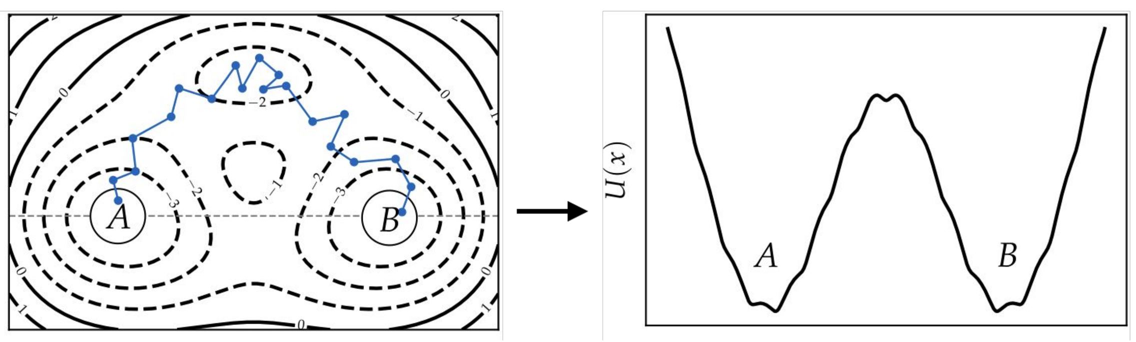
Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling

Yuanqi Du*, Michael Plainer*, Rob Brekelmans*, Chenru Duan, Frank Noé, Carla P. Gomes, Alan Aspuru-Guzik, Kirill Neklyudov

Overview

Goal

The goal is to efficiently sample possible transition paths connecting two low-energy states separated by a high-energy barrier.



Transition Path Sampling

We consider the transition dynamics in Brownian systems with Langevin dynamics.

Stochastic Differential Equation (SDE)

$$dx_t = b_t(x_t) \cdot dt + \Xi_t dW_t$$

Reference process (Langevin dynamics)

1st order

$$dx_t = -(\gamma M)^{-1} \nabla_x U(x_t) \cdot dt + (\gamma M)^{-\frac{1}{2}} \sqrt{2 k_B \mathcal{T}} \cdot dW_t$$

2nd order

$$d\bar{x}_t = \bar{v}_t \cdot dt$$
$$d\bar{v}_t = \Big(-M^{-1} \nabla_x U(\bar{x}_t) - \gamma \bar{v}_t \Big) \cdot dt + M^{-\frac{1}{2}} \sqrt{2 \gamma k_B \mathcal{T}} \cdot dW_t$$

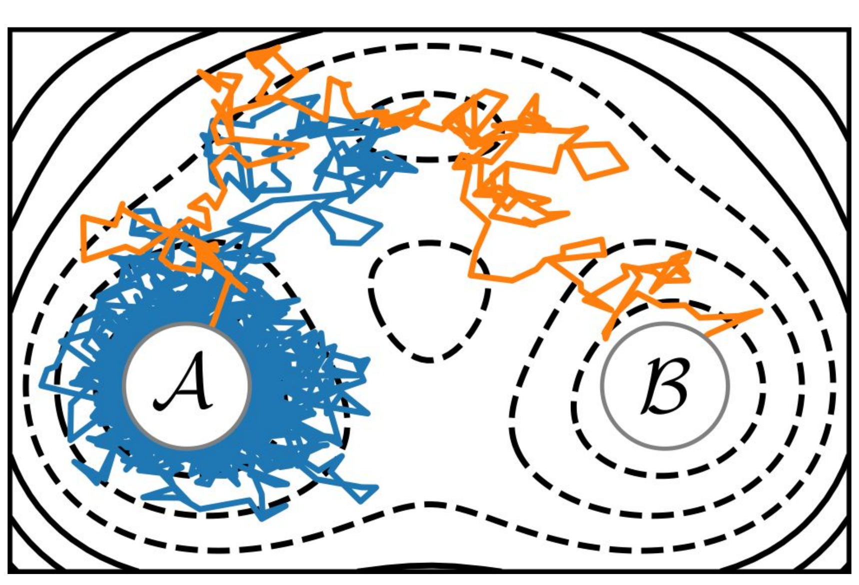
$$x_t = \begin{bmatrix} \bar{x}_t \\ \bar{v}_t \end{bmatrix}, \quad b_t(x_t) = \begin{bmatrix} \bar{v}_t \\ -M^{-1} \nabla_x U(\bar{x}_t) - \gamma \bar{v}_t \end{bmatrix},$$
$$G_t = \begin{bmatrix} 0 & 0 \\ 0 & M^{-\frac{1}{2}} \sqrt{2 \gamma k_B \mathcal{T}} \end{bmatrix}$$

