Numerical Methods for Stochastic Controls

nyz

January 2025

Contents

1 Problem Formation

[1]

Suppose $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is a filtered probability space satisfies usual condition. Denote:

$$\mathbf{S} = [0, T) \times \mathbb{R}^d, \quad U \subset \mathbb{R}^k.$$

$$b : \mathbf{S} \times U \to \mathbb{R}^d, \quad (t, x, u) \mapsto b(t, x, u)$$
$$\sigma : \mathbf{S} \times U \to \mathbb{R}^{d \times r}, \quad (t, x, u) \mapsto \sigma(t, x, u).$$

We assumme b, σ satistifies the following:

$$|b(t, x, u) - b(t, y, u)| + |\sigma(t, x, u) - \sigma(t, y, u)| \leq K|x - y|$$
$$|b(t, x, u)| + |\sigma(t, x, u)| \leq K(1 + |x| + |u|).$$

Define second order operator $\mathcal{L}^u$:

$$\mathcal{L}^u \varphi = b(t, x, u) \cdot D\varphi(t, x) + \frac{1}{2}\mathbf{tr}(\sigma\sigma^T(t, x, u)D^2\varphi(t, x, u)), \forall u \in U.$$

Definition 1 *admissible controls*

A progressively measurable process ν is called admissible if

$$\mathbb{E}[\int_0^T \nu_s^2 \mathrm{d}s] < \infty.$$

The set of all admissible processes is denoted by $\mathcal{A}_0$. Similarly, we define

$$\mathcal{A}_t = \{\nu \in \mathcal{A}_0 : \nu \perp \mathcal{F}_t\}.$$

For each $\nu \in \mathcal{A}$, the equation:

$$\mathrm{d}X_t = b(t, X_t, \nu_t)\mathrm{d}t + \sigma(t, X_t, \nu_t)\mathrm{d}B_t$$

admits unique strong solution for a given initial data. We use $X^{t,x,\nu}$ to denote the solution started at t with initial value x under control ν.

Definition 2 *gain functional*

We define the gain functional as follows:

$$J(t, x, \nu) = \mathbb{E}[\int_t^T f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + g(X_T^{t,x,\nu})]$$

If $f : \mathbf{S} \times U \to \mathbb{R}$, $g : \mathbb{R}^d \to \mathbb{R}$ satisfies quadratic growth condition

$$|f(t, x, u)| + |g(x)| \leq K(1 + |x| + |u|^2),$$

then the gain functional is well defined.

Definition 3 *optimal value function and optimal control*

We define the optimal value function as follows:

$$V(t,x) = \sup_{\nu \in \mathcal{A}_0} J(t,x,\nu).$$

If $V(t,x) = J(t,x,\nu^*_{t,x})$, we call $\nu^*_{t,x}$ an optimal control.

Theorem 4

$$\sup_{\nu \in \mathcal{A}_0} J(t,x,\nu) = \sup_{\nu \in \mathcal{A}_t} J(t,x,\nu)$$

Definition 5 *stochastic control problem*

For fixed t, x, we define the stochastic control problem as:

$$\begin{aligned} \text{maximize}_{\nu \in \mathcal{A}} \quad & J(t,x,\nu) \\ \text{subject to} \quad & \mathrm{d}X_t = b(t,X_t,\nu_t)\mathrm{d}t + \sigma(t,X_t,\nu_t)\mathrm{d}B_t \\ & X_t^{t,x,\nu} = x. \end{aligned}$$

There are two types of methods to study stochastic control problems, Hamilton-Jacobian-Bellman(HJB) method and Stochastic Maximal Principle(SMP). The first method study the local behaviour of optimal controls, while SMP describe the optimal criterion of optimal control (Just like the optimal condition in convex optimization problem).

