aligned}
&\mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - \phi^\star_{\mathscr{K},M}(X_{t_i}^{(\cdot)})|_2^2] \\
&\le \mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - p(X_{t_i}^{(\cdot)})^\top \beta^\star_M|_2^2] \\
&= \mathbb{E}[|\phi^\star(X_{t_i}^{(\cdot)}) - p(X_{t_i}^{(\cdot)})^\top \hat{\beta}^\star_M|_2^2] + \mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top (\beta^\star_M - \hat{\beta}^\star_M)|_2^2] \\
&\le M \inf_{\phi \in \mathscr{K}} \mathbb{E}[|\phi^\star(X_{t_i}) - \phi(X_{t_i})|^2] + 2\mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top (\tilde{\beta}^\star_M - \hat{\beta}^\star_M)|_2^2] \\
&\qquad + 2\mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top (\hat{\beta}^\star_M - \beta^\star_M)|_2^2]
\end{aligned}
$$

where $\hat{\beta}^\star_M = \arg\inf_{\beta \in \mathbb{R}^K} |\mathbb{E}[S^M_{t_i,T}(X^{(\cdot)}) | \{X_{t_i}^{(m)}\}_m] - p(X_{t_i}^{(\cdot)})|_2$

Propagation of error

$$\mathbb{E}[|p(X_{t_i}^{(\cdot)})^\top(\tilde{\beta}_M^\star - \hat{\beta}_M^\star)|_2^2] \leq M\mathbb{E}[|y_i(X_{t_i}) - \mathbb{E}[S_{t_i,T}^M(X)|X_{t_i}]|^2]$$

Now, $Y_i^M := \mathbb{E}[S_{t_i,T}^M(X)|X_{t_i}]$ solves linear discrete BSDE with driver

$$f_M(t_i, X_{t_i}) := \mathbb{E}[f(t_i, X_{t_i}, y_{i+1}^M(X_{t_{i+1}}), z_i^M(X_{t_i}))|\{X^{(m)}\}_m, X_{t_i}]$$

so the term above is treated with a priori estimates for discrete BSDE.

N.B. Compare with one step scheme, where

$$\hat{S}_{t_i,T}^M(X) = y_{i+1}^M(X_{t_{i+1}}) + f(t_i, X_{t_i}, y_{i+1}^M(X_{t_{i+1}}), z_i^M(X_{t_i}))\Delta_i$$

discrete BSDE property is lost $\Rightarrow$ large propagation of error.

✓ Similar analysis for Malliavin weights scheme.

USES

Least-squares regression

Method of normal equations:

$$\beta^* \in \arg\inf_{\beta\in\mathbb{R}^K} |X\beta - y|^2 \iff X^\top X\beta^* = X^\top y.$$

$\beta^\star = \arg\inf\{|\beta^*|_2\}$ is unique and given by $\beta^\star = A^\dagger y$ for $A = X^\top X$.

Condition number: $\kappa(B) = \max\sigma_0(B)/\min\sigma_0(B)$ determines *sensitivity* of solving a linear problem. I.e., $|B^\dagger(y+\epsilon) - B^\dagger y|_2/|B^\dagger y|_2$.

Cost $= O(K^2M)$ to form $X^\top X = \sum_{m=1}^M p(X_{t_i}^{(m)})p(X_{t_i}^{(m)})^\top$, can be done in parallel.

☠ For normal equations: $\kappa(A) = \kappa(X)^2$.

Least-squares regression

Method of QR factorization: multiplication by orthogonal matrix P doesn't change length,

$$|P(X\beta - y)|_2 = |X\beta - y|_2.$$

$\exists! Q = [Q_1 \ Q_2]$ orthogonal and $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ upper-right triangular (R_1 full rank) such that $X = QR$.

$$|X\beta - y|_2^2 = |QR\beta - y|_2^2 = |\cancel{Q^\top Q}\, R\beta - Q^\top y|_2^2 = |R_1\beta - Q_1^\top y|_2^2 + |Q_2 y|_2^2$$

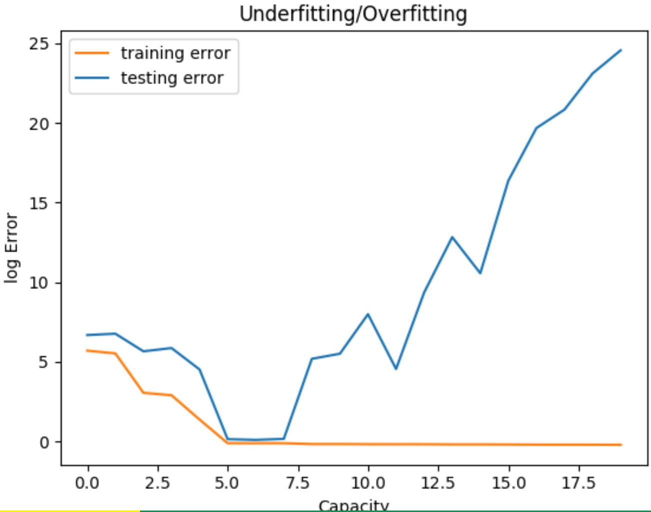
So $\beta^\star = R_1^{-1} Q_1^\top y$.

