Martingale Based Deep Neural Networks for high-dimensional PDE and Optimal Stochastic Control Problems [1]

W. Cai

Southern Methodist University

Illinois Institute of Technology
October 31, 2023

[1] joint work with Andrew He and Daniel Margolis

Outline of talk

- SDE based Deep Neural networks (DNNs) for PDEs
- Martingale problems for PDEs
- Martingale based DeepMartNet
- Numerical Results for PDEs
- DeepMartNet for Optimal Stochastic Controls
- Future work


Andrew He and Daniel Margolis

Review of SDE based Neural Networks for PDEs

Consider an Initial Value Problem (IVP) for Quasi-linear Problems

$$\partial_t u + \frac{1}{2}\mathrm{Tr}[\sigma\sigma^T \nabla\nabla u] + \mu \cdot \nabla u = \phi \tag{1}$$

with the condition $u(T, x) = g(x)$. Problem is to find solution at $x, t = 0$ $u(0, x)$, and the solution is related to FBSDE (Pardoux-Peng, 1990)

$$\begin{aligned} dX_t &= \mu(t, X_t, Y_t, Z_t)dt + \sigma(t, X_t, Y_t)dW_t, \\ X_0 &= \xi, \end{aligned} \tag{2}$$

$$\begin{aligned} dY_t &= \phi(t, X_t, Y_t, Z_t)dt + Z_t^T\sigma(t, X_t, Y_t)dW_t, \\ Y_T &= g(X_T), \end{aligned} \tag{3}$$

Namely,

$$Y_t = u(t, X_t), \quad Z_t = \nabla u(t, X_t). \tag{4}$$

Example I: Deep BSDE by J. Han, E, et al. (2016)

The Deep BSDE trains the network with input $X_0 = \xi$ and output $Y_0 = u(0, X)$. Apply the Euler–Maruyama scheme (EM) to the FBSDE (2) and (3), respectively,

$$X_{n+1} \approx X_n + \mu(t_n, X_n, Y_n, Z_n)\Delta t_n + \sigma(t_n, X_n, Y_n)\Delta W_n, \tag{5}$$

$$Y_{n+1} \approx Y_n + \phi(t_n, X_n, Y_n, Z_n)\Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n)\Delta W_n. \tag{6}$$

The missing Z_{n+1} will be approximated by a NN at t_{n+1}

$$\nabla u(t_n, X_n|\theta_n) \mapsto Z_n = \nabla u(t_n, X_n). \tag{7}$$

Loss function: with Ensemble average approximation

$$Loss_{bsde}(Y_0, \theta) = \mathbb{E} \left\| u(T, X_T) - g(X_T) \right\|^2. \tag{8}$$

where

$$u(T, X_T) = Y_N$$

Trainable parameters are $\{Y_0, \theta_n, n = 1, \cdots, N\}$.

Method 2: FBSNNs by M. Raissi

The FBSNNs trains the network with input pair $(t, X_0 = \xi)$ and output a NN $u_\theta(t, x)$.
Loss function is based on the differnence of two discrete Markov chains:

- Markov Chain one

$$\begin{aligned} X_{n+1} &= X_n + \mu(t_n, X_n, Y_n, Z_n)\Delta t_n + \sigma(t_n, X_n, Y_n)\Delta W_n, \\ Y_{n+1} &= u(t_{n+1}, X_{n+1}), \\ Z_{n+1} &= \nabla u(t_{n+1}, X_{n+1}). \end{aligned} \tag{9}$$

- Reference process by using the EM scheme

$$Y^\star_{n+1} = Y_n + \phi(t_n, X_n, Y_n, Z_n)\Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n)\Delta W_n. \tag{10}$$

- Loss function: a Monte Carlo approximation of

$$Loss_{fbsnn} = \mathbb{E}\left[\sum_{n=1}^{N} \|Y_n - Y_n^\star\|^2 + \|Y_N - g(X_N)\|^2 + \|Z_N - \nabla g(X_N)\|^2\right]. \tag{11}$$

Method 2 (continued): FBSNNs (An improvement by Zhang, Cai 2023)

- Markov chain one

$$
\begin{aligned}
X_{n+1} &= X_n + \mu(t_n, X_n, Y_n, Z_n)\Delta t_n + \sigma(t_n, X_n, Y_n)\Delta W_n, \\
Y_{n+1} &= Y_n + \phi(t_n, X_n, Y_n, Z_n)\Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n)\Delta W_n, \\
Z_{n+1} &= \nabla u(t_{n+1}, X_{n+1}).
\end{aligned} \tag{12}
$$

- Markov chain two

$$
Y^{\star}_{n+1} = u(t_{n+1}, X_{n+1}). \tag{13}
$$

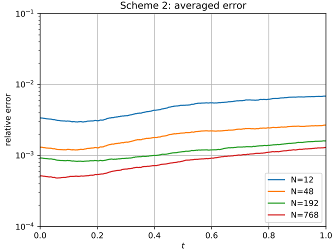
- The loss function is a Monte Carlo approximation of

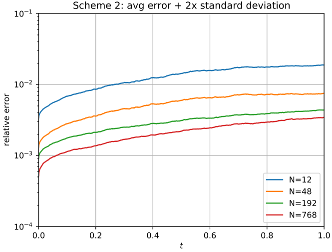
$$
\mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\|Y_n - Y_n^{\star}\|^2 + 0.02\,\|Y_N^{\star} - g(X_N)\|^2 + 0.02\,\|Z_N - \nabla g(X_N)\|^2\right]. \tag{14}
$$

Half-order convergence of $u_{\theta(x,t)}$ is observed due to the fact both processes defined are Markov chain.

