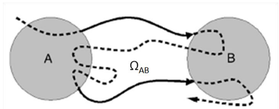
Neural Network Approaches to the Committor Problem

Luke Triplett (lft4@duke.edu), William Clark (wiclark1@davidson.edu) , Dmitry Pinchuk (dpinchuk@wisc.edu)

Background: Transition Path Theory

Consider a process governed by a stochastic differential equation (SDE) that evolves between some regions A and B. Below is a sample trajectory of the random process:



We look at a process governed by the SDE $dx = b(x)dt + \sqrt{\epsilon}dW$

The probability that at a given point x , the process will first visit B rather than A is given by the *committor* function q(x). The committor function q(x) satisfies the backward Kolmogorov equation and the following boundary conditions:

$$\begin{cases} b(x)\cdot\nabla q + \frac{\epsilon}{2}\Delta q = 0 \\ q|_{\partial A} = 0, \quad q|_{\partial B} = 1 \end{cases}$$

Neural Network approach 1: PINN

We show that physics informed neural networks (PINNs) [Raissi 2019] can be used to approximate the committor. Prior to our research, this method had not been applied to the committor problem. We model the committor for 2D (x,y) systems as follows:

$$q(x,y) \approx \underset{NN(x,y)}{\arg\min} \frac{1}{M} \sum_{(x,y)\in\Omega\setminus(A\cup B)} ||\mathcal{L}NN(x,y)||^2 + \frac{\alpha}{100}\sum_{(x,y)\in\partial A} ||NN(x,y)||^2 + \frac{\alpha}{100}\sum_{(x,y)\in\partial B} ||NN(x,y)-1||^2$$

M denotes the number of training points in the region outside of A and B. We also take 100 points uniformly sampled from each of the boundaries of A and B. With this model for the committor, we approximate the boundary conditions by including them in the above loss function. We also show the success of a hybrid model for the committor, a variation of the PINN method. The argument below is inspired by [Li,Lin, Ren 2019]:

$$q(x,y) \approx \underset{q_t(x,y)}{\arg\min} \frac{1}{M}\sum_{1}^{M} ||\mathcal{L}q_t(x,y)||^2$$

where

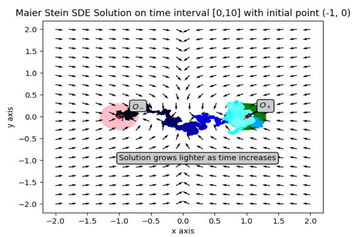
$$q_t(x,y) \approx (1-\chi_A(x,y))[(1-\chi_B(x,y))NN(x,y) + \chi_B(x,y)]$$

Sampling: We choose the training set to be a quasi-uniform sampling from the space where trajectories likely visit. Since trajectories spend more time in A and B, we use the enhanced sampling technique "metadynamics" to encourage trajectories to sample regions of lower density.

Advantages and Disadvantages:

1. Suitable for general SDEs including non-gradient SDEs
2. Suitable for complex regions A and B (not true for hybrid approach)
3. Does not require knowledge of the system's invariant density
4. Does not require knowledge of the sampling distribution, which allows us to easily use any training set of adequate size
5. Disadvantages: see advantages of the Variational Form Minimization.

Case Study: Maier-Stein System



The Maier-Stein (MS) System is a 2D non-gradient SDE admitting 2 stable equilibria (as illustrated) surrounded by the regions A in pink and B in green.

$$d\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}x - x^3 - \beta x y^2 \\ -(1+x^2)y\end{bmatrix}dt + \sqrt{\epsilon}dW$$

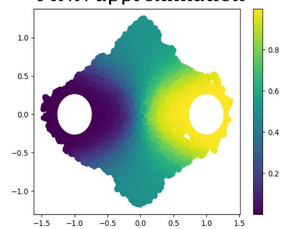
The sample trajectory shown begins at (-1,0) before travelling to region B.