Cost $= O(K^2 M)$ to compute the QR factorization.

✓ Condition number: $\kappa(R_1) = \kappa(X)$, much better than for normal equations!

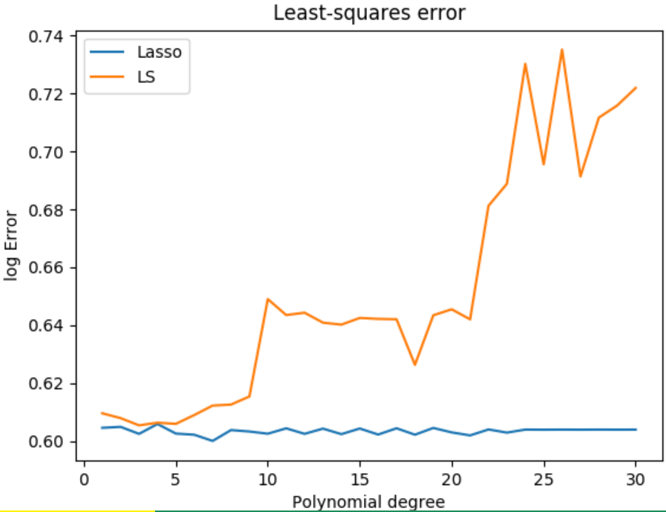
Choice of hypothesis space (Goodfellow et al-16)

How well does the coefficient generalize? Draw i.i.d. testing sample:

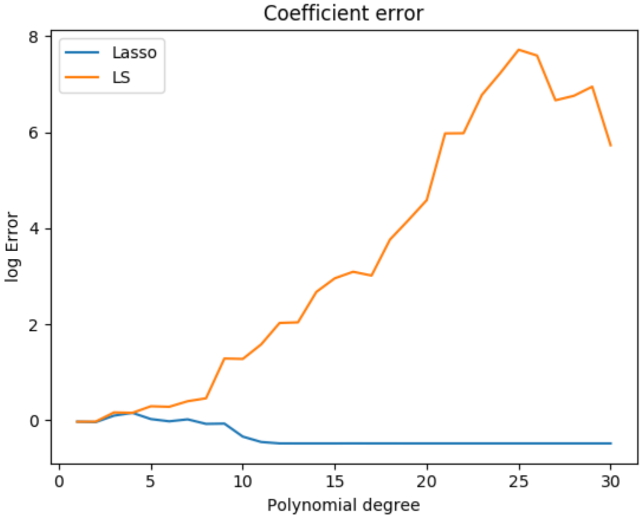


Regularization

Add “lasso" penalty $\mu|\beta|_1$ to the training loss function, unmodified testing loss:



Regularization



BSDE tricks

In high dimension, constrained by memory budget and computational time

- To conserve memory, re-simulate X trajectories at each time point ▸ simulation
- Use variance reduction schemes ▸ var ▸ Malli
- Reduce time points by high order scheme (Chassagneux-Crisan-14).
- If you don't care about conserving coefficients, use the one-step scheme to conserve coefficients ▸ OS
- Use the USES sampling method to increase basis stability and leverage HPC...

Multilevel scheme (Becherer-T.-14)

$$f(t,x,y,z) = \left(\sum_{k=1}^{d} z_k\right)\left(0 \vee y \wedge 1 - \frac{2+d}{2d}\right), \Phi(x) = \frac{\exp(T+\sum_{k=1}^{d} x_k)}{1+\exp(T+\sum_{k=1}^{d} x_k)}$$

Variance reduced scheme based on ▸ system ▸ var:

N	$MSE_{Y,\text{max}}$	$MSE_{Y,\text{av}}$	$MSE_{Z,\text{av}}$
4	0.0335796	0.0083949	0.0126556
8	0.0334017	0.00417521	0.00651092
16	0.0421584	0.0026349	0.00344173

Standard multistep forward :

N	$MSE_{Y,\text{max}}$	$MSE_{Y,\text{av}}$	$MSE_{Z,\text{av}}$
4	0.0353173	0.00882931	0.0351813
8	0.0372012	0.00465015	0.0289552
16	0.0474109	0.00296318	0.025199

Uniform Sub-Exponential Sandwiching (USES)

Stratified simulation

If X_{t_i} distribution explicit, stratified sampling possible.