1/2-order convergence for Extrapolation for better accuracy of Y_0

With modifed verions of FBSNN we have the error plots for $N = 12$, $N = 48$, $N = 192$ and $N = 768$:





u_θ^N - the trained network with N number of time steps.

	Modified FBSNN-1		Modified FBSNN-2	
N	u_θ^N	u_{ex}^N	u_θ^N	u_{ex}^N
12	2.91e-03		2.82e-03	
48	1.67e-03	**4.29e-04**	1.13e-03	5.57e-04
192	7.58e-04	**1.53e-04**	8.43e-04	5.55e-04
768	6.77e-04	5.97e-04	5.96e-04	3.49e-04

Table: Relative error of Y_0 from the network approximation and extrapolation.

Method III. fixed point of semi-group formulation for Eigenvalue Problem (Lu, etal, 2020)

$$\mathcal{L}\Psi \doteq \left(\frac{1}{2}\mathrm{Tr}[\sigma\sigma^T\nabla\nabla] + \mu\cdot\nabla\right)\Psi = \lambda\Psi$$

Reformulated as backward parabolic PDE

$$\partial_t u(t,x) + \mathcal{L}u(t,x) - \lambda u(t,x) = 0$$

$$u(T,x) = \Psi(x) \approx \Psi_\theta(x),$$

So

$$u(T-t,\cdot) = P_t^\lambda\Psi, \quad P_T^\lambda\Psi = \Psi$$

Loss function $= ||P_T^\lambda\Psi - \Psi||^2$
Evolution of PDE solution done by two SDEs

$$X_{n+1} = X_n + \sigma\Delta B_n$$

$$u_{n+1} = u_n + (\lambda\Psi_\theta - \mu^T\nabla\Psi_\theta)(X_n)\Delta t + \nabla\Psi_\theta(X_n)\Delta B_n$$

$$Loss_{semigroup} = E_{X_0\approx\pi_0}[|u_N - \Psi_\theta(X_N)|^2]$$

Martingale Problem Formulation for BVP of PDEs

$$\mathcal{L} = \mu^\top \nabla + \frac{1}{2} Tr(\sigma\sigma^\top \nabla\nabla^\top) = \mu^\top \nabla + \frac{1}{2} Tr(A\nabla\nabla^\top) \tag{15}$$

As a generator for SDE

$$\begin{aligned} d\mathbf{X}_t &= \mu(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t) \cdot d\mathbf{B}_{,t} \\ \mathbf{X}_t &= \mathbf{x}_0 \in D, \end{aligned} \tag{16}$$

Consider the Robin BVP

$$\begin{aligned} \mathcal{L}u + V(\mathbf{x}, u, \nabla u) &= f(x, u), \quad \mathbf{x} \in D \subset R^d, \\ \Gamma(u) = \gamma^\top \cdot \nabla u + cu &= g, \quad \mathbf{x} \in \partial D, \end{aligned} \tag{17}$$

where the unit vector

$$\gamma(x) = \frac{1}{2} A \cdot n,$$

and a shorthand

$$v(\mathbf{x}) = V(\mathbf{x}, u(\mathbf{x}), \nabla u(\mathbf{x}))$$

Reflecting Diffusion and Skorohod problem

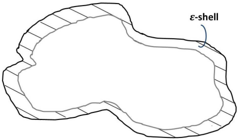
(Skorohod problem): A pair $(X^{ref}(t), L(t))$ is a solution to the Skorohod problem $S(X; D)$ if the following conditions are satisfied:

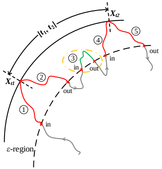
1. X^{ref} is a path in $\bar{D}$;
2. (**local time**) $L(t)$ is a nondecreasing function which increases only when $X^{ref} \in \partial D$, namely,

$$L(t) = \int_0^t I_{\partial D}(X^{ref}(s))L(ds), \tag{18}$$

3. The Skorohod equation holds:

$$S(X; D): \qquad X^{ref}(t) = X(t) - \int_0^t \gamma(X^{ref}(s))L(ds). \tag{19}$$





$$L(t) \approx \frac{\int_0^t I_{D_\epsilon}(X_s)ds}{2\epsilon}. \tag{20}$$

Martingale for BVP of Elliptic PDEs

$$\text{Denoting} \quad X(t) \leftarrow X^{ref}(t) \quad (i.e. semi - Martingale) \tag{21}$$

Using the Ito formula for the semi-martingale $X(t)$

$$du(X(t)) = \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}(X(t))dX_i(t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(X(t))\frac{\partial^2 u}{\partial x_i \partial x_j}(X(t))dt$$

$$\begin{aligned} du(X(t)) &= \mathcal{L}u(X(t)) - \gamma^{\intercal} \cdot \nabla u(u(X_t))L(dt) + \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_{ij}\frac{\partial u}{\partial x_i}(X(t))dB_i(t) \\ &= f(X(t), u(X(t)) - V(X(t), u(X(t)), \nabla u(X(t))) \\ &- [g(X(t)) - cu(X(t))]\, L(dt) + \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_{ij}\frac{\partial u}{\partial x_i}(X(t))dB_i(t) \quad (Martingale) \end{aligned}$$

