Fantastic Path RND and Where to Find Them

Jiajun He, Francisco Vargas

"Density Ratio" and Radon-Nikodym Derivative

🧞‍♂️ Don't freak out about the name Radon-Nikodym Derivative
--- it's just the "*density ratio*"

👉 Very informally, let $\mathbf{P}$ and $\mathbf{Q}$ be two measures with density p and q, their density ratio is the **Radon-Nikodym Derivative (RND),** denoted as

$$\frac{p(x)}{q(x)} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(x)$$

👉 The density is essentially the RND w.r.t to Lebesgue measure

$$p(x) = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mu}(x), q(x) = \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mu}(x)$$

👉 RND is helpful for spaces without Lebesgue measure

Stochastic Differential Equations

⏭ Forward SDE

$$\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

⏮ Backward SDE

$$\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}$$

💡 Intuitive understanding by **Eular-Maruyama Discretisation:**

$$X_{n+1} - X_n = f(X_n, t_n)\Delta t + \sigma_n \sqrt{\Delta t}\epsilon$$
$$X_{n+1} - X_n = g(X_{n+1}, t_{n+1})\Delta t + \sigma_{n+1} \sqrt{\Delta t}\epsilon'$$

From Gaussian Density Ratio to Path RND

$$X_{n+1} - X_n = f(X_n, t_n)\Delta t + \sigma_n \sqrt{\Delta t}\epsilon$$

❓ for a **discretised** path sample $\{X_1, X_2, \ldots X_N\}$, what is its density?

✅ **Transition density:** $p(X_{n+1}|X_n) = N(X_{n+1}|X_n + f(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$

✅ **Full path density:** $p(X_1, X_2, \ldots X_N) = p(X_1) \prod p(X_{n+1}|X_n)$

From Gaussian Density Ratio to Path RND

Now take a closer look at

$$N(X_{n+1} | X_n + f(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$\log p = \frac{-(\sigma_n \sqrt{\Delta t}\epsilon)^2}{2\sigma_n^2 \Delta t} - \log \sigma_n - \boxed{\frac{1}{2}\log \Delta t} + C$$

💀 density diverge when $\Delta t \to 0$

From Gaussian Density Ratio to Path RND

But what if we have another SDE:

$$p_1 = N(X_{n+1}|X_n + f(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$p_2 = N(X_{n+1}|X_n + h(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$\log p_1 - \log p_2 = \frac{(2X_{n+1} - 2X_n - h\Delta t - f\Delta t)(h\Delta t - f\Delta t)}{2\sigma_n^2 \Delta t}$$

😎 density ratio did NOT diverge when $\Delta t \to 0$

From Gaussian Density Ratio to Path RND

For solution X to one SDE: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$,

we **cannot** define its density $p(X_0)\prod p(X_{t+\mathrm{d}t}|X_t)$

But with another SDE: $\mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$,

we **can** define density ratio **(Radon-Nikodym Derivative)** as a whole:

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)$$

Forward-forward RND and Girsanov

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\mathbf{Q}: \ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) \approx \frac{p(X_0)\prod N_1(X_{n+1}|X_n)}{q(X_0)\prod N_2(X_{n+1}|X_n)}$$

Forward-forward RND and Girsanov

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\mathbf{Q}:\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) = \underbrace{\frac{p(X_0)}{q(X_0)}}_{\text{Initial density ratio}} \exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t}_{\text{Forward Ito Integral } \int a_t(X_t)\cdot \mathrm{d}X_t = \lim \sum a_n(X_n)\cdot (X_{n+1} - X_n)} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

Forward-backward RND

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) \approx \frac{p_0(X_0)\prod N_1(X_{n+1}|X_n)}{q_1(X_1)\prod N_2(X_n|X_{n+1})}$$

Forward-backward RND

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0\,(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Forward-backward RND

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0\,(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \underbrace{\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)}_{\lim \frac{\prod N_1(X_{n+1}|X_n)}{\prod N_2(X_n|X_{n+1})}}$$

A Side Note on Stochastic Intergrals

Ito forward integral

$$\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$

Ito backward integral

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Stratonovich integral

$$\int a_t(X_t) \circ \mathrm{d}X_t = \lim \sum \frac{a_n(X_n) + a_{n+1}(X_{n+1})}{2} \cdot (X_{n+1} - X_n)$$

A Side Note on Stochastic Intergrals

Ito forward integral

$$\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$

Ito backward integral

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Conversion rule:

$$\int a_t(X_t) \cdot \mathrm{d}X_t - \int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = -\int \sigma_t^2 \nabla \cdot a_t \mathrm{d}t$$

Time-reversal and Nelson's relation

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim p_1$$

$$\overleftarrow{\mathbf{Q}} = \mathbf{P}, \mathrm{i.e.}, \frac{\overleftarrow{\mathrm{d}\mathbf{Q}}}{\mathrm{d}\mathbf{P}} = 1$$

iff

$$g(\cdot, t) = f(\cdot, t) - \sigma_t^2 \nabla \log p_t(\cdot)$$

Time-reversal and Nelson's relation

$$\mathbf{P}: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim p_1$$

$$\overleftarrow{\mathbf{Q}} = \mathbf{P}, \text{i.e.}, \frac{\overleftarrow{\mathrm{d}\mathbf{Q}}}{\mathrm{d}\mathbf{P}} = 1$$

iff

$$g(\cdot, t) = f(\cdot, t) - \sigma_t^2 \nabla \log p_t(\cdot)$$

e.g., 0 in VE process **score**

Where to Find Path RND?

Path RND is no different than other Density Ratios

👉 You can do **Importance Sampling**
👉 You can do **Variational Inference**

lead to a variety of methods!