Easy to sample by simulating SDE

$$\rho(x_t = x \mid x_0 = A)$$

Difficult to sample when conditioned on endpoint

$$\rho(x_t = x \mid x_0 = A, x_T \in \mathcal{B})$$



Transition Path Sampling with Doob's h-transform

Design choices

- **Every sample matters**: our method avoids wasteful forward simulation and importance sampling (simulation-free)
- **Optimization over sampling**: we optimize over an expressive variational family of approximations through neural networks
- **Problem-informed parameterization**: we enforce the exact boundary condition, thus reducing the search space

Doob's h-transform

Doob's h-transform allows us to condition the underlying SDE on the event that the state $\mathcal{B}$ is reached at $t = T$.

Consider the unconditioned reference process $\mathbb{P}^{\text{ref}}_{0:T}$:

$$dx_t = b_t(x_t) \cdot dt + \Xi_t dW_t \qquad x_0 \sim \rho_0$$

With Doob's h-transform, we can condition it on the endpoint by adding an additional drift term to get $\mathbb{P}^{*}_{0:T}$:

where $h(x,t) := \rho(x_T \in \mathcal{B} \mid x_t = x)$ and $G_t := \frac{1}{2} \Xi_t \Xi_t^T$

$$dx_{t|0,T} = \big(b_t\,(x_{t|0,T}) + 2 G_t \boxed{\nabla_x \log h\,(x_{t|0,T}, t)} \big) \cdot dt + \Xi_t dW_t$$

intractable!

Variational Formulation of Doob's h-transform

Theorem

The following Lagrangian action function has a unique solution which matches the Doob h-transform

$$\mathcal{S} = \min_{q,v} \int_0^T dt \int dx\; q_{t|0,T}(x) \left\langle v_{t|0,T}(x), G_t v_{t|0,T}(x) \right\rangle$$

such that

$$\frac{\partial q_{t|0,T}(x)}{\partial t} = -\left\langle \nabla_x, q_{t|0,T}(x) \left(b_t(x) + 2 G_t v_{t|0,T}(x) \right) \right\rangle + \sum_{ij} (G_t)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} q_{t|0,T}(x)$$

with $q_0(x) = \delta(x - A) \qquad q_T(x) = \delta(x - B)$

At optimality:

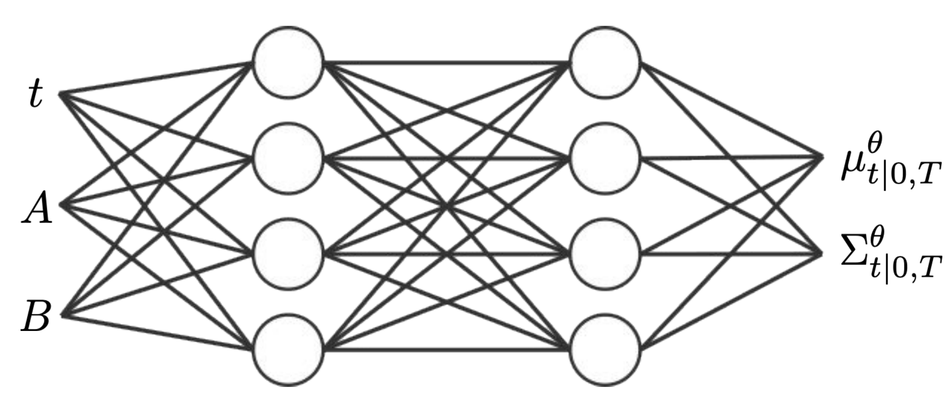
$$q^*_{t|0,T}(x) = \rho_{t|0,T}(x)$$
$$v^*_{t|0,T}(x) = \nabla_x \log h(x,t)$$

Computational Approach

Tractable Drift

We assume $q_{t|0,T}(x)$ to follow a Gaussian and train a neural network to predict the parameters.

