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

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

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

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

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

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

- Clearly $l_\pi(\cdot)$ is an approximation of $l(\cdot)$.
- The choice of $\pi$ will affect the efficiency of the approximation.
- The regularity and boundedness of $\Phi$, $f$, $b$, and $\sigma$ will influence the efficiency of the approximation.
Conditional expectation formulation

By taking conditional expectations in BSDE:

\[ Y_t = \mathbb{E} \left[ \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) \, ds \right] \bigg| \mathcal{F}_t \] a.s.

\[ = \arg \inf_{\Psi \in \mathcal{A}(t)} \mathbb{E} \left[ \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) \, ds - \Psi_t \right]^2 \]

where \( \mathcal{A}(t) = L_2(\mathcal{F}_t; \mathbb{R}) \). Markov property: replace \( \mathcal{A}(t) \) by

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

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

\[
l_{\pi, y}(t, \psi) = \mathbb{E}[|\psi(t, X_t) - \Phi(X_T) - \int_t^T f(s, X_s, Y_s, Z_s) ds|^2].
\]
The optimal discrete $Z$ is also a conditional expectation, \textit{BSDE}:

$$Z_{\pi}(t_i, x) := \arg\inf_{\phi \in \mathcal{A}_t} \mathbb{E} \left[ \int_{t_i}^{t_{i+1}} |\phi(X_{t_i}) - Z(s, X_s)|^2 ds \right]$$

$$= \frac{1}{t_{i+1} - t_i} \mathbb{E} \left[ \int_{t_i}^{t_{i+1}} Z_s ds | X_{t_i} = x \right]$$

$$= \mathbb{E} \left[ \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left( \Phi(X_T) - \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds \right) | X_{t_i} = x \right]$$
As before, we use orthogonality property of the conditional expectation

\[ E[|\phi(t_i, X_{t_i}) - Z_\pi(t_i, X_{t_i})|^2] \]

\[ + E[|Z_\pi(t_i, X_{t_i}) - \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left( \Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds \right)|^2] \]

\[ = E[|\phi(t_i, X_{t_i}) - \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left( \Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds \right)|^2] \]

\[ =: l_{\pi,z}(t_i, \phi). \]

The discrete loss is approximated by

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

\[ \times \quad \text{The loss function is still not tractable because of the integral.} \]
Equivalent continuous time representations
From the tower law,

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

\[ = \mathbb{E} \left[ Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s) ds \bigg| X_{t_i} = x \right]. \]

Likewise,

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

\[ = \mathbb{E} \left[ \frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \left( Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, X_s, Y_s, Z_s) ds \right) \bigg| X_{t_i} = x \right]. \]
Decomposition into a system

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

\[
\hat{y}_t = \Phi(X_T) - \int_t^T \hat{z}_s dW_s,
\]

\[
\check{Y}_t = \int_t^T f(s, X_s, \hat{y}_s + \check{Y}_s, \hat{z}_s + \check{Z}_s) ds - \int_t^T \check{Z}_s dW_s.
\]

Observe that \(Y_t = \hat{y}_t + \check{Y}_t\) and \(Z_t = \hat{z}_t + \check{Z}_t\).

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 \(\check{Y}(t, X_t) = \check{Y}_t, \check{Z}(t, X_t) = \check{Z}_t\) are generally smoother than their \(Y(t, x), Z(t, x)\) counterparts.
From the conditional expectation

\[
Y(t_i, x) = \mathbb{E} \left[ \Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds \mid X_{t_i} = x \right]
\]

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

\[
= Y(t_i, x)
\]

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

From the conditional expectation

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

\[
= \mathbb{E} \left[ \frac{W_{t_i+1} - W_{t_i}}{t_{i+1} - t_i} \right]
\]

\[
\times \left( \Phi(X_T) + \int_{t_i}^{T} f(s, X_s, Y_s, Z_s) ds - Y(t_i, x) - \int_{t_i+1}^{T} Z_s dW_s \right) \mid X_{t_i} = x
\]

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

✓ The integrand has low conditional variance zero. More to come...
Malliavin representation (Hu-Nualart-Song-11)

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

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

$$+ \mathbb{E} \left[ \int_{t_i}^T \partial_y f(s, X_s, Y_s, Z_s) D_t Y_s ds \bigg| X_t = x \right]$$

$$+ \mathbb{E} \left[ \int_{t_i}^T \nabla_z f(s, X_s, Y_s, Z_s) D_t Z_s ds \bigg| X_t = x \right]$$

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

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

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

for random variables

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

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

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

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

→ Important to approximate case of super-linear $f$ (Chassagneux-Richou-16) (Lionnet-dos Reis-Szpruch-15).
Discrete time approximation
Discretizing the integral

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

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

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 $\boldsymbol{Y}$ $\boldsymbol{Z}$

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

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

where

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

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

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

\[ Y_n = \Phi(X_T) \text{ and} \]

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

Discrete time analogue of \( \text{continuous time equations} \).