Importance Sampling with Path RND

Importance Sampling:

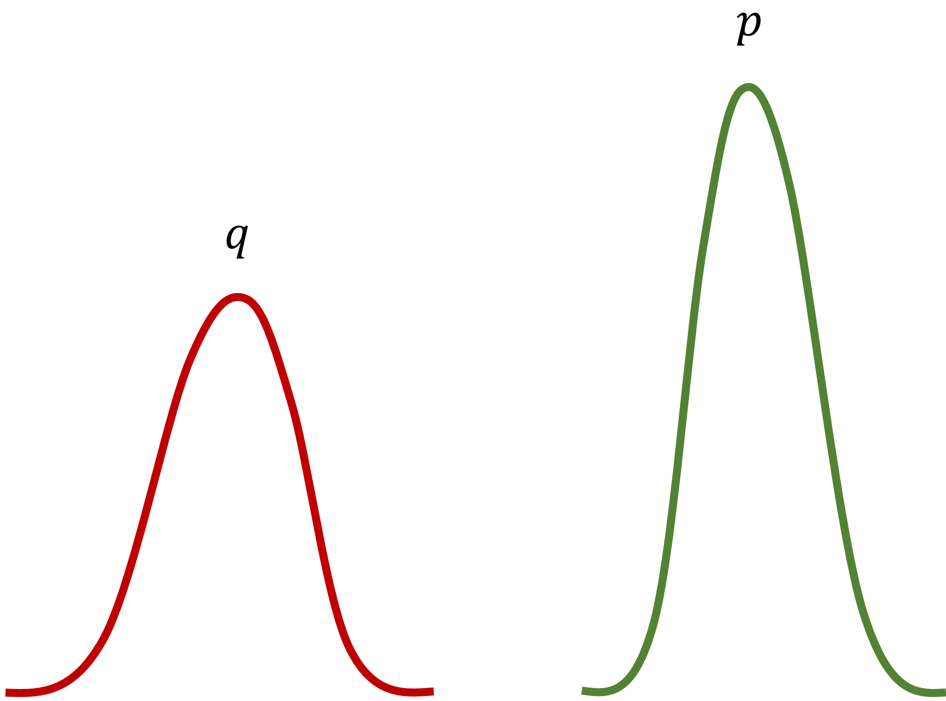
$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)}f(x)\right]$$

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

Importance Sampling with Path RND: AIS

Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q



Importance Sampling with Path RND: AIS

Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

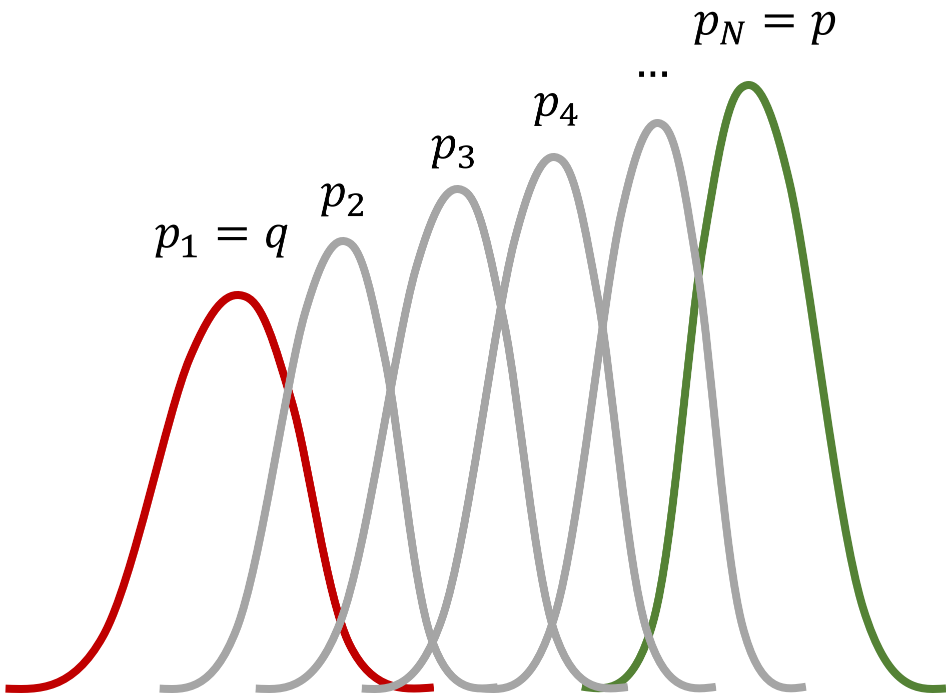
$X_1 \sim p_1 \qquad X_t \sim \mathrm{MCMC}_{p_t}(X_{t-1})$ **"Forward Process"**

$X_N \sim p_N \qquad X_{t-1} \sim \mathrm{MCMC}_{p_t}(X_t)$ **"Backward Process"**

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N)p(x_N)\prod p(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int \frac{f(x_N)p(x_N)\prod p(x_{n-1}|x_n)}{q(x_1)\prod p(x_n|x_{n-1})} q(x_1)\prod p(x_n|x_{n-1})\mathrm{d}x$$

Proposal



Importance Sampling with Path RND: AIS

Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

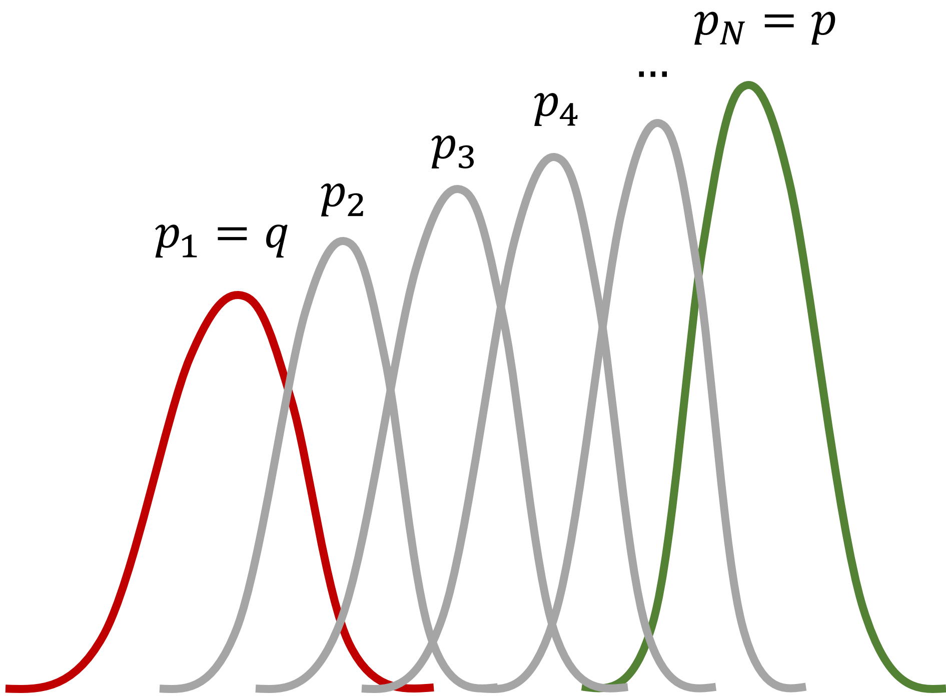
$X_1 \sim p_1$ $\quad$ $X_t \sim \mathrm{ULA}(X_{t-1})$ $\quad$ **"Forward Process"**