A Martingale M_t^u is found to be

$$\begin{aligned} M_t^u &\doteq u(X_t) - u(X_0) - \int_0^t [f(X_s, u(X_s)) - V(X_s, u(X_s), \nabla u(X_s))]\, ds \\ &+ \int_0^t [g(X_s) - cu(X_s)]\, L(ds) = \int_0^t \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_{ij}\frac{\partial u}{\partial x_i}(X_s)dB_i(s), \end{aligned}$$

Dirichlet BVP

For Dirichlet problem of (48) with a boundary condition

$$\Gamma[u] = u = g, x \in \partial D, \tag{22}$$

the underlying diffusion process is the original diffusion process (48), but killed at the boundary at the first exit time

$$\tau_D = \inf\{t, X_t \in \partial D\}, \tag{23}$$

and it can be shown that in fact

$$\tau_D = \inf\{t > 0, L(t) > 0\}. \tag{24}$$

$M^u_{t\wedge\tau_D}$ remains a Martingale, which will not involve the integral with respect to local time L(t), i.e.

$$M^u_{t\wedge\tau_D} = u(X_{t\wedge\tau_D}) - u(X_0) - \int_0^{t\wedge\tau_D} [f(X_s, u(X_s)) - V(X_s, u(X_s), \nabla u(X_s))]ds. \tag{25}$$

For the case of linear PDE, i.e. $f(x,u) = f(x), V = 0$, by taking expectation, we get Feynman-Kac formula for Dirichlet problem

$$u(x) = E[g(X_{\tau_D})] - E[\int_0^{\tau_D} f(X_s)ds]. \tag{26}$$

Local Solution by Feynman-kac formula for mixed Laplace BVP

$$\Delta u = 0 \qquad \text{in} \quad \Omega \backslash \Omega_0, \tag{27}$$

$$\frac{\partial u}{\partial n} - cu = \phi_3(x) \qquad \text{on} \quad \Gamma_3 = \cup_{l=1}^{8} E_l, \tag{28}$$

$$\frac{\partial u}{\partial n} = \phi_2(x) \qquad \text{on} \quad \Gamma_2 = \partial\Omega \backslash \Gamma_3, \tag{29}$$

$$u = \phi_1(x) \qquad \text{on} \quad \Gamma_1 = \partial\Omega_0, \tag{30}$$

$$u_{Mix}(x) = E^x \left\{ \int_0^{\tau_{\Gamma_1}} \hat{e}_c(t)\phi_{2,3}(X_t)dL(t) \right\} + E^x(\hat{e}_c(\tau_{\Gamma_1})\phi_1(X_{\tau_{\Gamma_1}})). \tag{31}$$

$$\hat{e}_c(t) := e^{\int_0^t c(X_s)dL(s)}. \tag{32}$$

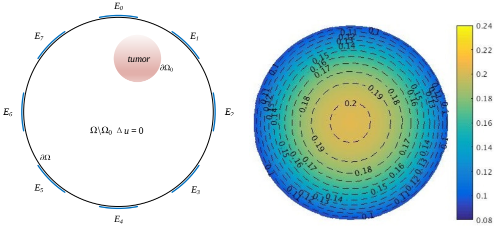


Figure: Potential on one electrode by Feynman-Kac formula with reflecting Brownian Motion (Ding, Cai, et al, 2023, JCP)

Martingale Problem Formulation of BVP

- A probabilistic weak form of the Robin BVPs is that M_t^u is a Martingale.
- Classic weak form: For every test function $\phi(x) \in C^2_{\partial D} = \{\phi : \phi \in C^2(D) \cap C^1(\overline{D}), (\gamma \cdot \nabla + c)\, \phi = 0\}$, we have

$$\int_D u(x)\mathcal{L}^*\phi dx = \int_D [f(x,u(x)) - V(x,u(x),\nabla u(x))]\, \phi(x) dx + \int_{\partial D} \phi(x)[\mu^\intercal \cdot nu + g(x) - cu(x)] ds_x, \tag{33}$$

where

$$\mathcal{L}^*\phi = \frac{1}{2} Tr(\nabla\nabla^\intercal A)\phi - div(\mu\phi). \tag{34}$$

The equivalence between the probabilistic weak form and the classic weak form are proven for the Schrodinger operator $\mathcal{L}u = \frac{1}{2}\Delta u + qu$ for Neumann problem (Hsu 1984) and Robin problem (V. G. Papanicolaou 1990).

For simplicity of discussion, let us assume that $s \leq t \leq \tau_D$, by the Martingale property of $M_t = M_t^u$ of (25), we have

$$E[M_t|\mathcal{F}_s] = M_s, \tag{35}$$

which implies for any measurable set $A \in \mathcal{F}_s$,

$$E[M_t|A] = M_s = E[M_s|A], \tag{36}$$

thus,

$$E[(M_t - M_s)\,|A] = 0, \tag{37}$$

i,e,

$$\int_A (M_t - M_s)\, P(d\omega) = 0, \tag{38}$$

Martingale Loss

For a given time interval $[0, T]$, we define a partition

$$0 = t_0 < t_1 < \cdots < t_i < t_{i+1} < \cdots < t_N = T, \tag{39}$$

and M-discrete realizations

$$\Omega' = \{\omega_m\}_{m=1}^{M} \subset \Omega \tag{40}$$

of the Ito process using Euler-Maruyama scheme with M-realizations of the Brownian motions $\mathbf{B}_i^{(m)}$, $0 \leq m \leq M$,

$$\mathbf{X}_i^{(m)}(\omega_m) \sim X(t_i, \omega_m), 0 \leq i \leq N,$$

where

$$\mathbf{X}_{i+1}^{(m)} = \mathbf{X}_i^{(m)} + \mu(\mathbf{X}_i^{(m)})\Delta t_i + \sigma(\mathbf{X}_\mathbf{i}^{(\mathbf{m})}) \cdot \Delta \mathbf{B}_i^{(m)}, \tag{41}$$

$$\mathbf{X}_0^{(m)} = \mathbf{x}_0 \tag{42}$$

where $\Delta t_i = t_{i+1} - t_i$,

$$\Delta \mathbf{B}_i^{(m)} = \mathbf{B}_{i+1}^{(m)} - \mathbf{B}_i^{(m)}.$$

Martingale Loss for DeepMartNet (Dirichlet BVP)