2 Related Results

2.1 Dynamic Programming Principle

In this subsection, we will study the local behaviour value function. We denote:

$$\mathcal{T}_{t,T} = \{\tau \text{ is a stopping time } : \tau(\omega) \in [t,T] \text{ and } \tau \perp \mathcal{F}_t\}.$$

Theorem 6 *DPP for J*

For each $\nu \in \mathcal{A}$ and $\tau \in \mathcal{T}_{t,T}$, we have

$$J(t,x,\nu) = \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\nu_s}, \nu_s)\mathrm{d}s + J(\tau, X_\tau^{t,x,\nu}, \nu)\right]$$

Proof

$$\begin{aligned}
J(t,x,\nu) &= \mathbb{E}[\int_t^T f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + g(X_T^{t,x,\nu})] \\
&= \mathbb{E}[\int_t^T f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + g(X_T^{t,x,\nu})] \\
&= \mathbb{E}[\int_t^\tau f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + \int_\tau^T f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + g(X_T^{t,x,\nu})] \\
&= \mathbb{E}[\int_t^\tau f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + \int_\tau^T f(s, X_s^{\tau, X_\tau^{t,x,\nu},\nu}, \nu_s)\mathrm{d}s + g(X_T^{t, X_\tau^{t,x,\nu},\nu})] \\
&= \mathbb{E}[\int_t^\tau f(s, X_s^{t,x,\nu}, \nu_s)\mathrm{d}s + J(\tau, X_\tau^{t,x,\nu}, \nu)].
\end{aligned}$$

■

Theorem 7 *DPP for V*

Assume V is continuous, then

$$V(t,x) = \sup_{\nu \in \mathcal{A}_t} \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\nu_s}, \nu_s)\mathrm{d}s + V(\tau, X_\tau^{t,x,\nu})\right]$$

Proof According to Theorem 6, for fixed $\nu \in \mathcal{A}_t$, we have

$$J(t,x,\nu) \leq \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\nu_s}, \nu_s)\mathrm{d}s + V(\tau, X_\tau^{t,x,\nu})\right],$$

this implies $V(t,x) \leq$ RHS. For another side, we assume there exists an epsilon optimal control $\nu_\varepsilon \in \mathcal{A}_t$, i.e.

$$J(\tau, X_\tau^{t,x,\nu_\varepsilon}, \nu_\varepsilon) \geq V(\tau, X_\tau^{t,x,\nu_\varepsilon}) - \varepsilon.$$

We can further assume $\nu_\varepsilon(s,\omega) = u(s,\omega), \forall (s,\omega) \in [t,\tau]$. Then

$$\begin{aligned} V(t,x) &\geq J(t,x,\nu_\varepsilon) \\ &= \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\nu_s^\varepsilon}, \nu_s^\varepsilon)\mathrm{d}s + J(\tau, X_\tau^{t,x,\nu^\varepsilon}, \nu_s^\varepsilon)\right] \\ &\geq \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,u_s}, u_s)\mathrm{d}s + V(\tau, X_\tau^{t,x,u_s})\right] - \varepsilon. \end{aligned}$$

This provides the required inequality, since the above inequality holds for all $u \in \mathcal{A}$ and ε. ■

For a more rigorous proof of DPP, please refer to [4].

Now, we derive the infinitisimal version of DPP based on Ito's formulae. We first introduce the Hamilton map $H : \mathbf{S} \times \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}$ by

$$H(t,x,g,h) = \sup_{u\in U}\left\{f(t,x,u) + b(t,x,u)\cdot g + \frac{1}{2}\mathbf{tr}(\sigma\sigma^T(t,x,u)h)\right\}.$$

For all $\varphi \in C^{1,2}$, the Ito's formulae implies:

$$\varphi(s, X_s^{t,x,\nu}) - \varphi(t, X_t^{t,x,\nu}) = \int_t^s (\partial_t + \mathcal{L}^{\nu_r})\varphi(r, X_r^{t,x,\nu})\mathrm{d}r + \int_t^s D\varphi(r, X_r^{t,x,\nu})\cdot\sigma(r, X_r^{t,x,\nu}, \nu_r)\mathrm{d}B_r.$$

