Active Importance Sampling for Variational Objectives Dominated by Rare Events: Consequences for Optimization and Generalization

MSML Conference 2021

Grant M. Rotskoff⋆, Andrew R. Mitchell⋆, Eric Vanden-Eijnden†

⋆ Stanford University, † Courant Insitute

July 29, 2021

Stanford University

rotskoff@stanford.edu

https://statmech.stanford.edu

- Characterizing transition paths in condensed matter physics, a rare event problem
- Two perspectives:
  1. Spectral methods (existence of a gap, metastability,…)
  2. Transition path theory / potential theory (Cf. Bovier, E., et al.)

A simple example

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}dW_t \tag{1}$$

ergodic w/r/t

$$\rho(x) = Z^{-1}e^{-\beta V(x)} \tag{2}$$

Define the *committor* function as the conditional probability $(X_t = x)$

$$q(x) = \mathbb{P}_x(t_B < t_A) \tag{3}$$

Backward Kolmogorov equation

Let L be the infinitesimal generator for the dynamics

$$0 = Lq = -\nabla V \cdot \nabla q + \Delta q \quad q(x) = 0, x \in A \quad q(x) = 1, x \in B \tag{4}$$

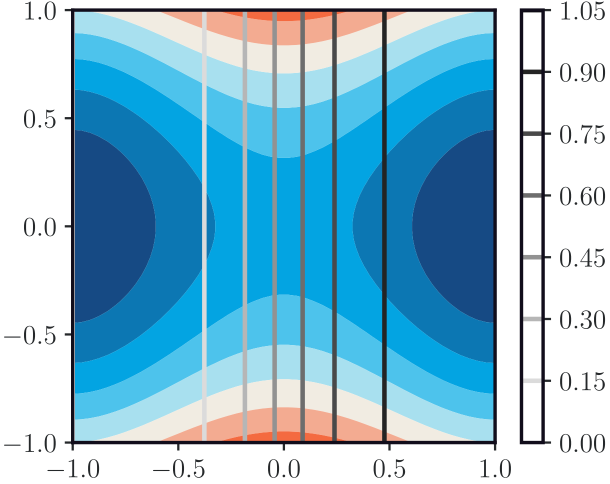
In 1D, we can solve directly for q,

$$q(x) = \frac{\int_a^x e^{\beta V(x)} dx}{\int_a^b e^{\beta V(x)} dx} \tag{5}$$

Check by inspection:

$$\partial_x V \partial_x q = \frac{\partial_x V e^{\beta V(x)} dx}{\int_a^b e^{\beta V(x)} dx} = \partial_x^2 q \tag{6}$$

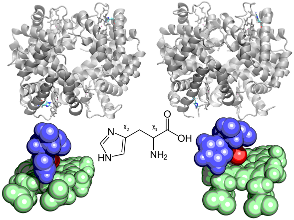
Explicit results for a simple double well



Insection of High-dim PDE & Rare Events

Committor PDE—a quintessentially high dimensional problem

- Canonical example: transitions between two conformations of a biological molecule



1. transition time is *long* relative to simulation time
2. importance sampling is key, but *not tractable* in high-dimensional systems
3. focus on low-lying eigenvalues not always right perspective

Main Contributions

- Under very general assumptions, we show that importance sampling asymptotically improves the generalization error.
- We describe an algorithm for *active* importance sampling that enables variance reduction for the estimator of the loss function, even in high-dimensional settings.
- We demonstrate numerically that this algorithm performs well both on low and high-dimensional examples and that, even when the total amount of data is fixed, optimizing the variational objective fails when importance sampling is not used

Variational principle:

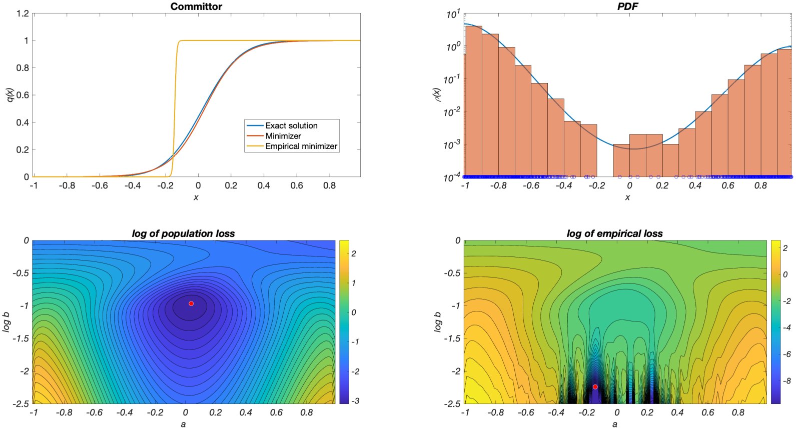
$$\inf_q C[q] \quad \text{subj. to} \quad q(A) = 0 \quad q(B) = 1 \tag{7}$$

with

$$C[q] \equiv \int_{\mathbb{R}^d} |\nabla q(\boldsymbol{x})|^2 e^{-\beta V(\boldsymbol{x})} d\boldsymbol{x} \tag{8}$$

Represent q with a neural network, estimate (8), optimize. Several other approaches based on this idea: Khoo et al, Li et al.