The increment of the M_t over $[t_i, t_{i+k}]$ can be approximated by

$$\begin{aligned} M_{t_{i+k}} - M_{t_i} =& u(\mathbf{X}_{i+k}) - u(\mathbf{X}_i) - \int_{t_i}^{t_{i+k}} \mathcal{L}u(\mathbf{X}_z)dz \\ \doteq& u(\mathbf{X}_{i+k}) - u(\mathbf{X}_i) - \Delta t \sum_{l=0}^{k} \omega_l \mathcal{L}u(\mathbf{X}_{i+l}) \\ =& u(\mathbf{X}_{i+k}) - u(\mathbf{X}_i) - \Delta t \sum_{l=0}^{k} \omega_l (f(\mathbf{X}_{i+l}, u(\mathbf{X}_{i+l})) - v(\mathbf{X}_{i+l})). \quad (43) \end{aligned}$$

Adding back the exit time τ_D, note that

$$M_{t_{i+k}\wedge\tau_D} - M_{t_i\wedge\tau_D} = u(\mathbf{X}_{t_{i+k}\wedge\tau_D}) - u(\mathbf{X}_{t_i\wedge\tau_D}) - \int_{t_i\wedge\tau_D}^{t_{i+k}\wedge\tau_D} \mathcal{L}u(\mathbf{X}_z)dz = 0$$

if both $t_{i+k}, t_i \geq \tau_D$.

DeepMartNet for Dirichlet BVP

As $E[M_{t_{i+k}} - M_{t_i}] \approx 0$, for a randomly selected $A_i \in \Omega' = \mathcal{F}_{t_i}$ (mini-batches)

$$Loss_{mart}(\theta) = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{i=0}^{N-1}(M_{t_{i+k}\wedge\tau_D} - M_{t_i\wedge\tau_D})^2 \tag{44}$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}\frac{I(t_i \leq \tau_D)}{|A_i|^2}\sum_{m=1}^{|A_i|}\Big(u_\theta(\mathbf{X}_{i+k}^{(m)}) - u_\theta(\mathbf{X}_i^{(m)}) -$$

$$\Delta t\sum_{l=0}^{k}\omega_l(f(\mathbf{X}_{i+l}^{(m)}, u_\theta(\mathbf{X}_{i+l}^{(m)})) - v_\theta(\mathbf{X}_{i+l}^{(m)}))\Big)^2 \tag{45}$$

DeepMartNet solution $- u_{\theta^*}(x), \theta^* = argminLoss_{mart}(\theta)$

Martingale Loss for Robin BVP

$$Loss_{mart}(\theta) = \frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\Big(u_\theta(\mathbf{X}_{i+k}^{(m)}) - u_\theta(\mathbf{X}_i^{(m)}) -$$

$$\Delta t \sum_{l=0}^{k}\omega_l(f(\mathbf{X}_{i+l}^{(m)}, u_\theta(\mathbf{X}_{i+l}^{(m)})) - v_\theta(\mathbf{X}_{i+l}^{(m)})) \quad (46)$$

$$-\sum_{l=0}^{k}\omega_l(g(\mathbf{X}_{i+l}^{(m)}) - cu_\theta(\mathbf{X}_{i+l}^{(m)}))L(\Delta t_{i+l})\Big)^2, \quad (47)$$

where

$$v_\theta(\mathbf{x}) = V(\mathbf{x}, u_\theta(\mathbf{x}), \nabla u_\theta(\mathbf{x})).$$

DeepNetMart for Dirichlet eigenvalue problems

When the RHS $f(x,u) = \lambda u$, we will have an eigenvalue problem

$$\mathcal{L}u + V(\mathbf{x}, u, \nabla u) = \lambda u, \quad \mathbf{x} \in D \subset R^d, \tag{48}$$
$$\Gamma(u) = u = 0, \quad \mathbf{x} \in \partial D,$$

and the Martingale loss becomes

$$Loss_{mart}(\lambda, \theta) = \frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\Big(u_\theta(\mathbf{X}_{i+k}^{(m)}) - u_\theta(\mathbf{X}_i^{(m)}) - \Delta t\sum_{l=0}^{k}\omega_l(\lambda u_\theta(\mathbf{X}_{i+l}^{(m)}) - v_\theta(\mathbf{X}_{i+l}^{(m)}))\Big)^2 \tag{49}$$

DeepMartNet eigenvalue solution – $(\lambda^*, u_{\theta^*}(x))$

$$(\lambda^*, \theta^*) = argmin \quad Loss_{mart}(\lambda, \theta)$$

Mini-Batchs in SGD and Martingale Property

- Martingale property implies that for any measurable set $A \in \mathcal{F}_s$,

$$E[M_t|A] = M_s, \quad (50)$$

  requires the equation holds for any random member of $\mathcal{F}_s$, a native mechanism for the mini-batch in the SGD. Martingale based NN provides an ideal fit for deep learning of high-d PDEs.