Theorem 8 *The Hamilton-Jacobian-Bellman Equation*

Assume $V \in C^{1,2}$, then we have

$$-\partial_t V - H(t,x,DV(t,x), D^2V(t,x)) = 0.$$

Theorem 9 *The Verification Argument*

Suppose $v \in C^{1,2}$ solves:

$$\partial_t v + H(t,x,Dv(t,x), D^2v(t,x)) = 0, \quad v(T,x) = g(x)$$

v and f has quadratic growth. If the following conditions are satisfied:

- There exists a maximizer of Hamilton map $\hat{\pi}$

$$(\partial_t + \mathcal{L}^{\hat{\pi}(t,x)})v(t,x) + f(t,x,\hat{\pi}(t,x)) = 0$$

- The SDE has pathwise uniqueness property:

$$\mathrm{d}X_s = b(s,X_s,\pi(s,X_s)\mathrm{d}s + \sigma(s,X_s,\pi(s,X_s))\mathrm{d}B_s, \quad X_0 = x.$$

- $\hat{\nu}_s := \hat{\pi}(s,X_s) \in \mathcal{A}_0$.

Then $v = V$ and $\hat{\nu}_s$ is an optimal control.

Proof For each $\nu \in \mathcal{A}_t$, define $\tau_n^\nu = (T - \frac{1}{n}) \wedge \inf\{s > t : |X_s^{t,x,\nu} - x| \geq n\}$. By Ito's formulae, we have

$$v(t,x) = v(\tau_n^\nu, X_{\tau_n^\nu}^{t,x,\nu}) - \int_t^{\tau_n} (\partial_t + \mathcal{L}^{\nu_r})v(r,X_r^{t,x,\nu})\mathrm{d}r - \int_t^{\tau_n} Dv(r,X_r^{t,x,\nu} \cdot \sigma(r,X_r^{t,x,\nu},\nu_r)\mathrm{d}B_r.$$

In addition, we have the following:

$$\partial_t v(r,X_r^{t,x,\nu}) + \mathcal{L}^{\nu_r} v(r,X_r^{t,x,\nu}) + f(r,X_r^{t,x,\nu},\nu_r) \leq 0.$$

This implies

$$\begin{aligned} v(t,x) &= \mathbb{E}[v(\tau_n^\nu, X_{\tau_n^\nu}^{t,x,\nu}) - \int_t^{\tau_n^\nu} (\partial_t + \mathcal{L}^{\nu_r})v(r,X_r^{t,x,\nu})\mathrm{d}r] \\ &\geq \mathbb{E}[v(\tau_n^\nu, X_{\tau_n^\nu}^{t,x,\nu}) + \int_t^{\tau_n^\nu} f(r,X_r^{t,x,\nu},\nu_r)\mathrm{d}r]. \end{aligned}$$

Now we can apply DCT to conclude:

$$v(t,x) \geq \mathbb{E}[g(X_T^{t,x,\nu}) + \int_t^T f(r,X^{t,x,\nu},\nu_r)\mathrm{d}r], \quad \forall \nu \in \mathcal{A}_t.$$

This implies $v(t,x) \geq V(t,x)$. The desired result immediately follows from the fact $v(t,x) = J(t,x,\hat{\nu}) \leq V(t,x)$. ■

Remark: solution strategy

- Construct a solution v of HJB and check that it is $C^{1,2}$.
- For each (t,x), find a maximizer $\pi(t,x)$.

- ...

