Experiments-1

May 12, 2021

0.1 Information Bottleneck

First, we recall the algorithm of Past-future information bottleneck.

For a given long trajectory $\{x_1, x_t, \ldots, x_T\}$, find a mapping r so that the $MI\mathcal{I}\left(r_t \| x_{t+\tau}\right)$ is (approximately) minimized.

By modeling r, f by neural networks, we can obtain the ideal bottleneck variable by solving

$$max_{r,f}\mathbb{E}_J[f(r_t, x_{t+\tau})] - \log \mathbb{E}_I[\exp f(r_t, x_{t+\tau})],$$

where $\mathbb{E}_J$ denotes the expectation over the joint distribution of $\left(r_t, x_{t+\tau}\right)$, and $\mathbb{E}_I$ denotes the expectation with r_t and $x_{t+\tau}$ being independently sampled.

0.1.1 Model 1

Suppose we have samples $(X_t)_{N\times 2}$, and the first dimension of X_t obey normal distribution, i.e. $X_t[:, 0] \sim \mathcal{N}(0, t)$. The second dimension of X_t equals the first dimension plus a random noise, i.e. $X_t[:, 1] = X_t[:, 0] + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$. On the other hand, by properties of Brownian Motion, $X_{t+\tau} - X_t \sim \mathcal{N}(0, \tau)$. We have $X_{t+\tau} = X_t + \mathcal{N}(0, \tau)$. Then we add a nonlinear transformation to the data, which is

$$\begin{aligned} X_t &= \exp(-\frac{\|w^T X_t + b\|^2}{2}) \\ X_{t+\tau} &= \exp(-\frac{\|w^T X_{t+\tau} + b\|^2}{2}) \end{aligned}$$

, where $w \sim U[-1, 1]$, and $b \sim U[0, 1]$.

As we know, the second dimension of X_t is a duplication of the first dimension plus some random noise. Therefore, the first dimension of X_t is sufficient enough to predict $X_{t+\tau}$. The predictive information bottleneck we trained, defined as $r_t = r(X_t)$ should be a function of the first dimension of X_t. To test this, we make the scatter plot of r_t versus $X_t[:, 0]$.

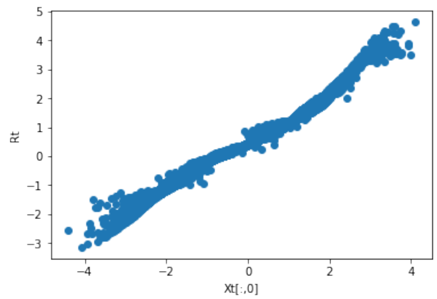

Figure 1: Pib test1:$r(X_t)$ vs $X_t[:,0]$

From the figure above, we can see $r(X_t)$ correspond to $X_t[:,0]$. In this way, we can infer that $r(X_t)$ is a low-dimensional sufficient statistics for predicting $X_{t+\tau}$.

0.1.2 Model2: Mueller Potential

We consider the noisy gradient flow induced by V in the form of the over-damped Langevin equation:

$$\dot{X}_t = -\nabla V(X_t) + \sqrt{2\beta^{-1}}\eta(t) \tag{1}$$

,where $\eta(t) \sim \mathcal{N}(0,1)$ is a random noise, and $V(X_t)$ is the rugged Mueller potential in two dimensions(2D).

$$V(x_1,x_2) = \sum_{i=1}^{4} D_i \exp[a_i(x_1-X_i)^2+b_i(x_1-X_i)(x_2-Y_i)+c_i(x_2-Y_i)^2]+\gamma sin(2k\pi x_1)sin(2k\pi x_2) \tag{2}$$

The parameters γ and k control the toughness of the energy landscape.

In this example, we generate the data at the artificial temperature $k_BT' = 10 or 20$ separately by solving the Langevin equation 1 using the Euler-Maruyama scheme with the time step $\Delta t = 10^{-5}$.

Other parameters are as follows: $\gamma = 9$, $k = 5$, $a_{i=1:4} = [-1,-1,-6.5,0.7]$, $b_{i=1:4} = [0,0,11,0.6]$, $D_{i=1:4} = [-200,-100,-170,15]$, $X_{i=1:4} = [1,0,-0.5,-1]$, $Y_{i=1:4} = [0,0.5,1.5,1]$, $\beta = \frac{1}{k_BT}$.

The scatter plots of the samples are as follows, where the color represents $V(X_t)$.