$$q^\theta_{t|0,T} = \mathcal{N}(\mu^\theta_{t|0,T}, \Sigma^\theta_{t|0,T})$$



$$\theta^* = \operatorname*{argmin}_\theta \mathbb{E}_{x \sim q^\theta_{t|0,T}} \left[\mathcal{S}(x, \theta) \right]$$

$$\mu^{(\theta)}_{t|0,T} = (1-t)A + t\,B + t(1-t)\text{NNET}_\theta(t, A, B)_{[:D]}$$
$$\Sigma^{(\theta)}_{t|0,T} = t(1-t)\texttt{diag}\big(\text{NNET}_\theta(t, A, B)_{[D:]}\big) + \sigma^2_{\min}\mathbb{I}.$$

Optimization Want SDE whose evolution follows $q_{t|0,T}(x)$ and is of the form:

$$dx_{t|0,T} = \underbrace{\left(b_t(x_{t|0,T}) + 2 G_t\; v_{t|0,T}(x_{t|0,T}) \right)}_{u_{t|0,T}} \cdot dt + \Xi_t\; dW_t$$

where $v_{t|0,T}(x_{t|0,T}) \approx \nabla_x \log h(x_{t|0,t}, t)$

Assuming Gaussian $q_{t|0,T}(x)$, we can solve for corresponding

$$u^{(q)}_{t|0,T}(x) := \frac{\partial \mu_{t|0,T}}{\partial t} + \left[\frac{1}{2} \frac{\partial \Sigma_{t|0,T}}{\partial t} \Sigma^{-1}_{t|0,T} - G_t\, \Sigma^{-1}_{t|0,T} \right] (x - \mu_{t|0,T})$$

$$v^{(q)}_{t|0,T}(x) = \frac{1}{2} (G_t)^{-1} \left(u^{(q)}_{t|0,T}(x) - b_t(x) \right)$$

and substitute this vector field into the objective above

Algorithm 3 Variational Doob h-Transform (Single Sample with Gaussian Paths)

Input: Reference drift b_t, diffusion matrix $G_t = \Xi_t \Xi_t^T$, conditioning endpoints $x_0 = A, x_T = B$

while not converged **do**

Sample x_t for $t \sim \mathcal{U}(0,T)$ from current $q^{(\theta)}_{t|0,T}$ using

$$x_t = \mu^{(\theta)}_{t|0,T} + \Sigma^{(\theta)}_{t|0,T}\, \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, \mathbb{I}_D),$$
$$\mu^{(\theta)}_{t|0,T} = (1 - \frac{t}{T})A + \frac{t}{T}\,B + \frac{t}{T}(1 - \frac{t}{T})\text{NNET}_\theta(t, A, B)_{[:D]},$$
$$\Sigma^{(\theta)}_{t|0,T} = \frac{t}{T}(1 - \frac{t}{T})\texttt{diag}\big(\text{NNET}_\theta(t, A, B)_{[D:]}\big) + \sigma^2_{\min}\mathbb{I}.$$

Calculate $u_{t|0,T}(x_t)$ using the output of the neural network

$$u^{(q,\theta)}_{t|0,T}(x_t) = \frac{\partial \mu^{(\theta)}_{t|0,T}}{\partial t} + \left[\frac{1}{2} \frac{\partial \Sigma^{(\theta)}_{t|0,T}}{\partial t} (\Sigma^{(\theta)}_{t|0,T})^{-1} - G_t\, (\Sigma^{(\theta)}_{t|0,T})^{-1} \right] (x_t - \mu^{(\theta)}_{t|0,T})$$

Calculate $v_{t|0,T}(x_t)$ using $u^{(q,\theta)}_{t|0,T}(x_t)$ and the base drift $b_t(x_t)$

$$v^{(q,\theta)}_{t|0,T}(x_t) = \frac{1}{2} (G_t)^{-1} \left(u^{(q,\theta)}_{t|0,T}(x_t) - b_t(x_t) \right)$$

Calculate $\mathcal{S}$

$$\mathcal{S} = \langle v^{(q,\theta)}_{t|0,T}(x_t), G_t\, v^{(q,\theta)}_{t|0,T}(x_t) \rangle$$

- Note: $x_t \sim q^{(\theta)}_{t|0,T}$ and thus need to backpropagate through $b_t(x_t)$ using reparameterization trick