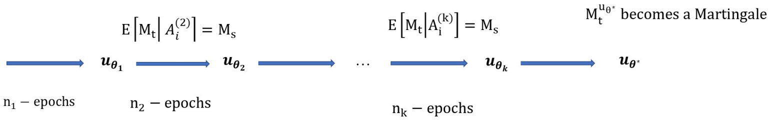


Figure: DeepMartNet training and Martingale property

- (Size of mini-batch A_i in) $\int_{A_i} (M_t - M_s)\, P(d\omega) = 0.$
  The size of mini-batch $|A_i|$ should be large enough to give an accurate sampling of the continuous distribution

$$M/20 \leq |A_i| \leq M/5$$

  where M is the total number of paths used.

Beyond Feynman-Kac formula

- Traditional (Pre-ML) Feynman-Kac formula based Monte Carlo method gives solution at one single point x_0, where the paths originate

$$u(x_0) = E[g(X_{\tau_D})] - E[\int_0^{\tau_D} f(X_s)ds]. \tag{51}$$

or

$$u_{Mix}(x_0) = E\left\{\int_0^{\tau_{\Gamma_1}} \hat{e}_c(t)\phi_{2,3}(X_t)dL(t)\right\} + E(\hat{e}_c(\tau_{\Gamma_1})\phi_1(X_{\tau_{\Gamma_1}})). \tag{52}$$

- DeepMartNet uses the same number of paths starting from one point x_0 to produce global solution of the PDEs.

$$\text{DeepMartNet solution} - u_{\theta^*}(x), \theta^* = argminLoss_{mart}(\theta)$$

$$Loss_{mart}(\theta) = \frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\Big(u_\theta(\mathbf{X}_{i+k}^{(m)}) - u_\theta(\mathbf{X}_i^{(m)}) - \Delta t\sum_{l=0}^{k}\omega_l(f(\mathbf{X}_{i+l}^{(m)}, u_\theta(\mathbf{X}_{i+l}^{(m)})) - v_\theta(\mathbf{X}_{i+l}^{(m)}))\Big)^2. \tag{53}$$

where $\mathbf{X}_0^{(m)} = x_0, m = 1, \cdots, M$.

Numerical Results (Dirichlet BVP of the Poisson-Boltzmann Equation)

$$\begin{cases} \Delta u(x) + cu(x) = f(x), & x \in D \\ u(x) = g(x), & x \in \partial D \end{cases} \tag{54}$$

where $c < 0$ with an high-d exact solution given

$$u(x) = \sum_{i=1}^{d} \cos(\omega x_i), \quad \omega = 2. \tag{55}$$

For Brownian motions $W^{(j)}, j = 1, \cdots, M$ We define

$$Loss^{\text{mart}}(\theta) := \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{|A_i|^2} \left(\sum_{j=1}^{|A_i|} u_\theta(W^{(j)}_{t_{i+1}}) - u_\theta(W^{(j)}_{t_i}) \right. \tag{56}$$

$$\left. - \frac{1}{2} \left(f(W^{(j)}_{t_i}) - cu(W^{(j)}_{t_i}) \right) \mathbb{I}(t_i \leq \tau^{(j)}_D) \Delta t \right)^2, \tag{57}$$

$$Loss^{\text{F-K}} = (u_\theta(x_0) - u(x_0))^2$$

$$u(x_0) \approx \frac{1}{M} \sum_{j=1}^{M} \left(g\left(W^{(j)}_{\tau_D}\right) \mathrm{e}^{\frac{c\tau_D}{2}} + \frac{1}{2} \sum_{i=0}^{N-1} f\left(W^{(j)}_{t_i}\right) \mathrm{e}^{\frac{ct_i}{2}} \mathbb{I}\left(t_i \leq \tau^{(j)}_D\right) \Delta t \right),$$

PBE in a d=20 dim unit cube

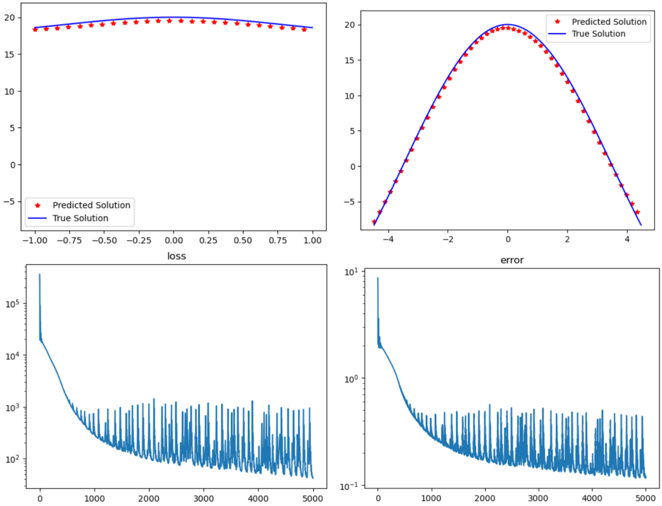
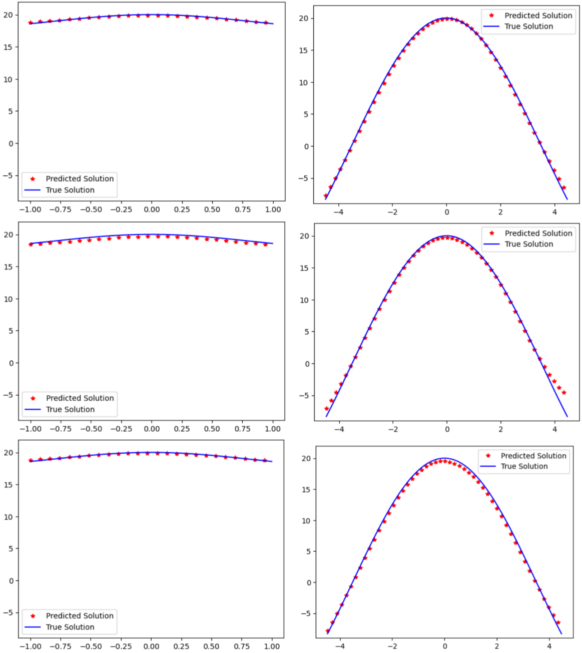