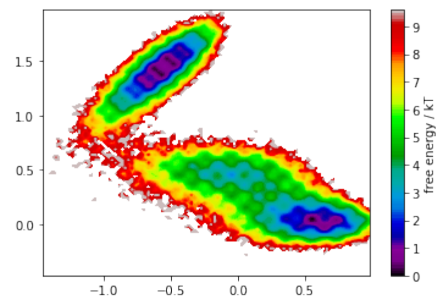

Figure 2: Mueller Potential: Sample of X_t, $k_B T = 10$

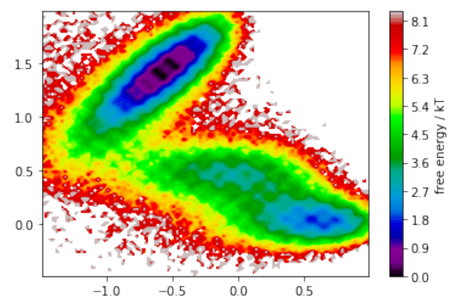

Figure 3: Mueller Potential: Sample of X_t, $k_B T = 20$

Then we get the results:

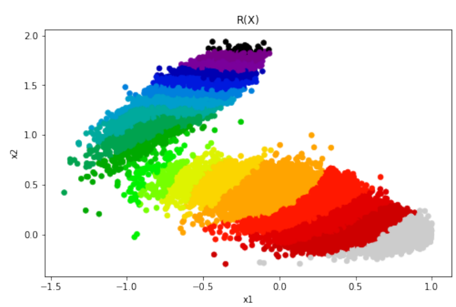

Figure 4: Pib Mueller Potential: Result of X_t, $k_B T = 10$

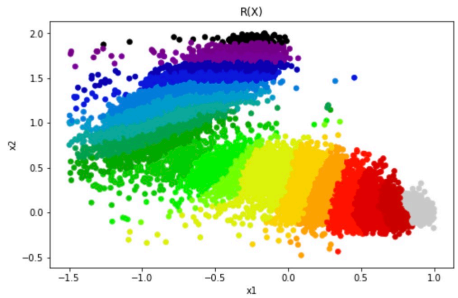

Figure 5: Pib Mueller Potential: Result of X_t, $k_B T = 20$

In Figure 4 and Figure 5, two dimensions of X_t are scattered, where the color represents $r(X_t)$.

We can see that $r(X_t)$ is perpendicular to $V(X_t)$, thus $r(X_t)$ keeps the information of predicting $X_{t+\tau}$ from X_t.

0.2 Observable Information Bottleneck

In this part, we introduce the results for Observable Information bottleneck.

First, we recall the algorithm.

In many practical applications, we are only interested in an (deterministic or random) observable y_t of the state instead of the whole state x_t.

For such cases, we propose a new bottleneck variable r_t as a solution to

$$\max_r \mathcal{L}(r) = \mathcal{I}\left(r_t || r_{t+\tau}, y_{t\tau}\right) - \mathcal{I}\left(x_t || r_{t+\tau}, y_{t+\tau}\right)$$

We can obtain the following equivalent formulation of OIB based on the Donsker-Varadhan representation:

$$\begin{array}{ll} \max_{r,f} \min_{f_x} & \mathbb{E}_J\left[f\left(r_t; r_{t+\tau}, y_{t+\tau}\right) - f_x\left(x_t; r_{t+\tau}, y_{t+\tau}\right)\right] \\ & -\log \mathbb{E}_I\left[\exp f\left(r_t; r_{t+\tau}, y_{t+\tau}\right)\right] \\ & +\log \mathbb{E}_I\left[\exp f_x\left(x_t; r_{t+\tau}, y_{t+\tau}\right)\right] \end{array},$$

where r, f, f_x are also three neural networks.

0.2.1 Model 1

We sample data $X_t \sim \mathcal{N}(0, t)$, and $X_{t+\tau} - X_t \sim \mathcal{N}(0, \tau)$. In addition, the deterministic observable Y_t of the state equals the first dimension of the state. i.e. $Y_t = X_t[:, 0]$, and $Y_{t+\tau} = X_{t+\tau}[:, 0]$. Therefore, the ideal information bottleneck variable $r_t = r(X_t)$ should highly correlated to the first dimension of X_t, or it should be a function of $X_t[:, 0]$.

In this part, we plot the $r(X_t)$ trained by observable information bottleneck algorithm and the first dimension of X_t to verify their correlation.

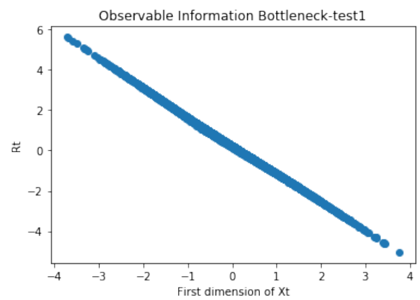

Figure 6: OPib test1:$r(X_t)$ vs $X_t[:, 0]$

From Figure 6, we can see that r_t is highly correlated to the first dimension of X_t, which proves our algorithm is efficient in this case.