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 \( \text{continuous time equations} \).
Convergence result

\( (A1) \) \( \Phi(\cdot) \) is \( \theta_\Phi \)-Hölder continuous;

\( (A2) \) \( L_f, C_f \in [0, \infty) \) and \( \theta_L, \theta_C \in [0, 1) \) s.t.
\[
|f(t, x, 0, 0)| \leq C_f (T - t)^{\theta_C - 1}, \text{ and}
\]
\[
|f(t, x_1, y_1, z_1) - f(t, x_2, y_2, z_2)| \leq L_f \frac{|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|}{(T - t_i)^{(1 - \theta_L)/2}};
\]

\( (A3) \) \( b(t, x) \) 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 \( \beta < \gamma \land \theta_\Phi \land \theta_L \), let \( \gamma = \theta_C \land (2\theta_C \land \theta_\Phi + \theta_L) \),

\( (\text{Gobet-Makhlouf-10})(\text{T.-15}) \) show

\[
\inf l(\psi, \phi) \leq O(n^{-1}) 1_{\theta_\Phi + \gamma + \beta \geq 1} + O(n^{-\gamma}) 1_{\theta_\Phi + \gamma + \beta < 1}.
\]
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 \(\Delta W\) terms).
Discrete time approximation

Picard scheme for One-step/multistep implicit schemes

One-step scheme from \textbf{(Gobet-Lemor-Warin-05)}:

\[
Y_{q+1,i} = \mathbb{E}_i[Y_{q,i+1}] + f(t_i, X_{t_i}, Y_{q,i}, Z_{q,i}) \Delta_i
\]

\[
Z_{q+1,i} = \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i} Y_{q+1,i}].
\]

Multistep scheme of \textbf{(Bender-Denk-07)}

\[
Y_{q+1,i} = \mathbb{E}_i[\Phi(X_T) + \sum_{j \geq i} f(t_{j}, X_{t_{j}}, Y_{q,j}, Z_{q,j}) \Delta_j]
\]

\[
Z_{q+1,i} = \mathbb{E}_i[\frac{\Delta W_i}{\Delta_i} (\Phi(X_T) + \sum_{j \geq i+1} f(t_{j}, X_{t_{j}}, Y_{q,j}, Z_{q,j}) \Delta_j)].
\]
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 \(j = 1, \ldots, q\), 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 approximation

**Discrete time Malliavin weights scheme**


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

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

New loss function for $Z$:

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

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

(A2) $L_f, C_f \in [0, \infty)$ and $\theta_L, \theta_C \in [0, 1)$ s.t.

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

and

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

(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

$$\inf l(\psi, \phi) \leq O(n^{-1}) 1_{\theta_\phi + \gamma + \beta \geq 1} + O(n^{-\gamma}) 1_{\theta_\phi + \gamma + \beta < 1}.$$
Least-squares regression
Let $S_{t_i, T}(\omega) = S_i(\omega(t_i), \ldots \omega(t_n))$ be a random functional and define

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

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

Generally can’t compute $\mathbb{E}[\cdot]$ for a search policy.
Estimating the measure by empirical measure,

\[
l_\pi(t, \phi) \approx l_{\pi,M}(t, \phi) := \frac{1}{M} \sum_{m=1}^{M} |\phi(X_{t_i}^{(m)}) - S_{t_i,T}(m, X^{(m)})|^2
\]

\[
\Rightarrow \arg \inf_{\phi \in \mathcal{A}_{t_i}} l_\pi(t, \phi) \approx \arg \inf_{\phi \in \mathcal{A}_{t_i}} l_{\pi,M}(t, \phi)
\]

\[\mathcal{A}_{t_i} \] infinite dimensional, not suitable for a search policy.
Two stage approximation:

- Choosing finite dimensional hypothesis space $\mathcal{H} \subset A_{t,i}$,

$$\arg \inf_{A_{t,i}} l_\pi (t_i, \phi) \approx \arg \inf_{\mathcal{H}} l_\pi (t_i, \phi);$$

approximation error

$$\mathbb{E}[|\phi^* (X_{t_i}) - \phi_\mathcal{H} (X_{t_i})|^2] = \inf_{\mathcal{H}} \mathbb{E}[|\phi^* (X_{t_i}) - \phi (X_{t_i})|^2].$$

because $\forall \phi \in \mathcal{H}$

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

⚠️ The choice of hypothesis space is crucial: good space “is close to” the solution.
General setup