Removes sources of instability:

- random sample size per cell in piecewise basis simulation
- high condition number due to poor basis selection.

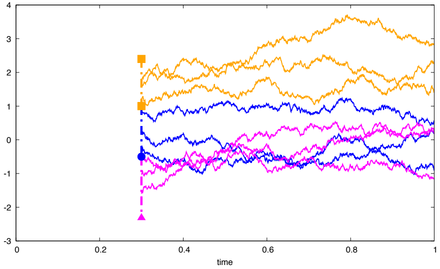
✓ In piecewise basis, cell-by-cell simulation also reduce simulation memory budget constraint and parallel processing across cells reduces computation time.

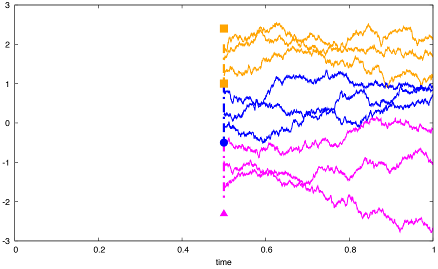
✗ X_{t_i} distribution is rarely explicit.

Generic method for Markov X

Function $y_i(\cdot)$ determined by *transition function* of X after t_i; doesn't care about X_{t_i} law. ▸ DP

Simulations $\{X^{(i,m)} \ : \ m = 1, \ldots, M\}$ started from an arbitrary random variable at time t_i.





✗ Need to conserve law of $\{X^{(i)}\}_i$ to estimate propagation of error.
▸ Err

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 $j \geq i$,

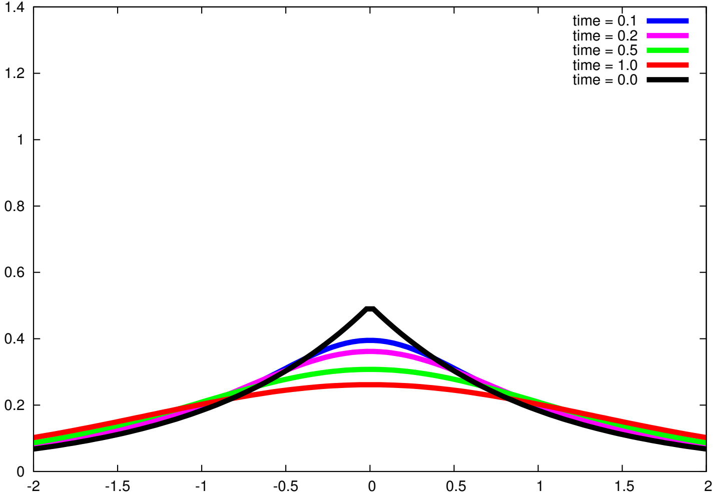
$$\frac{\mathbb{E}[|\phi(X_i)|^2]}{C_p} \leq \mathbb{E}[|\phi(X_j)|^2] \leq C_p \mathbb{E}[|\phi(X_i)|^2].$$

Suitable densities: Laplace, logistic, twisted exponential, Parato type,... (Gobet-T.-16) (Gobet-Salas-T.-Vázquez-16).

Huge advantage: easy stratified simulation.

Sufficient conditions on random initial value

For initial density $p(x) = 0.5 \times \exp(-|x|)$, density of particles is almost stationary:



Piecewise constant $d = 6$

- 12 core CPU processor with 2.9GHz, $-O3$ compiler optimization.
- Nvidia GeForce GTX Titan Black 6GB memory.
- #C=$(\#$ cells$)^{1/d} = \left\lfloor 2\sqrt{N} \right\rfloor$.

Δ_t	#C	K	M	$MSE_{Y,\text{max}}$	$MSE_{Y,\text{av}}$	$MSE_{Z,\text{av}}$	CPU	GPU
0.2	4	4096	25	-2.707882	-2.784022	-0.477751	0.29	1.94
0.1	6	46656	100	-3.195937	-3.294488	-1.133834	13.72	2.44
0.05	8	262144	400	-3.505867	-3.664396	-1.795697	775.33	52.20

Piecewise affine high dimensional examples

- 12 core CPU processor with 2.9GHz, $-O3$ compiler optimization.
- Nvidia GeForce GTX Titan Black 6GB memory.
- $\#\mathtt{C} = 2$.

d	K	M	$MSE_{Y,\mathsf{max}}$	$MSE_{Y,\mathsf{av}}$	$MSE_{Z,\mathsf{av}}$	CPU	GPU
15	32768	5000	−2.981181	−3.106590	−1.574532	578.88	139.60
16	65536	6000	−2.795353	−2.959375	−1.588716	1411.75	429.53
17	131072	5000	−2.772595	−2.936549	−1.371146	2580.06	793.61
18	262144	4000	−2.845755	−2.918057	−1.114600	4275.13	1589.30
19	524288	3200	−2.726427	−2.851617	−0.839849	7245.91	4370.31