2.2 Nonlinear Feynman-Kac Formulae

As discussed above, when the control does not appear in σ, the resulting HJB reduces to a semilinear PDE. In this section, we give a probabilistic description of PDE in the form of

$$\begin{cases} (\partial_t + \mathcal{L})v(t,x) + f(t,x,v(t,x),\sigma^T Dv(t,x)) = 0 \\ v(T,x) = g(x), \end{cases} \tag{1}$$

Theorem 10

Let $v \in C^{1,2}$ be a classical solution of 1. Then the pair (Y,Z) defined by

$$Y_t = v(t,X_t^x), \quad Z_t = \sigma^T Dv(t,X_t^x)$$

is the solution of

$$\begin{cases} X_t = x + \int_0^t b(s,X_s)\mathrm{d}s + \int_0^t \sigma(s,X_s)\mathrm{d}B_s \\ Y_t = g(X_T) + \int_t^T f(s,X_s,Y_s,Z_s)\mathrm{d}s - \int_t^T Z_s \mathrm{d}B_s. \end{cases}$$

Remark: We have already the BSDE in the form of

$$\mathrm{d}_t = -f(t,\omega,Y_t,Z_t)\mathrm{d}s + Z_t \mathrm{d}B_t.$$

Thus we can define

$$f'(t,\omega,y,z) = f(t,X_t(\omega),y,z),$$

thus the decoupled FBSDE reduces to a BSDE with generator f'.

Theorem 11

Suppose $Y^{t,x}, Z^{t,x}$ is the solution of

$$\begin{cases} X_s = x + \int_t^s b(r,X_r)\mathrm{d}r + \int_t^s \sigma(r,X_r)\mathrm{d}B_r \\ Y_s = g(X_T) + \int_s^T f(r,X_r,Y_r,Z_r)\mathrm{d}s - \int_s^T Z_r \mathrm{d}B_r. \end{cases}$$

Define $v(t,x) = Y_t^{t,x}$, and suppose $v \in C^{1,2}$, then v is a classical solution of 1.

2.3 Pontryagin/Stochastic Maximum Principle

Consider $t = 0$ and we define

$$J_x(\nu) = \mathbb{E}[g(X_T^{x,\nu}],$$

and

$$V_x = \sup_{\nu \in \mathcal{A}_0} J_x(v),$$

where X^x, ν is the solution of

$$\mathrm{d}X_t = b(t, X_t, \nu_t)\mathrm{d}t + \sigma(t, X_t, \nu_t)\mathrm{d}B_t.$$

$$H(t, x, y, z, u) = b(t, x, u) \cdot y + \mathbf{tr}(\sigma^T(t, x, u)z).$$

$$\hat{H}(t, x, y, z) = \sup_{u \in U} H(t, x, y, z, u).$$

Theorem 12 *SMP*

Let $\hat{\nu} \in \mathcal{A}_0$ such that

- there is a solution $(\hat{Y}, \hat{Z}) \in \mathbb{H}^2$ of

$$Y_t = Dg(\hat{X}_T) + \int_t^T D_x H(s, \hat{X}_s, Y_s, Z_s)\mathrm{d}s - \int_t^T Z_s \mathrm{d}B_s.$$

- $$H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t) = \sup_{u \in U} H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, u)$$

- $$(x, u) \mapsto H(t, x, \hat{Y}_t, \hat{Z}_t, u) \text{ and } g \text{ are concave,}$$

then $V_x = J_x(\hat{\nu})$.

Proof

$$\begin{aligned}
J_x(\hat{\nu}) - J_x(\nu) &= \mathbb{E}[g(\hat{X}_T) - g(X_T^\nu)] \\
&\geq \mathbb{E}[(\hat{X}_T - X_T^\nu) Dg(\hat{X}_T)] \\
&= \mathbb{E}[(\hat{X}_T - X_T^\nu)\hat{Y}_T] \\
&= \mathbb{E}\left[\int_0^T (H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t) - H(t, X_t^\nu, \hat{Y}_t, \hat{Z}_t, \nu_t) - (\hat{X}_t - X_t^\nu) \cdot D_x H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t))\right] \\
&\geq \mathbb{E}\left[\int_0^T (H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t) - H(t, X_t^\nu, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t) - (\hat{X}_t - X_t^\nu) \cdot D_x H(t, \hat{X}_t, \hat{Y}_t, \hat{Z}_t, \hat{\nu}_t))\right] \\
&\geq 0.
\end{aligned}$$

■

3 Numerical Methods Based on DPE

3.1 Mixed Difference Scheme

See e.g. [4].