$X_N \sim p_N$ $\quad$ $X_{t-1} \sim \mathrm{ULA}(X_t)$ $\quad$ **"Backward Process"**

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N)p(x_N)\prod \boxed{N}(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int \frac{f(x_N)p(x_N)\prod \boxed{N}(x_{n-1}|x_n)}{q(x_1)\prod \boxed{N}(x_n|x_{n-1})} q(x_1)\prod \boxed{N}(x_n|x_{n-1})\mathrm{d}x$$

Proposal



Importance Sampling with Path RND: AIS

Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

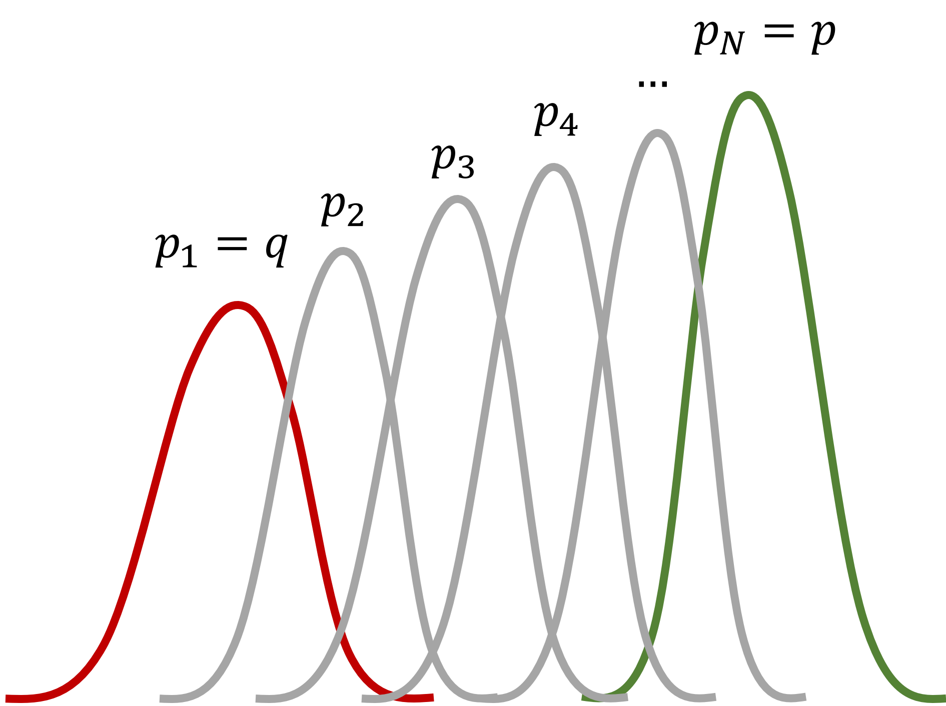
$X_1 \sim p_1 \qquad X_t \sim \mathrm{ULA}(X_{t-1})$ "Forward Process"

$X_N \sim p_N \qquad X_{t-1} \sim \mathrm{ULA}(X_t)$ "Backward Process"

Taking the limit… (∞ intermediate distributions)

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N)p(x_N)\prod N(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int f(x_N)\frac{p(x_N)\prod N(x_{n-1}|x_n)}{q(x_1)\prod N(x_n|x_{n-1})}q(x_1)\prod N(x_n|x_{n-1})\mathrm{d}x$$



Importance Sampling with Path RND: AIS

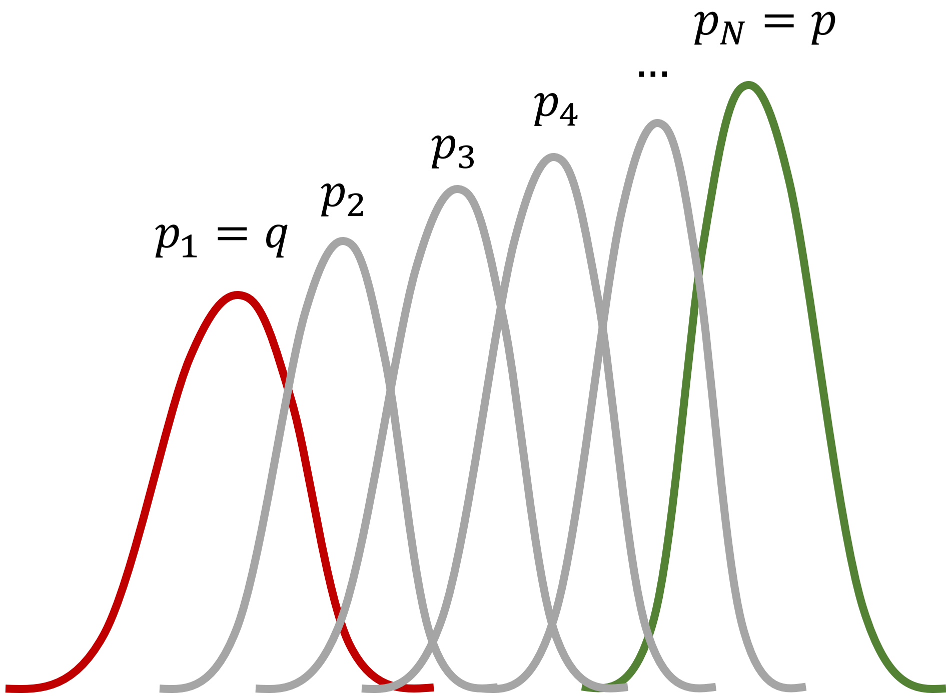
Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