Figure: $D = [-1, 1]^d, d = 20$. The total number of paths is $M = 100,000$, and mini-batch size $M_0 = 1000$; $\Delta t = 0.01$. (Upper left): true and predicted value of u at the diagonal of $\mathbb{R}^d$; (Upper right): true and predicted value of u at the first coordinate axis of $\mathbb{R}^d$. (Lower left): The loss L history; (lower right): The relative error L^2 history

Effect of starting point $x_0 = (l, 0, \cdots, 0), l = 0.1, 0.3, 0.7$



PBE in a d=100 dim unit ball

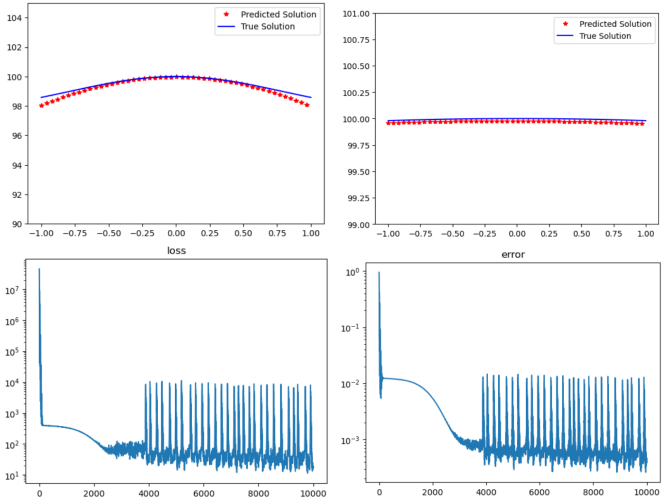


Figure: D = unit ball $\in R^100$. The total number of paths is $M = 100,000$, and mini-batch size $M_0 = 1000$; $\Delta t = 0.005$. (Upper left): true and predicted value of u at the diagonal of $\mathbb{R}^d$; (Upper right): true and predicted value of u at the first coordinate axis of $\mathbb{R}^d$. (Lower left): The loss L history; (lower right): The relative error L^2 history

Eigenvalue Problem for Fokker-Planck equations

$$\mathcal{L}'\psi = -\Delta\psi - \nabla\cdot(\psi\nabla W) + c\psi = -\Delta\psi - \nabla W\cdot\nabla\psi - \Delta W\psi + c\psi = (\lambda_0 + c)\psi = \lambda_c\psi, \tag{58}$$

where the eigenfunction for the $\lambda_c = c$- eigenvalue is simply

$$\psi(x) = e^{-W(x)}. \tag{59}$$

Re-written as

$$\mathcal{L}\psi = \frac{1}{2}\Delta\psi + \frac{1}{2}\nabla W\cdot\nabla\psi = -\left(\frac{1}{2}\Delta W - \frac{1}{2}c + \lambda\right)\psi. \tag{60}$$

Set generator for the SDE $\mathcal{L}$ with drift and diffusion as

$$\mu = \frac{1}{2}\nabla W \quad \text{and} \quad \sigma = I_{d\times d},$$

And, the V is given by

$$V = -\frac{1}{2}\Delta W. \tag{61}$$

The Martingale loss for this case will be

$$loss_1 = \frac{1}{\Delta t}\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\Big(u_\theta(\mathbf{x}_{\mathbf{i+1}}^{(m)}) - u_\theta(\mathbf{x}_{\mathbf{i}}^{(m)}) \tag{62}$$

$$+(\frac{1}{2}\Delta W(\mathbf{x}_{\mathbf{i}}^{(m)}) - \frac{1}{2}c + \lambda)u_\theta(\mathbf{x}_{\mathbf{i}}^{(m)})\Delta t\Big)^2. \tag{63}$$

Quadratic potential $W(x) = ||x||^2, x \in R^d$

a 5 Dimensional Fokker Planck

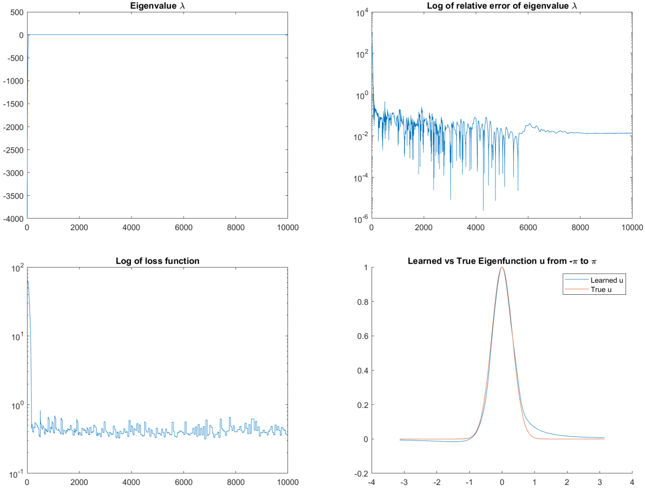


Figure: Eigenvalue $c = 10$, trapezoid $k = 3$, number of paths $M = 9000$, and number of time steps $N = 1350$; relative eigenvalue error of 0.013, an $L^2_{RMS} = 0.025$ with 10000 epochs.