3.2 Policy Iteration

Policy iteration, also known as Howard improvement algorithm, is an efficient tool for solving HJB. [2]

For the convenience of calculations, we define H, H^π and f^π as:

$$\begin{cases}
H(t, x, d, h, u) = f(t, x, u) + b(t, x, u) \cdot d + \frac{1}{2}\mathrm{tr}[\sigma\sigma^T(t, x, u)h], \\
H^\pi(t, x, v) = \partial_t v + H(t, x, \nabla v(t, x), \nabla^2 v(t, x), \pi(t, x)), \\
f^\pi(t, x) = f(t, x, \pi(t, x)).
\end{cases} \tag{2}$$

We first initialize $\pi_0 \in U^{\mathbf{S}}$. In the policy evaluation procedure, we solve the following linear PDE:

$$\begin{aligned}
\partial_t v + \left\{ f(t, x, \pi_k(t, x)) + b(t, x, \pi_k(t, x)) \cdot \nabla v + \frac{1}{2}\mathrm{tr}[\sigma\sigma^T(t, x, \pi_k(t, x))\nabla^2 v] \right\} = 0, \\
v(T, x) = g(x),
\end{aligned} \tag{3}$$

and we denote the solution by V^{π_k}. (One can shows that $V^{\pi_k}(t, x) = J(t, x, \nu_k)$, where

$\nu_k(t,\omega) = \pi_k(t, X(t,\omega))$. Next, we improve the control by maximize the Hamilton map:

$$\pi_{k+1}(t,x) = \underset{u\in U}{\operatorname{argmax}} \left\{ f(t,x,u) + b(t,x,u)\cdot \nabla V^{\pi_k}(t,x) + \frac{1}{2}\mathbf{tr}[\sigma\sigma^T(t,x,u)D^2V^{\pi_k}(t,x)] \right\}.$$

Theorem 13

The resulting value function V^{π_k} is monotone increasing.

Proof For all stopping time τ, we have the following:

$$\begin{aligned} V^{\pi_k}(t,x) &= \mathbb{E}\left[V^{\pi_k}(\tau, X_\tau^{t,x,\pi_{k+1}})\right] - \mathbb{E}\left[\int_t^\tau (\partial_t + \mathcal{A}^{\pi_{k+1}})V^{\pi_k}(s, X_s^{t,x,\pi_{k+1}})\mathrm{d}s\right] \\ &= \mathbb{E}\left[V^{\pi_k}(\tau, X_\tau^{t,x,\pi_{k+1}})\right] + \mathbb{E}\left[\int_t^\tau f^{\pi_{k+1}}(s, X_s^{t,x,\pi_{k+1}})\mathrm{d}s\right] - \mathbb{E}\left[\int_t^\tau H^{\pi_{k+1}}(s, X_s^{t,x,\pi_{k+1}}, V^{\pi_k})\mathrm{d}s\right] \\ &\le \mathbb{E}\left[V^{\pi_k}(\tau, X_\tau^{t,x,\pi_{k+1}})\right] + \mathbb{E}\left[\int_t^\tau f^{\pi_{k+1}}(s, X_s^{t,x,\pi_{k+1}})\mathrm{d}s\right], \end{aligned}$$

where $\mathcal{A}^{\pi_k}$ is the generator of the semigroup of $X^{t,x,\pi_{k+1}}$ and the first line follows from Dynkin's formulae. Then we take τ as a deterministic map: $\omega \mapsto T$ and hence

$$\begin{aligned} V^{\pi_k}(t,x) &\le \mathbb{E}\left[\int_t^T f\left(s, X_s^{t,x,\pi_{k+1}}, \pi_{k+1}(s, X_s^{t,x,\pi_{k+1}})\right)\mathrm{d}s + V^{\pi_k}(T, X_T^{t,x,\pi_{k+1}})\right] \\ &= \mathbb{E}\left[\int_t^T f\left(s, X_s^{t,x,\pi_{k+1}}, \pi_{k+1}(s, X_s^{t,x,\pi_{k+1}})\right)\mathrm{d}s + g(X_T^{t,x,\pi_{k+1}})\right] \\ &= V^{\pi_{k+1}}(t,x). \end{aligned}$$