$X_1 \sim q = p_1 \quad X_t \sim \mathrm{ULA}(X_{t-1})$ **"Forward Process"**

$X_N \sim p = p_N \quad X_{t-1} \sim \mathrm{ULA}(X_t)$ **"Backward Process"**

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N)p(x_N)\prod N(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int f(x_N)\frac{p(x_N)\prod N(x_{n-1}|x_n)}{q(x_1)\prod N(x_n|x_{n-1})}q(x_1)\prod N(x_n|x_{n-1})\mathrm{d}x$$



Importance Sampling with Path RND: AIS

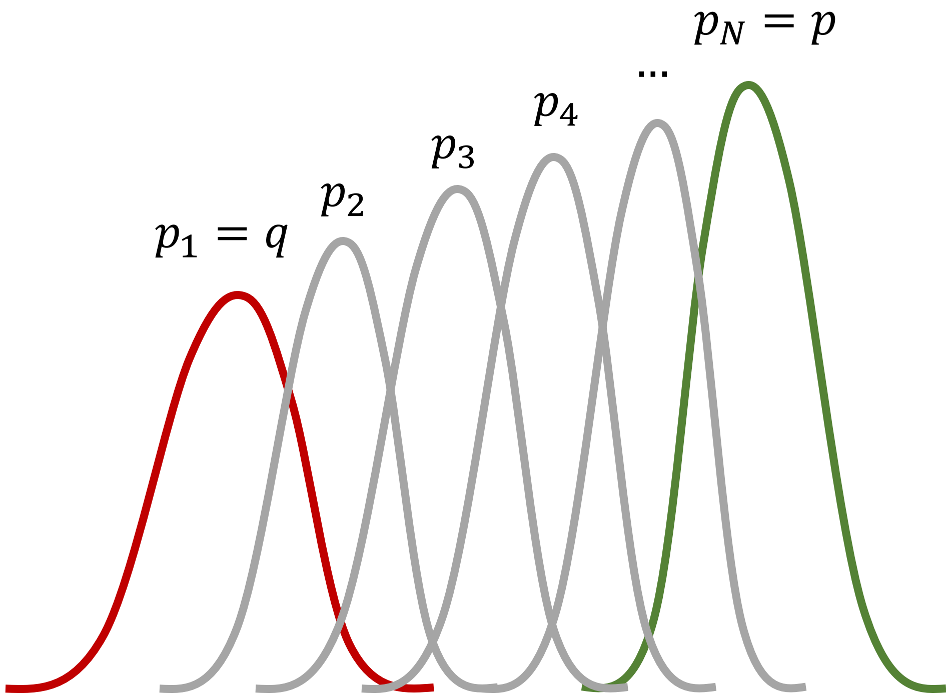
Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

$X_1 \sim q = p_1$ $\quad$ $\mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\mathrm{d}\overrightarrow{W_t},$

$X_N \sim p = p_N$ $\quad$ $\mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\mathrm{d}\overleftarrow{W_t},$

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N)p(x_N)\prod N(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int f(x_N)\frac{p(x_N)\prod N(x_{n-1}|x_n)}{q(x_1)\prod N(x_n|x_{n-1})} q(x_1)\prod N(x_n|x_{n-1})\mathrm{d}x$$



Importance Sampling with Path RND: AIS

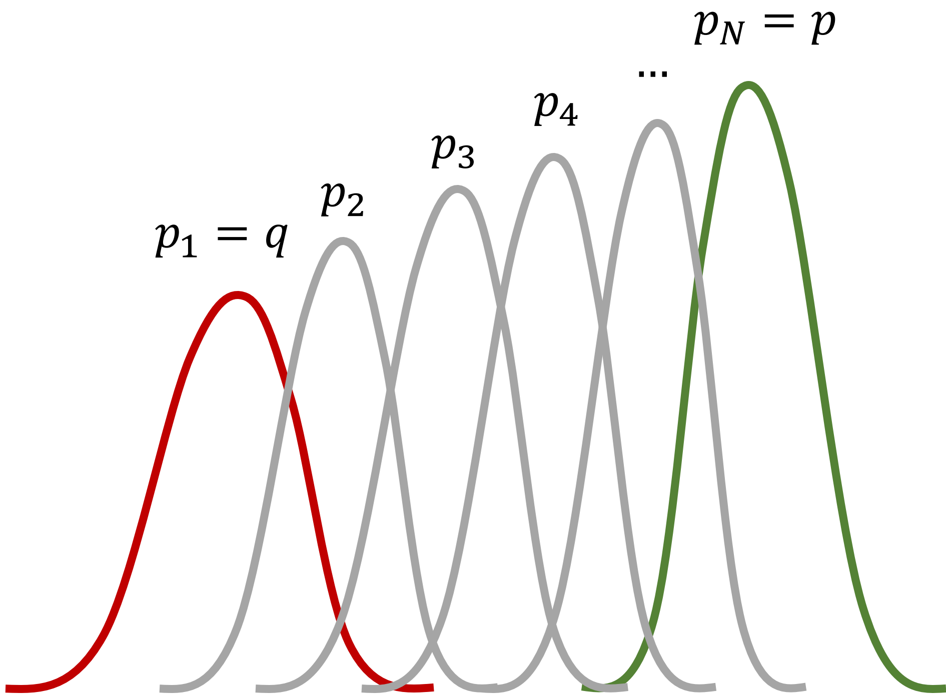
Estimate: $\mathbf{E}_{x\sim p}[f(x)]$ with proposal q

$X_1 \sim q = p_1$ $\quad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\mathrm{d}\overrightarrow{W_t},$