- Approximate the probability measure with the empirical measure

$$\arg \inf_{\mathcal{A}_t} l_\pi(t_i, \phi) \approx \arg \inf_{\mathcal{KH}} l_{\pi,M}(t_i, \phi)$$

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

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

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

✓ The right-hand side is a least-squares problem (least-squares regression): finally a tractable algorithm!
Assume $\phi^*(x) := \mathbb{E}[S_{t_i,T}(X)|X_{t_i} = x]$ is bounded by $L$.

Define $\phi^*_\mathcal{K},M(x) = \mathcal{T}_L(p(x)\top \beta^*_M)$.

$$\mathbb{E}[|\phi^*(X_{t_i}) - \phi^*_\mathcal{K},M(X_{t_i})|^2]$$

$$= \mathbb{E}[|\phi^*(X_{t_i}) - \phi^*_\mathcal{K},M(X_{t_i})|^2 - \frac{2}{M} |\phi^*(X_{t_i}^{(\cdot)}) - \phi^*_\mathcal{K},M(X_{t_i}^{(\cdot)})|^2]$$

$$+ \mathbb{E}[\frac{2}{M} |\phi^*(X_{t_i}^{(\cdot)}) - \phi^*_\mathcal{K},M(X_{t_i}^{(\cdot)})|^2]$$

$$\leq \mathbb{E}[\sup_{\phi \in \mathcal{K}} \left( \mathbb{E}[|\phi^*(X_{t_i}) - \mathcal{T}_L(\phi(X_{t_i}))|^2] - \frac{2}{M} |\phi^*(X_{t_i}^{(\cdot)}) - \mathcal{T}_L(\phi(X_{t_i}^{(\cdot)}))|^2 \right) +$$

$$+ \mathbb{E}[\frac{2}{M} |\phi^*(X_{t_i}^{(\cdot)}) - \phi^*_\mathcal{K},M(X_{t_i}^{(\cdot)})|^2]$$
Concentration of measure

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

$$\mathbb{E}[\sup_{\phi \in \mathcal{H}} \left( \mathbb{E}[|\phi^*(X_{t_i}) - \mathcal{T}_L(\phi(X_{t_i}))|^2] - \frac{2}{M} |\phi^*(X_{t_i}) - \mathcal{T}_L(\phi(X_{t_i}))|^2 \right) + \frac{2028(K + 1) \log(3M)L^2}{M}].$$

✓ 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

\[
\mathbb{E}[|\phi^*(X_{t_i}) - \phi_{\mathcal{K},M}(X_{t_i})|^2] \\
\leq \mathbb{E}[|\phi^*(X_{t_i}) - p(X_{t_i})^\top \beta^*_M|^2] \\
= \mathbb{E}[|\phi^*(X_{t_i}) - p(X_{t_i})^\top \beta^*_M|^2] + \mathbb{E}[|p(X_{t_i})^\top (\beta^*_M - \hat{\beta}^*_M)|^2] \\
\leq M \inf_{\phi \in \mathcal{K}} \mathbb{E}[|\phi^*(X_{t_i}) - \phi(X_{t_i})|^2]
\]

where \( \hat{\beta}^*_M := \arg \inf_{\beta \in \mathbb{R}^K} |\phi^*(X_{t_i}^{(m)}) - p(X_{t_i}^{(m)})^\top \beta|^2 \)
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

\[
\frac{1}{M} |p(X_{t_i}^{(1)})^\top (\beta^*_M - \hat{\beta}^*_M)|^2 = |\beta^*_M - \hat{\beta}^*_M|^2
\]

\[
= \frac{1}{M^2} \sum_{m_1,m_2=1}^{M} \sum_{k=1}^{K} p_k(X_{t_i}^{(m_1)}) p_k(X_{t_i}^{(m_2)}) \\
\times (S_{t_i}, T(X^{(m_1)}) - \phi^*(X_{t_i}^{(m_1)}))(S_{t_i}, T(X^{(m_2)}) - \phi^*(X_{t_i}^{(m_2)}))
\]
Taking conditional expectations w.r.t. \( \{X^{(m)}_{t_i}\}_m \) and then expectations,

\[
\frac{1}{M} \mathbb{E}[\|p(X^{(\cdot)}_{t_i})^\top (\beta^*_M - \hat{\beta}^*_M)\|_2^2] \leq K \sup_x \frac{\nabla (S_{t_i}, T(X) | X_{t_i} = x) K}{M}.
\]

✓ Impact of variance is captured in this estimate, where it was not in the concentration of measure.
(Gobet-T.-16) Let $p_k(x) = 1_{H_k}(x)$, $\{H_k \subset \mathbb{R}^d\}_{k=1,\ldots,K}$.