25 Dimensional Fokker Planck

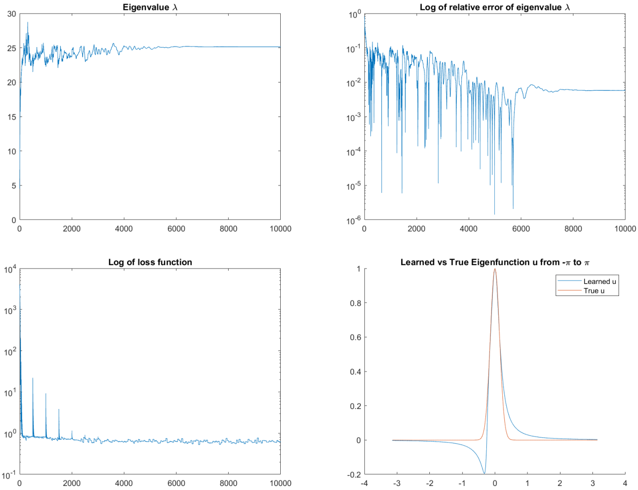


Figure: Eigenvalue $c = 50$, trapezoid $k = 3$, number of paths $M = 30000$, and number of time steps $N = 1800$; relative eigenvalue error of 0.0057, an $L^2_{RMS} = 0.058$ with 10000 epochs.

a 50 Dimensional Fokker Planck

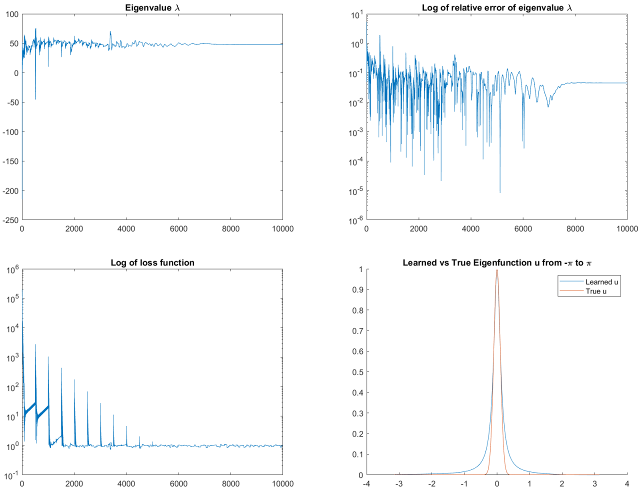


Figure: Eigenvalue $c = 100$, trapezoid $k = 3$, number of paths $M = 7500$, and number of time steps $N = 900$; relative eigenvalue error of 0.045, an $L^2_{RMS} = 0.041$ with 10000 epochs.

200 Dimensional Fokker Planck

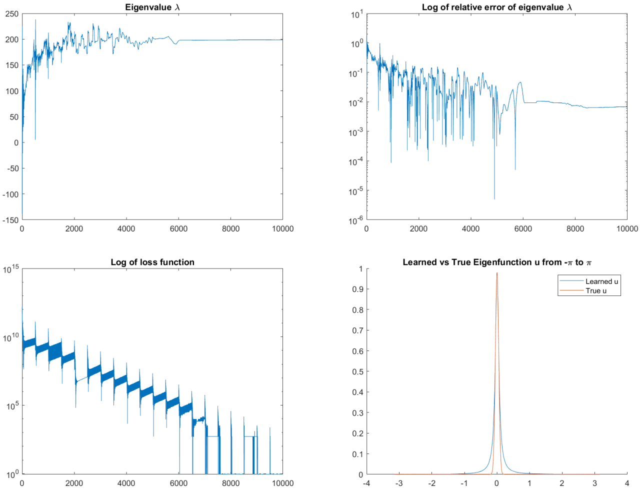


Figure: Eigenvalue $c = 400$, trapezoid $k = 3$, number of paths $M = 24000$, and number of time steps $N = 1350$; relative eigenvalue error of 0.0067, an $L^2_{RMS} = 0.029$ with 10000 epochs.