$X_N \sim p = p_N$ $\quad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\mathrm{d}\overleftarrow{W_t},$

$$\mathbf{E}_{x\sim p}[f(x)] = \int f(x_N) p(x_N) \prod N(x_{n-1}|x_n)\mathrm{d}x$$

$$= \int f(x_N) \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} \mathrm{d}\mathbf{Q}$$



$$\mathbf{E}_{X\sim\mathbf{P}}[f(X)] = \mathbf{E}_{X\sim\mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) f(X)\right]$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} =$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} \right)$$

👉 conversion rule

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t\right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t}\right)$$

👈 conversion rule

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t\right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{dt} \right)$$

👉 Ito’s lemma

$$\mathrm{d}f_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f)\mathrm{d}t + \nabla U_t \cdot \mathrm{d}X_t$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\begin{aligned}
\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} &= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t)\mathrm{d}t \right)
\end{aligned}$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right)$$

$$= \frac{Z_0 \exp(-U_1(X_1))}{Z_1 \exp(-U_0(X_0))} \exp\left(U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t)\mathrm{d}t \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t\right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t\right)$$

$$= \frac{Z_0 \cancel{\exp(-U_1(X_1))}}{Z_1 \cancel{\exp(-U_0(X_0))}} \exp\left(\cancel{U_1(X_1) - U_0(X_0)} + \int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right)$$

$$= \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t \right)$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q$ $\qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p$ $\qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

💡 Crooks Fluctuation Theorem

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

💡 Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}}\right] = \mathbf{E}_{\mathbf{Q}}\left[\frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = 1$$