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

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

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

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

\[ S_{t_i,T}^M(X) = \Phi(X_T) + \sum_{j \geq i} f(t_j, X_{t_j}, y^M_{j+1}(X_{t_{j+1}}), z^M_j(X_{t_j})) \Delta_j : \]
\[
E[|\phi^*(X_{ti}) - \phi^*,M(X_{ti})|^2] \\
\leq E[|\phi^*(X_{ti}) - p(X_{ti})^T \beta^*_M|^2] \\
= E[|\phi^*(X_{ti}) - p(X_{ti})^T \hat{\beta}^*_M|^2] + E[|p(X_{ti})^T (\beta^*_M - \hat{\beta}^*_M)|^2] \\
\leq M \inf_{\phi \in \mathcal{K}} E[|\phi^*(X_{ti}) - \phi(X_{ti})|^2] + 2E[|p(X_{ti})^T (\beta^*_M - \hat{\beta}^*_M)|^2] \\
+ 2E[|p(X_{ti})^T (\hat{\beta}^*_M - \beta^*_M)|^2]
\]

where \( \hat{\beta}^*_M = \arg \inf_{\beta \in \mathbb{R}^K} E[S_{t_i,T}^M(X^{(\cdot)})|\{X_{t_i}^{(m)}\}_m] - p(X_{ti})\)
Propagation of error

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

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.

\(\checkmark\) Similar analysis for Malliavin weights scheme.
Least-squares regression

Method of normal equations:

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

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

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 \( X^\top X = \sum_{m=1}^M p(X_{ti}^{(m)})p(X_{ti}^{(m)})^\top \), can be done in parallel.

\[ \kappa(A) = \kappa(X)^2. \]

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

$$\|X\beta - y\|_2^2 = \|QR\beta - y\|_2^2 = \|Q^\top Q \ R\beta - Q^\top y\|_2^2 = \|R_1\beta - Q_1^\top y\|_2^2 + \|Q_2y\|_2^2$$

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

Cost = $O(K^2M)$ 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:
Add “lasso” penalty $\mu \| \beta \|_1$ to the training loss function, unmodified testing loss:
Regularization

Coefficient error

![Graph showing coefficient error comparison between Lasso and LS methods across different polynomial degrees. The graph plots log error against polynomial degree, with Lasso and LS methods represented by different lines.](image)
BSDE tricks
In high dimension, constrained by memory budget and computational time

- To conserve memory, re-simulate $X$ trajectories at each time point.
- Use variance reduction schemes.
- 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.
- Use the USES sampling method to increase basis stability and leverage HPC...
Multilevel scheme (Becherer-T.-14)

\[
f(t, x, y, z) = \left( \sum_{k=1}^{d} z_k \right) \left( 0 \lor y \land 1 - \frac{2+d}{2d} \right), \quad \Phi(x) = \frac{\exp(T + \sum_{k=1}^{d} x_k)}{1 + \exp(T + \sum_{k=1}^{d} x_k)}
\]

Variance reduced scheme based on system:

\[
N \quad \text{MSE}_{Y,\text{max}} \quad \text{MSE}_{Y,\text{av}} \quad \text{MSE}_{Z,\text{av}}
\]

<table>
<thead>
<tr>
<th>N</th>
<th>MSE_{Y,\text{max}}</th>
<th>MSE_{Y,\text{av}}</th>
<th>MSE_{Z,\text{av}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0335796</td>
<td>0.0083949</td>
<td>0.0126556</td>
</tr>
<tr>
<td>8</td>
<td>0.0334017</td>
<td>0.0041752</td>
<td>0.0065109</td>
</tr>
<tr>
<td>16</td>
<td>0.0421584</td>
<td>0.0026349</td>
<td>0.0034417</td>
</tr>
</tbody>
</table>

Standard multistep forward:

\[
N \quad \text{MSE}_{Y,\text{max}} \quad \text{MSE}_{Y,\text{av}} \quad \text{MSE}_{Z,\text{av}}
\]

<table>
<thead>
<tr>
<th>N</th>
<th>MSE_{Y,\text{max}}</th>
<th>MSE_{Y,\text{av}}</th>
<th>MSE_{Z,\text{av}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0353173</td>
<td>0.00882931</td>
<td>0.0351813</td>
</tr>
<tr>
<td>8</td>
<td>0.0372012</td>
<td>0.00465015</td>
<td>0.0289552</td>
</tr>
<tr>
<td>16</td>
<td>0.0474109</td>
<td>0.00296318</td>
<td>0.025199</td>
</tr>
</tbody>
</table>
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
- 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.