Adaptive importance sampling scheme

Change of probability measure

SDE satisies $dX_t = b_t dt + \sigma_t dW_t$, approximation scheme is

$$Y(t,x) := \mathbb{E}_t[\Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds]$$

SDE satisies $d\tilde{X}_t = \tilde{b}_t dt + \sigma_t dW_t$, approximation scheme is

$$Y(t,x) := \mathbb{E}_t[\Phi(\tilde{X}_T)\mathscr{L}_{t_i,T}(\tilde{b}) + \int_t^T f(s,\tilde{X}_s,Y_s,Z_s)\mathscr{L}_{t,s}(\tilde{b})ds]$$

Optimal choice to minimize variance (Gobet-T.-15):

$$\tilde{b}_t = b_t + \sigma_t \frac{Z_t}{Y_t};$$

How to obtain particles $\tilde{X}_{t_i}^{(m)}$ without $\{(Y_t, Z_t) \ : \ t \leq t_i\}$?

Use stationarity of the distribution:
letting $\tilde{X}_i$ have distribution $\lambda(dx) = \prod_{j=1}^d 0.5 \times \exp(-|x_j|)dx$, simulate paths $\{\tilde{X}_i, \tilde{X}_{i+1}, \ldots, \tilde{X}_N\}$.

Defining $d\mathscr{L}_t(h) := \mathscr{L}_t(h)h_t dW_t$,

$$S(t,T) =(\mathscr{L}_t(h))^{-1}\left(Y_T\mathscr{L}_T(h) + \int_t^T f(s,Y_s,Z_s)\mathscr{L}_s(h)ds\right)$$

$$= \mathscr{Y}_t + (\mathscr{L}_t(h))^{-1}\int_t^T \mathscr{L}_s(h)[-f(s,Y_s,Z_s)ds + Z_s dW_s]$$

$$- (\mathscr{L}_t(h))^{-1}\int_t^T [\mathscr{L}_s(h)Y_s h_s dW_s^{(h)} + \mathscr{L}_s(h)Y_s h_s^\top Z_s ds]$$

$$+ (\mathscr{L}_t(h))^{-1}\int_t^T f(s,Y_s,Z_s)\mathscr{L}_s(h)ds$$

$$= Y_t + (\mathscr{L}_t(h))^{-1}\int_t^T \mathscr{L}_s(h)(Z_s - Y_s h_s)dW_s^{(h)}.$$

Choosing $h = Z/Y$, the $\mathscr{F}_{t_i}$-conditional variance of $S(t,T)$ is zero under the changed probability.

Fully implementable scheme

Setting $\mathscr{L}_{i,j} = \exp\left(-\sum_{k=i+1}^{j-1}\left\{\frac{Z_k^M(\tilde{X}_k)^\top \Delta W_k}{Y_k^M(\tilde{X}_k)} + \frac{|Z_k^M(\tilde{X}_k)|^2\Delta_k}{2|Y_k^M(\tilde{X}_k)|^2}\right\}\right)$,

$$Y_i(\tilde{X}_i) := \underbrace{\mathbb{E}_i[\Phi(\tilde{X}_N)\mathscr{L}_{i,N} + \sum_{j=i}^{N-1} f_j(\tilde{X}_j, Y_{j+1}(\tilde{X}_{j+1}))\mathscr{L}_{i,j}\Delta_j]}_{\approx S(t_i,T)}$$

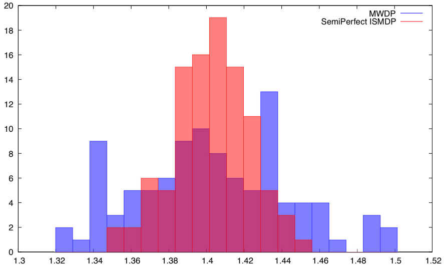
$Z_k^M(x)$ obtained without importance sampling with a Malliavin Weight's scheme:

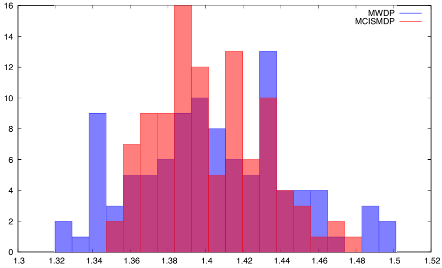
$$Z_i(X_i) := \mathbb{E}_i[\Phi(X_N)H_N^i + \sum_{j=i+1}^{N-1} H_j^i f_j(X_j, Y_{j+1})\Delta_j)],$$

Limitations:

- No (efficient) importance sampling available for the Z component.
- Can't include Z dependence in the driver due to the propagation of non-variance reduction.

Why the approximation of Z is important





Thank You!