From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

From AIS to Jarzynski Equation & Crooks Theorem

$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

💡 Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

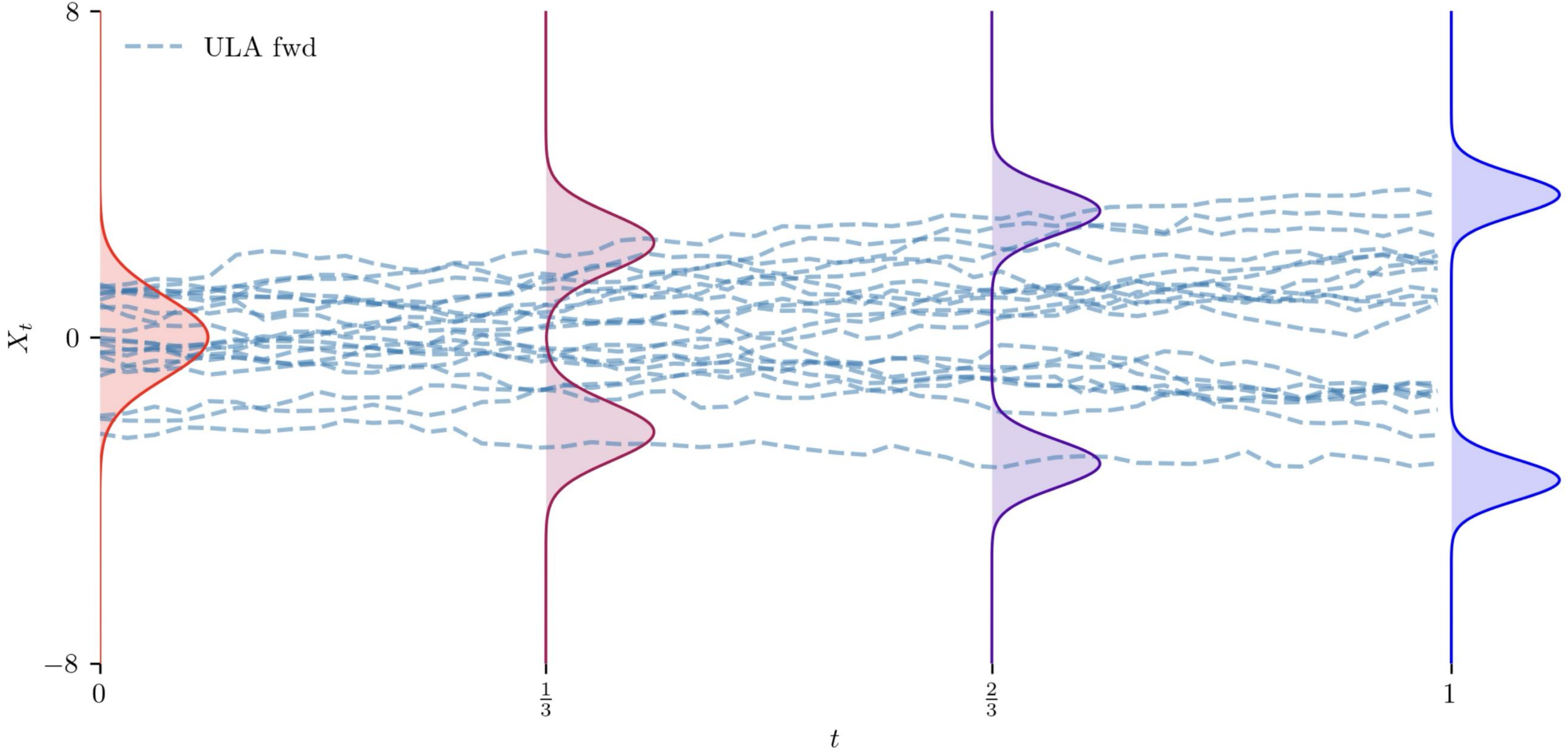
💡 Jarzynski Equation

From Jarzynski to Escorted Jarzynski

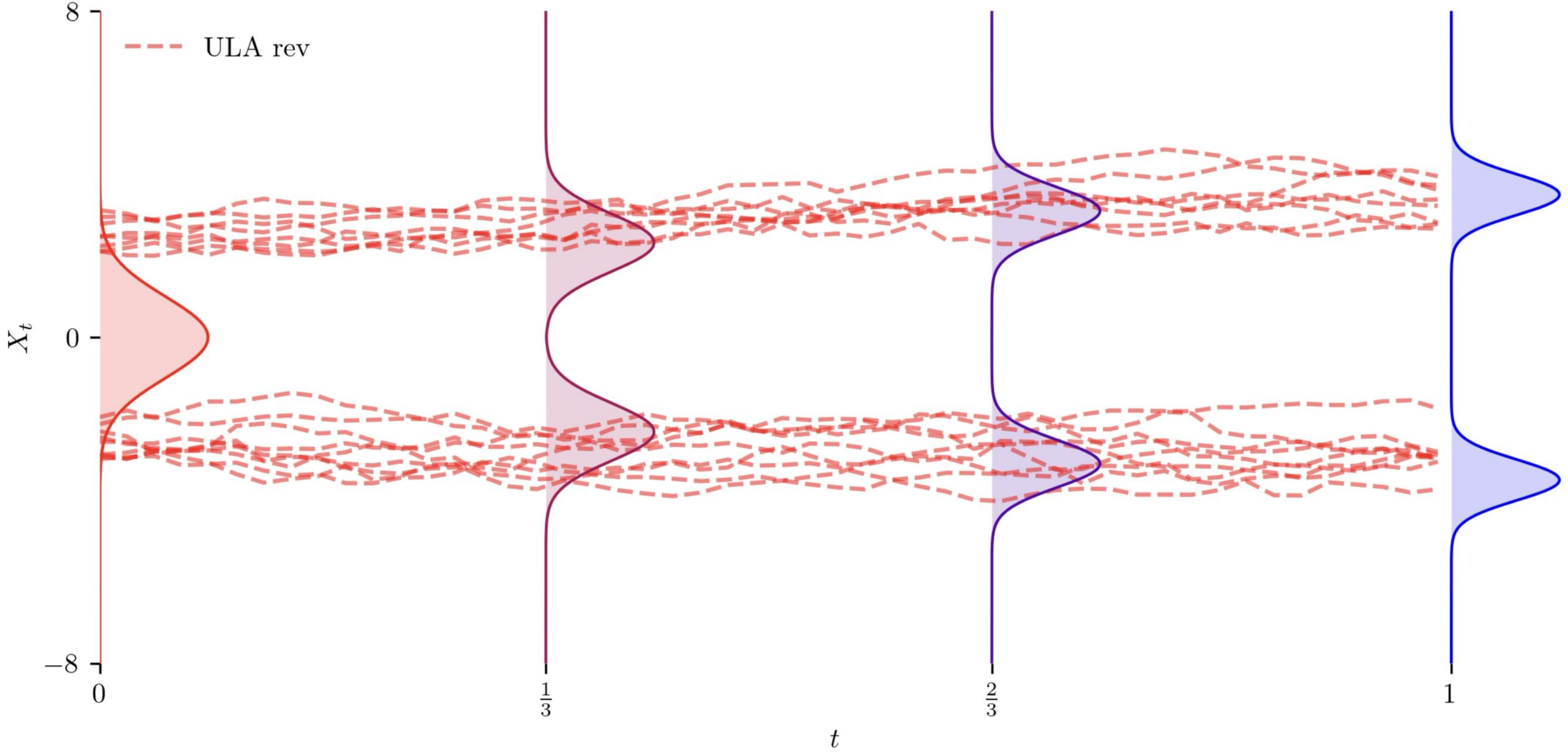
$X_0 \sim q$ $\qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$