Necessity of collecting rare data

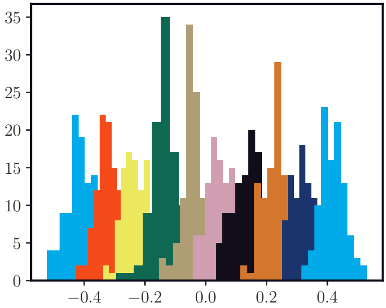


$I(q) = \int_{x_1}^{x_2} |q'(x)|^2 e^{-\beta V(x)} dx$, with $V(x) = (1 - x^2)^2 + x/10$, $\beta = 8$, and x_1, x_2 at the minima of $V(x)$, two parameters a and b for sigmoid

Reweighting the biased sampling

$$C[q] \approx \frac{1}{L}\sum_{l=1}^{L} |\nabla q|^2 e^{-\beta V} e^{-\beta G(u_l)} d\boldsymbol{x} \tag{11}$$

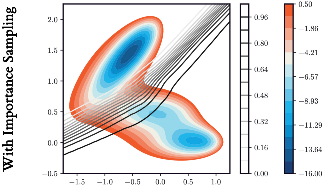
- Sample with overlap to compute the reweighting factor $G(u_l)$
- Standard stuff (use your favorite importance sampling method)

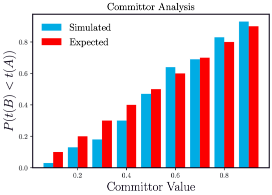


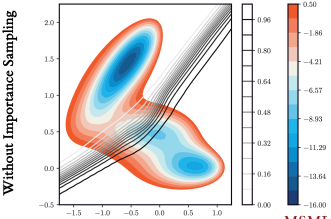
$$\frac{1}{M}\sum_{m=1}^{M} \frac{\exp\left(-\beta V(\boldsymbol{x}_{m,l+1}) + \frac{k}{2}(q(\boldsymbol{x}_{m,l+1}) - u_{l+1})^2\right)}{\exp\left(-\beta V(\boldsymbol{x}_{m,l}) + \frac{k}{2}(q(\boldsymbol{x}_{m,l}) - u_l)^2\right)} \tag{12}$$

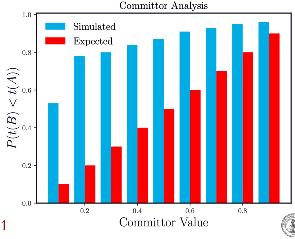
Müller-Brown Potential

$$V_{\mathrm{MB}}(\boldsymbol{x}) = \sum_{i=1}^{4} A_i \exp\left((\boldsymbol{x}-\mu_i)^T \Sigma_i^{-1} (\boldsymbol{x}-\mu_i)\right) \tag{13}$$





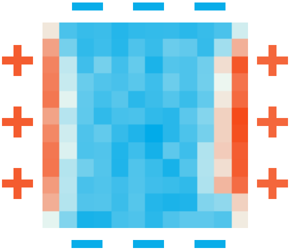




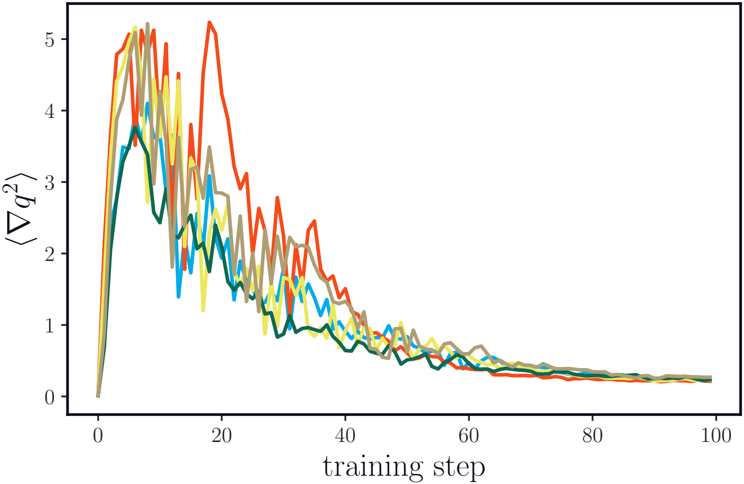
Ising-like model. Fix Dirichlet boundary conditions create metastability.

$$E[\rho] = \int \frac{D}{2}|\nabla\rho(z)|^2 + \frac{1}{4}(1 - \rho(z)^2)^2 dz \tag{14}$$

$$\partial_t \rho(z) = D\Delta\rho(z) + \rho(z) - \rho(z)^3 \tag{15}$$



Evolution of the loss (10 realizations of the experiment)



Optimization and identification of the transition states

Sampled transition paths ($q = -0.5, \ldots, 0.5$)