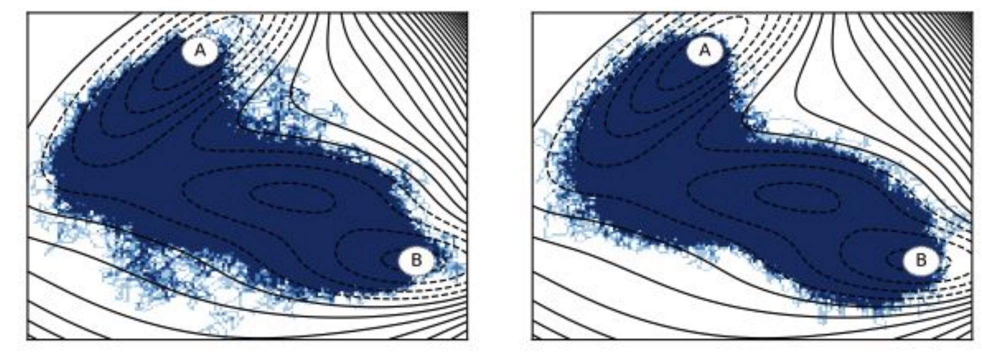
Update $\theta \leftarrow \text{optimizer}(\theta, \nabla_\theta \mathcal{S})$

end while

return

2D Toy Experiments

Müller-Brown Potential



(a) MCMC (b) Ours

Method	# Evaluations (↓)	Max Energy (↓)	MinMax Energy (↓)
MCMC (variable length)	3.53M	-13.77 ± 16.43	-40.75
MCMC	1.03B	-17.80 ± 14.77	-40.21
Ours	1.28M	-14.81 ± 13.73	-40.56

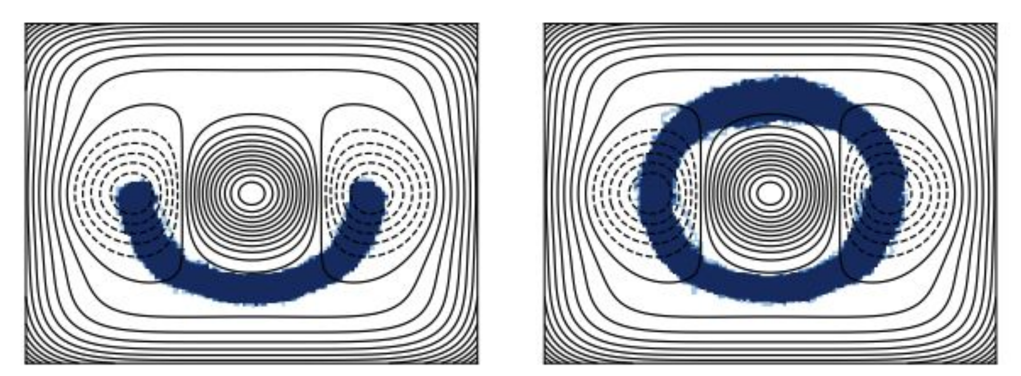
Mixture of Gaussian Parameterization

We introduce a mixture of Gaussians to capture more complex distributions and more paths.