$X_1 \sim p$ $\qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$

The evolving of samples is slower than that of energy



ULA fwd
8
0
−8
X_t
0
$\frac{1}{3}$
$\frac{2}{3}$
1
t

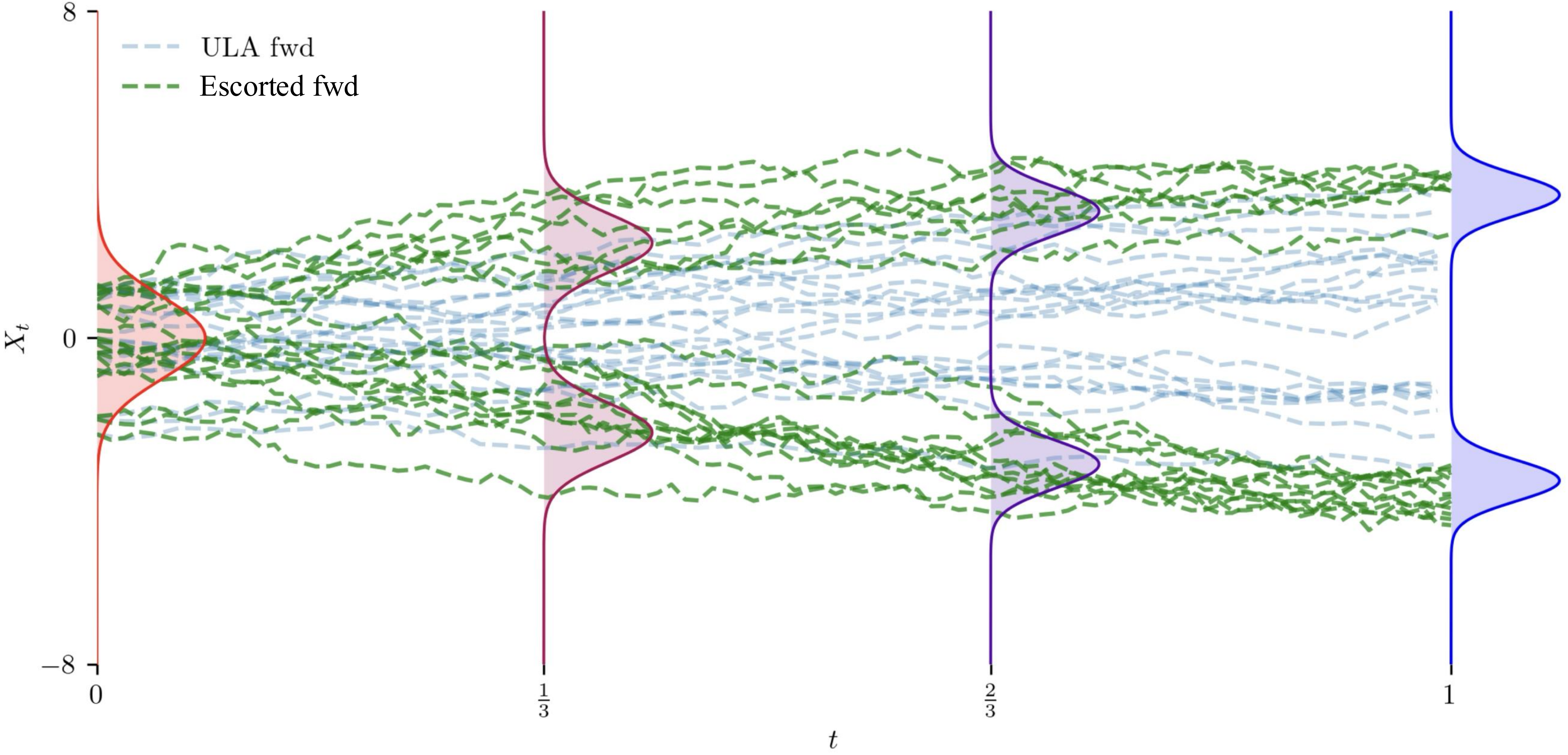

ULA rev
8
0
−8
X_t
0
$\frac{1}{3}$
$\frac{2}{3}$
1
t

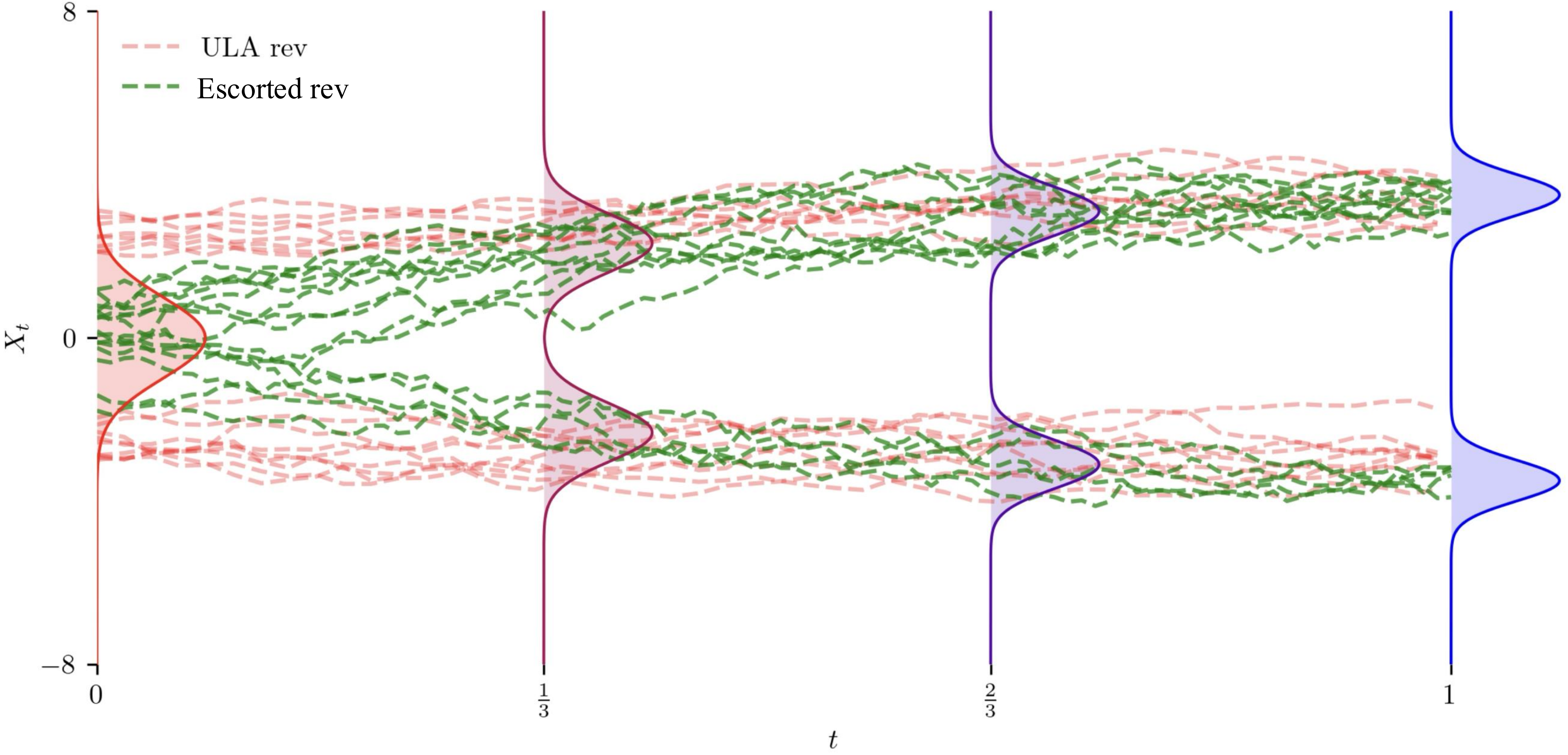
From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

The evolving of samples is closer to than that of energy


ULA fwd
Escorted fwd
8
0
−8
X_t
0
$\frac{1}{3}$
$\frac{2}{3}$
1
t



ULA rev
Escorted rev
8
0
−8
X_t
0
$\frac{1}{3}$
$\frac{2}{3}$
1
t

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} =$$

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot \mathrm{d}X_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 \,\mathrm{d}t + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 \,\mathrm{d}t \right)$$

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot \mathrm{d}X_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 \,\mathrm{d}t + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 \,\mathrm{d}t \right)$$

…conversion rule…

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot \mathrm{d}X_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 \,\mathrm{d}t + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 \,\mathrm{d}t \right)$$

…conversion rule…

…Ito’s Lemma…

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot \mathrm{d}X_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 \,\mathrm{d}t + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 \,\mathrm{d}t \right)$$

…conversion rule…

…Ito’s Lemma…

…cancel U_1 and U_0…

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot \mathrm{d}X_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 \,\mathrm{d}t + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 \,\mathrm{d}t \right)$$

…conversion rule…

…Ito’s Lemma…

…cancel U_1 and U_0…

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t \right)$$

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + \color{red}{u_t(X_t)}]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + \color{red}{u_t(X_t)}]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

💡 Controlled Crooks Fluctuation Theorem

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\, \overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\, \overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation

Can also be derived via PDEs [1] / Feynman-Kac formula [2]:

[1] Albergo, M. S., & Vanden-Eijnden, E (2025). NETS: A Non-equilibrium Transport Sampler. *ICML 2025.*

[2] Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., ... & Neklyudov, K. (2025). Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *ICML 2025*.