We approximate q with a neural network of the form

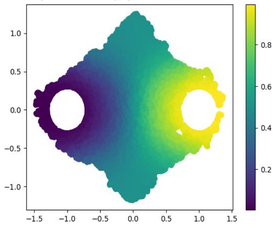
$$NN(x,y) = \underbrace{\delta}_{\text{sigmoid}}\left(\underbrace{v}_{1\times 20\text{ vector of weights}}\ \underbrace{\sigma}_{\tanh}\Big(\underbrace{W}_{20\times 2\text{ matrix of weights}}\begin{bmatrix}x\\y\end{bmatrix} + \underbrace{u}_{20\times 1\text{ bias vector}}\Big)\right)$$

Approximations:

PINN approximation



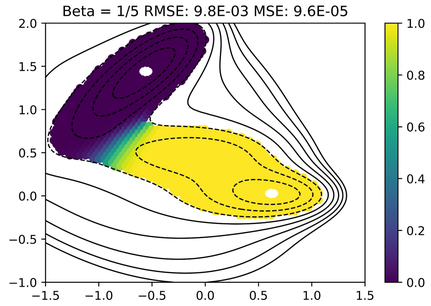
Hybrid Approximation



Case Study: Overdamped Langevin dynamics with the Mueller potential

The process X in overdamped Langevin dynamics evolves according to the following stochastic differential equation, where V is the Mueller potential, dw is Brownian motion and beta inverse is temperature: $dX_t = -\nabla V dt + \sqrt{2\beta^{-1}}dw$

Using the PINN framework with 1 hidden layer with 20 neurons, and a tanh activation function, the following estimation of the committor is achieved for constant beta = 0.2 on a uniform mesh training dataset of about 30,000 points within the -21 contour of the Mueller potential:



Similar results have been achieved for beta = 0.1 and smaller training datasets.

Ongoing work:

- Explore the effects of neural net architecture on error
- Apply methods of weighing training points to reduce error in the transition region

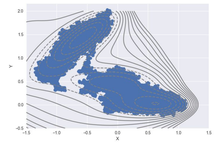
NN approach 2: Variational Form Minimization

Advantages and Disadvantages:

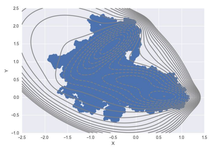
Though the PINN approach has proven to be successful, a drawback is that it can have slow run time in high dimensions. The loss function includes a Laplacian term, which is computationally the same as computing d^2 derivatives. This motivates an approach using the variational form of the PDE, which gives that $q(x) = \text{argmin}_f \int_{\Omega_{ab}} \|\nabla f(x)\|^2 e^{-\beta V(x)} dx$

Artificial Temperature Sampling:

As input for the neural network, we need a distribution of points that approximates the probability distribution of the system (a downside of this approach). To do this we run simulations of the process for many iterations. When the temperature is low, there will not be enough outside the metastable states to model effectively. To solve this problem, we can run the simulation for a temperature that is higher than our target temperature, then adjust the loss function of the neural net accordingly.



a) Simulation at Temperature 10



b) Simulation at Temperature 20

Direct Sampling:

When running simulations, there is a tendency for the process to get stuck in certain regions. To help minimize this effect, we introduce direct sampling. This approach starts with a quasi-uniform initial sample as was used in the PINN framework, then adds points around each initial point based upon the invariant probability there and the total number of points we want.

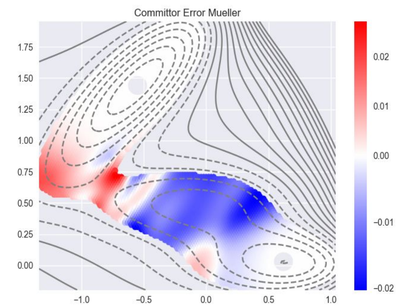
$$\mathbb{E}[\#\text{ of points added around } x_i] = \frac{P(x_i)}{\sum_{j=1}^{N_s} P(x_j)} * N_d$$

Method	Temp=10	Temp=20
Artificial Temp	0.024	0.025
Standard	0.56	0.048

c) RMSE using artificial temperature (without direct sampling)

Error Testing:

For low dimensional systems, there are techniques such as the finite element method which can be used to estimate the committor to a high degree of accuracy. We use this method to test the neural network.



Results: The best results were found using a neural network with 2 hidden layers and 20 neurons per layer. We achieved the following error plot compared to the FEM solution for the Mueller potential at temperature 10 after 1000 epochs of training. The root mean squared error was 0.007 and the maximum absolute error was 0.024