■

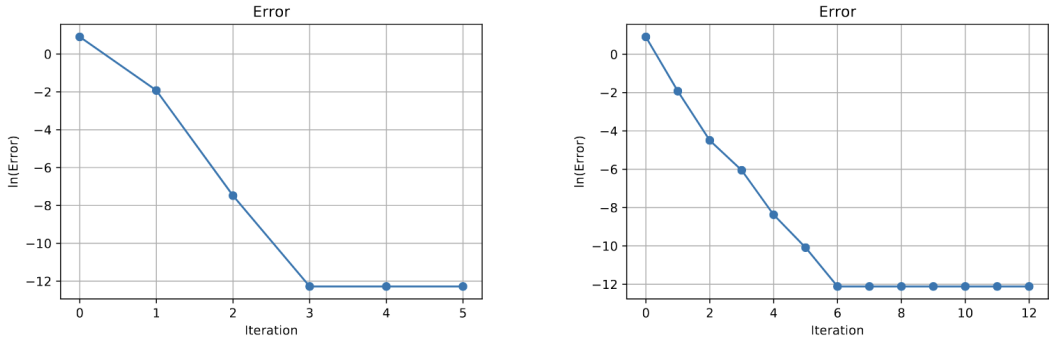

Figure 1: Enter Caption

4 Numerical Methods Based on SMP

4.1 Backward Eular Schemes

See Zhang chapter5.

4.2 Penalty Method

We can convert the following decoupled FBSDE

$$\begin{cases} X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s, \\ Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)\mathrm{d}s - \int_t^T Z_s \mathrm{d}B_s \end{cases}$$

to an optimization problem by penalizing. To be specific,

$$\begin{aligned} \min_{y_0, Z} \quad & \mathbb{E}[|Y_T - g(X_T)|^2] \\ \text{s.t} \quad & X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s, \\ & Y_t = y_0 - \int_0^t f(s, X_s, Y_s, Z_s)\mathrm{d}s + \int_0^t Z_s \mathrm{d}B_s \end{aligned}$$

In practice, we can assume $Z_t = u^\theta(t, X_t)$, and thus the equivalent optimization problem is given by:

$$\begin{aligned} \min_{y_0, \theta} \quad & \mathbb{E}[|Y_T^\theta - g(X_T)|^2] \\ \text{s.t} \quad & X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s, \\ & Y_t = y_0 - \int_0^t f(s, X_s, Y_s, u^\theta(s, X_s))\mathrm{d}s + \int_0^t u^\theta(s, X_s)\mathrm{d}B_s \end{aligned}$$

In this case, there exists F, such that

$$Y_T = F(T, B_\cdot, \theta).$$

A Reinforcement Learning

B PINNs

Consider the following PDE:

$$\begin{cases} Au(x) = 0, & x \in D, \\ u(x) = g(x), & x \in \partial D. \end{cases}$$

We assume the classical solution can be approximated by u^θ. Suppose X_D is a random variable values in D and $X_{\partial D}$ ranges in ∂D. To obtain the best parameter, we need to convert this equation to least square problems. To accomplish this, we define the loss function as

$$L(\theta) = \omega_1 \mathbb{E}[|Au^\theta(X_D)|^2] + \omega_2 \mathbb{E}[|(u^\theta - g)(X_{\partial D})|^2].$$

By minimizing $L(\theta)$(through SGD or LBFGS), we obtain an approximated solution. For more information about PINNs, we refer to [3].

References

[1] R. Carmona. *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*. SIAM, 2016.

[2] B. Kerimkulov, D. Siska, and L. Szpruch. Exponential convergence and stability of howard's policy improvement algorithm for controlled diffusions. *SIAM Journal on Control and Optimization*, 58(3):1314–1340, 2020.

[3] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[4] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media, 2012.