DeepMartNet for Optimal Stochastic Control

Feedback control: Consider SDE

$$d\mathbf{X}_t = \mu(t, \mathbf{X}_t, u_t)dt + \sigma(t, \mathbf{X}_t) \cdot d\mathbf{B}_t, \quad 0 \le t \le T \tag{64}$$

- $u_t \in \mathcal{U}$, $\{\mathcal{F}_t\}_{t \ge 0}$-predictable processes taking values in $U \subset R^m$.
- The running cost
$$c : \Omega \times [0, T] \times U \to R, \tag{65}$$
- A feedback control
$$c(\omega, t, u) = c(\mathbf{X}_t(\omega), t, u), \tag{66}$$
- Terminal cost
$$\xi(\omega) = \xi(\mathbf{X}_T(\omega)) \tag{67}$$

$u_t \in \mathcal{U}$

Optimal stochastic control

The optimal control problem: find a control u^*

$$u^* = \arg\inf_{u \in \mathcal{U}} J(u). \tag{68}$$

where the total expected cost is then defined by

$$J(u) = E_u[\xi + \int_{[0,T]} c(\mathbf{X}_t(\omega), t, u_t)dt]. \tag{69}$$

Define the expected remaining cost for a given control u

$$J(\omega, t, u) = E_u[\xi(\mathbf{X}_T(\omega)) + \int_{[t,T]} c(\mathbf{X}_t(\omega), t, u_t)dt|\mathcal{F}_t] \tag{70}$$

and a value process

$$V_t(\omega) = \inf_{u \in \mathcal{U}} J(\omega, t, u), \text{ and } E[V_0] = \inf_{u \in \mathcal{U}} J(u) = J(u^*), \tag{71}$$

and a cost process

$$M_t^u(\omega) = \int_{[0,t]} c(\mathbf{X}_s(\omega), s, u_s)ds + V_t(\omega). \tag{72}$$

Martingale Optimality Principle

The Martingale optimality principle is stated in the following theorem (Elliot, 2015).

Theorem

(Martingale optimality principle) M_t^u *is a* P^u*-super-martingale.* M_t^u *is a* P^u*-martingale if and only if control* $u = u^*$ *(the optimal control),and*

$$E[V_0] = E_u[M_0^{u^*}] = \inf_{u \in \mathcal{U}} J(u).$$

BSDE for Value Process $V_t(\omega)$

The value process $V_t(\omega)$ satisfies a backward SDE (BSDE)

$$\begin{cases} dV_t = -H(t, \mathbf{X}_t, \mathbf{Z}_t)dt + \mathbf{Z}_t dB_t, 0 \le t < T \\ V_T(\omega) = \xi(\mathbf{X}_T(\omega)) \end{cases}, \tag{73}$$

where the Hamiltanian$H(t, \mathbf{x}, \mathbf{z}) = \inf_{u \in \mathcal{U}} f(t, \mathbf{x}, \mathbf{z}; u)$

$$f(t, \mathbf{x}, \mathbf{z}; u) = c(\mathbf{x}, t, u) + \mathbf{z}\alpha(t, \mathbf{x}, u), \quad \alpha(t, \mathbf{x}, u) = \sigma^{-1}(t, \mathbf{x})\mu(t, \mathbf{x}, u).$$

From Pardoux-Peng BSDE theory

$$V_t(\omega) = v(t, \mathbf{X}_t(\omega))$$
$$Z_t(\omega) = \nabla v(t, \mathbf{X}_t(\omega))\sigma(t, \mathbf{X}_t(\omega))$$

where the value function $v(t, \mathbf{x})$ satisfies a HJB equation

$$\begin{cases} 0 = \frac{\partial v}{\partial t}(t, \mathbf{x}) + \mathcal{L}v(t, \mathbf{x}) + H(t, \mathbf{x}, \nabla_x v\sigma(t, \mathbf{x})), \quad 0 \le t < T, \mathbf{x} \in R^d \\ v(T, \mathbf{x}) = \xi(\mathbf{x}) \end{cases}. \tag{74}$$

DeepMartNet for Optimal Stochastic feedback Control

Approximate the optimal control by a neural network

$$u_t(\omega) = u_t(\mathbf{X}(\omega)) \sim u_{\theta_1}(t, \mathbf{X}(\omega)), \tag{75}$$

and the value function by another network

$$v(t, \mathbf{x}) \sim v_{\theta_2}(t, \mathbf{x}). \tag{76}$$

$$l(\theta_1, \theta_2) = l_{ctr}(\theta_1) + l_{val}(\theta_2)$$

where

$$\begin{aligned} l_{ctr}(\theta_1) &= \frac{1}{N}\sum_{i=0}^{N-1}\left(E[M^u_{t_{i+1}} - M^u_{t_i}]\right)^2 \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\left(c(X_{t_i}, t_i, u_{\theta_1}(t_i, \mathbf{X}_i^{(m)}))\Delta t_i + v_{\theta_2}(t_{i+1}, \mathbf{X}_{i+1}^{(m)}) - v_{\theta_2}(t_i, \mathbf{X}_i^{(m)})\right)^2 \end{aligned} \tag{77}$$

$$l_{val}(\theta_2) = \frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{|A_i|^2}\sum_{m=1}^{|A_i|}\left(\begin{array}{c} v_{\theta_2}(t_{i+1}, \mathbf{X}_{i+1}^{(m)}) - v_{\theta_2}(t_i, \mathbf{X}_i^{(m)}) + \\ H(t_i, \mathbf{X}_i^{(m)}, \nabla_x v_{\theta_2}(t_i, \mathbf{X}_i^{(m)})\sigma(t, \mathbf{X}_i^{(m)}))\Delta t_i \end{array}\right)\right)^2 \tag{78}$$

$$+\beta\frac{1}{M}\sum_{m=1}^{M}(v_{\theta_2}(T, \mathbf{X}_N^{(m)}) - \xi(\mathbf{X}_N^{(m)}))^2.$$

Future work

1. Apply DeepMartNet for various PDEs problem HJB, Black-Scholes, Fokker-Planck equation, Committor functions in TPT
2. ground state of many electron systems, non-Hermitian operator (electron under magnetic field)
3. Stochastic controls, financial applications