$$q_{t|0,T}(x) = \sum_k w^k q^k_{t|0,T}(x)$$

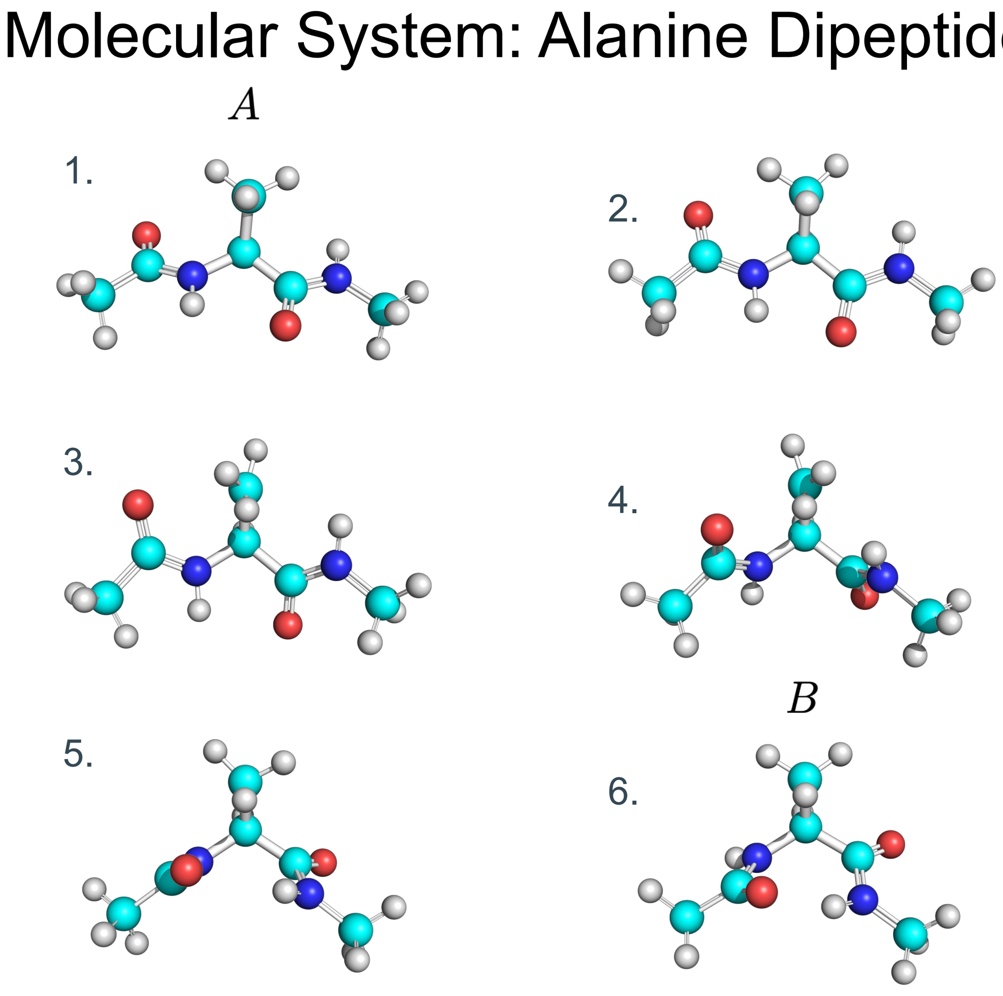
$$u^{(q)}_{t|0,T}(x) = \sum_{k=1}^{K} \frac{w^k q^k_{t|0,T}(x)}{\sum_{j=1}^K w^j q^j_{t|0,T}(x)} u^{(q,k)}_{t|0,T}(x)$$

Dual-Channel Double-Well Potential



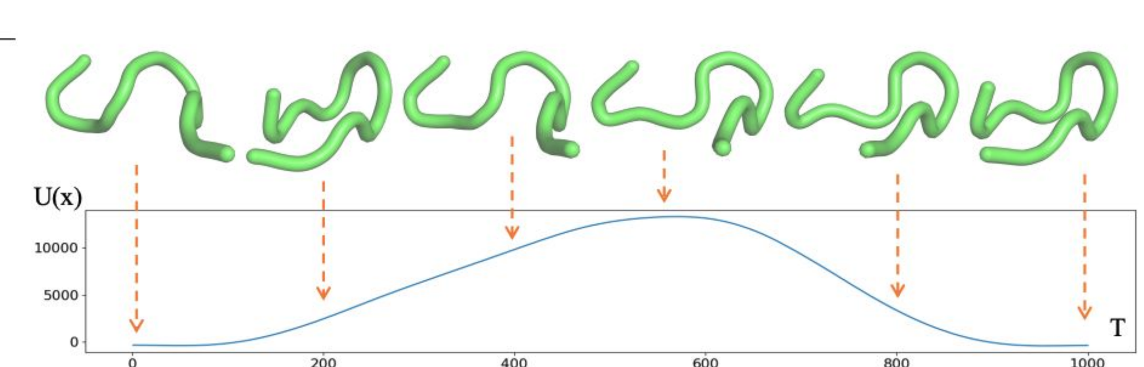
(a) Single Gaussian (b) Mixture of Gaussians

Molecular System: Alanine Dipeptide



Method	States	# Evaluations (↓)	Max Energy (↓)	MinMax Energy (↓)
MCMC (variable length)	CV	25.82M	1,212.81 ± 19,444.46	28.67
MCMC*	CV	1.29B	288.46 ± 128.31	60.52
MCMC (variable length)	relaxed	80.23M	269.16 ± 248.51	39.11
MCMC	relaxed	N/A	N/A	N/A
MCMC (variable length)	exact	N/A	N/A	N/A
MCMC	exact	N/A	N/A	N/A
Ours (Cartesian)	exact	38.40M	804.24 ± 0.20	803.62
Ours (Cartesian, 2 Mixtures)	exact	51.20M	828.77 ± 27.34	803.44
Ours (Internal)	exact	51.20M	352.20 ± 0.04	352.08
Ours (Internal, 2 Mixtures)	exact	51.20M	371.16 ± 82.88	239.66

Molecular System – Chignolin



OpenReview Code