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.
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 $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ 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:

![Graph showing the density of particles over time, with a peak at time = 0.5 and a gradual decrease towards time = 1.0.](chart.png)
Piecewise constant $d = 6$

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

<table>
<thead>
<tr>
<th>$\Delta_t$</th>
<th>$#C$</th>
<th>$K$</th>
<th>$M$</th>
<th>$MSE_{Y,\text{max}}$</th>
<th>$MSE_{Y,\text{av}}$</th>
<th>$MSE_{Z,\text{av}}$</th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>4</td>
<td>4096</td>
<td>25</td>
<td>-2.707882</td>
<td>-2.784022</td>
<td>-0.477751</td>
<td>0.29</td>
<td>1.94</td>
</tr>
<tr>
<td>0.1</td>
<td>6</td>
<td>46656</td>
<td>100</td>
<td>-3.195937</td>
<td>-3.294488</td>
<td>-1.133834</td>
<td>13.72</td>
<td>2.44</td>
</tr>
<tr>
<td>0.05</td>
<td>8</td>
<td>262144</td>
<td>400</td>
<td>-3.505867</td>
<td>-3.664396</td>
<td>-1.795697</td>
<td>775.33</td>
<td>52.20</td>
</tr>
</tbody>
</table>
Piecewise affine high dimensional examples

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

<table>
<thead>
<tr>
<th>(d)</th>
<th>(K)</th>
<th>(M)</th>
<th>(MSE_{Y,\text{max}})</th>
<th>(MSE_{Y,\text{av}})</th>
<th>(MSE_{Z,\text{av}})</th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>32768</td>
<td>5000</td>
<td>-2.981181</td>
<td>-3.106590</td>
<td>-1.574532</td>
<td>578.88</td>
<td>139.60</td>
</tr>
<tr>
<td>16</td>
<td>65536</td>
<td>6000</td>
<td>-2.795353</td>
<td>-2.959375</td>
<td>-1.588716</td>
<td>1411.75</td>
<td>429.53</td>
</tr>
<tr>
<td>17</td>
<td>131072</td>
<td>5000</td>
<td>-2.772595</td>
<td>-2.936549</td>
<td>-1.371146</td>
<td>2580.06</td>
<td>793.61</td>
</tr>
<tr>
<td>18</td>
<td>262144</td>
<td>4000</td>
<td>-2.845755</td>
<td>-2.918057</td>
<td>-1.114600</td>
<td>4275.13</td>
<td>1589.30</td>
</tr>
<tr>
<td>19</td>
<td>524288</td>
<td>3200</td>
<td>-2.726427</td>
<td>-2.851617</td>
<td>-0.839849</td>
<td>7245.91</td>
<td>4370.31</td>
</tr>
</tbody>
</table>
Adaptive importance sampling scheme
SDE satisfies $dX_t = b_t dt + \sigma_t dW_t$, 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 satisfies \( d\tilde{X}_t = \tilde{b}_t dt + \sigma_t dW_t \), approximation scheme is

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

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

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

How to obtain particles \( \tilde{X}_{i}^{(m)} \) 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 \( d\mathcal{L}_t(h) := \mathcal{L}_t(h)h_t dW_t \),

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

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

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

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

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

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

Setting \( \mathcal{L}_{i,j} = \exp \left( - \sum_{k=i+1}^{j-1} \left\{ \frac{Z_{k}^{M}(\tilde{X}_{k})^{\top} \Delta W_{k}}{Y_{k}^{M}(\tilde{X}_{k})} + \frac{|Z_{k}^{M}(\tilde{X}_{k})|^{2} \Delta k}{2|Y_{k}^{M}(\tilde{X}_{k})|^{2}} \right\} \right) \),

\[
Y_{i}(\tilde{X}_{i}) := \mathbb{E}_{i}[\Phi(\tilde{X}_{N}) \mathcal{L}_{i,N} + \sum_{j=i}^{N-1} f_{j}(\tilde{X}_{j}, Y_{j+1}(\tilde{X}_{j+1})) \mathcal{L}_{i,j} \Delta j]
\]

\[ \approx S(t_{i}, T) \]

\( Z_{k}^{M}(x) \) 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!