0.2.2 Mueller Potential

In this example, we also generate the data at the artificial temperature $k_B T' = 10 or 20$ separately by solving the Langevin equation 1 using the Euler-Maruyama scheme with the time step $\Delta t = 10^{-5}$. The way of generating data is the same as that in Information Bottleneck. Apart from this, we add a dimension of X_t, which obey mixture gaussian distribution. To be more specific, third dimension of $X_t \sim 0.5 * \mathcal{N}(1,1) + 0.5 * \mathcal{N}(3,1)$.

We take the first second dimensions of X_t as the deterministic observable Y_t. i.e, $Y_t = X_t[:, 0, 1]$, and $Y_{t+\tau} = X_{t+\tau}[:, 0, 1]$. The goal of this algorithm is to train an information bottleneck variable $r(X_t)$ which contains sufficient information to predict $Y_{t+\tau}$.

Other parameters are as follows: $\gamma = 9$, $k = 5$, $a_{i=1:4} = [-1, -1, -6.5, 0.7]$, $b_{i=1:4} = [0, 0, 11, 0.6]$, $D_{i=1:4} = [-200, -100, -170, 15]$, $X_{i=1:4} = [1, 0, -0.5, -1]$, $Y_{i=1:4} = [0, 0.5, 1.5, 1]$, $\beta = \frac{1}{k_B T}$.

We scatter the samples of X_t, and use different colors to represent the first and second dimensions of r_t respectively. The results are as follows.

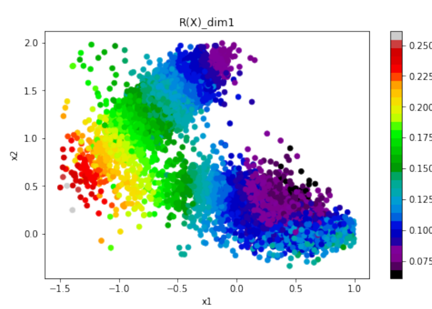

Figure 7: OPib Mueller Potential: Result of X_t, 1st dim of r_t, $k_B T = 20$

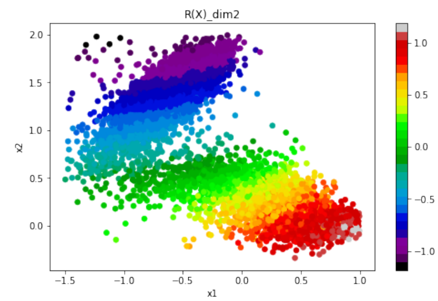

Figure 8: OPib Mueller Potential: Result of X_t, 2rd dim of r_t, $k_BT = 20$

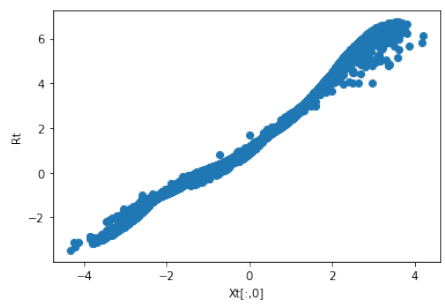

Figure 9: JS test1

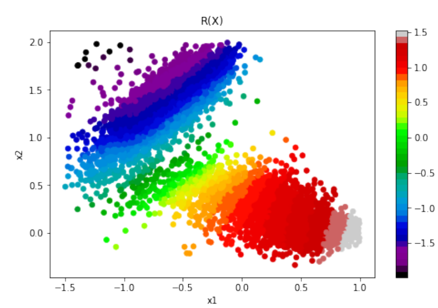

Figure 10: JS Mueller Potential

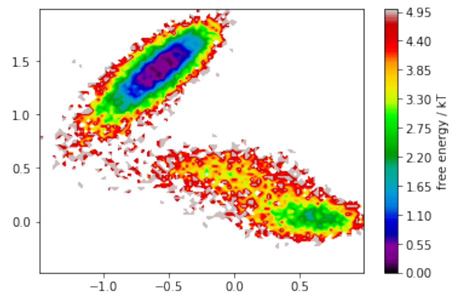

Figure 11: Approximate potential function-real

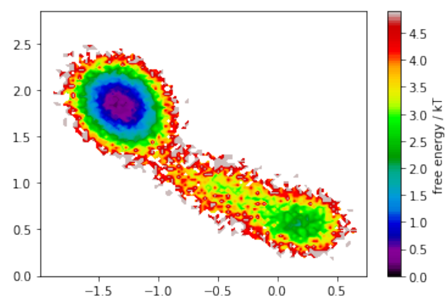

Figure 12: Approximate potential function-estimate