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

💡 match the forward & backward process

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

- 💡 match the forward & backward process
- 💡 match the marginal p_t of sampling process to U_t

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

- 💡 match the forward & backward process — 👉 CMCD [1]
- 💡 match the marginal p_t of sampling process to U_t — 👉 NETS [2]

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024.*

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025.*

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

- 💡 match the forward & backward process — 👉 CMCD [1]
- 💡 match the marginal p_t of sampling process to U_t — 👉 NETS [2]

👉 Setting 2: access sample and energy for q and p (e.g., aim to estimate Z_0/Z_1)

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024.*

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025.*

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find u_t?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

- 💡 match the forward & backward process — 👉 CMCD [1]
- 💡 match the marginal p_t of sampling process to U_t — 👉 NETS [2]

👉 Setting 2: access sample and energy for q and p (e.g., aim to estimate Z_0/Z_1)

- 💡 learn ∇U_t and u_t at the same time (e.g., by stochastic interpolant)

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024.*

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025.*

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Question: How to find $\boldsymbol{u_t}$?

👉 Setting 1: access sample for q + density of q + energy of p (neural samplers)

- 💡 match the forward & backward process 👉 CMCD [1]
- 💡 match the marginal p_t of sampling process to U_t 👉 NETS [2]

👉 Setting 2: access sample and energy for q and p (e.g., aim to estimate Z_0/Z_1)

- 💡 learn ∇U_t and u_t at the same time (e.g., by stochastic interpolant) 👉 FEAT [3]

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024.*
[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025.*
[3] He, J., Du, Y., Vargas, F., Wang, Y., Gomes, C. P., Hernández-Lobato, J. M., & Vanden-Eijnden, E. (2025). FEAT: Free energy Estimators with Adaptive Transport. *arXiv.*

Importance Sampling with Path RND

AIS $$\mathbf{E}_{X\sim\mathbf{P}}[f(X)] = \mathbf{E}_{X\sim\mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

Escorted Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

Importance Sampling with Path RND

AIS $$\mathbf{E}_{X\sim\mathbf{P}}[f(X)] = \mathbf{E}_{X\sim\mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

Escorted Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

$$Z = \int \exp(-U(X))\mathrm{d}X$$

Importance Sampling with Path RND

AIS $$\mathbf{E}_{X\sim\mathbf{P}}[f(X)] = \mathbf{E}_{X\sim\mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

Escorted Jarzynski $$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

$$Z = \int \exp(-U(X))\mathrm{d}X$$

$$p(X) = \int p(Z, X)\mathrm{d}Z$$

💡 $p(X)$ is the normalization factor for $p(Z|X) \propto p(Z, X)$

Importance Sampling with Path RND

$$\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N(0, v)$$

❓ What is $p_0(x)$ at $t = 0$?

Importance Sampling with Path RND

$$\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N(0, v)$$

❓ What is $p_0(x)$ at $t = 0$?

$$x \sim N(X|Z), p(Z) \sim N(0, v)$$

❓ What is $p(x)$?

Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$p(x) = \mathbf{E}_{z \sim p(z|x)}[1]p(x)$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$\begin{aligned} p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\ &= \mathbf{E}_{z\sim p(z|x)}[p(x)] \end{aligned}$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right]
\end{aligned}
$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right]
\end{aligned}
$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

Let's rewrite it in a path!

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right]
\end{aligned}
\qquad
\begin{aligned}
p_0(x) &= \\
&= \\
&= \\
&=
\end{aligned}
$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

Let's rewrite it in a path!

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right]
\end{aligned}
$$

$$
\begin{aligned}
p_0(x) &= \mathbf{E}_{X_{1:N}\sim p(X_{1:N}|X_0=x)}[1]p_0(x) \\
&= \\
&= \\
&=
\end{aligned}
$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

🤗 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\ &= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\ &= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\ &= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right] \end{aligned}$$

$$\begin{aligned} p_0(x) &= \mathbf{E}_{X_{1:N}\sim p(X_{1:N}|X_0=x)}[1]p_0(x) \\ &= \mathbf{E}_{X_{1:N}\sim p(X_{1:N}|X_0=x)}[p_0(x)] \\ &= \\ &= \end{aligned}$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

🤗 **Let’s rewrite it in a path!**

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z \sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z \sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z \sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\
&= \mathbf{E}_{z \sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right]
\end{aligned}
$$

$$
\begin{aligned}
p_0(x) &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)}[1]p_0(x) \\
&= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)}[p_0(x)] \\
&= \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)}\left[\frac{p_0(x)p(X_{1:N}|X_0=x)}{q(X_{1:N}|X_0=x)}\right] \\
&=
\end{aligned}
$$

Importance Sampling with Path RND

Marginal density relation with importance sampling:

🤗 **Let's rewrite it in a path!**

$$
\begin{aligned}
p(x) &= \mathbf{E}_{z\sim p(z|x)}[1]p(x) \\
&= \mathbf{E}_{z\sim p(z|x)}[p(x)] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(x)p(z|x)}{q(z|x)}\right] \\
&= \mathbf{E}_{z\sim q(z|x)}\left[\frac{p(z)p(x|z)}{q(z|x)}\right]
\end{aligned}
$$

$$
\begin{aligned}
p_0(x) &= \mathbf{E}_{X_{1:N}\sim p(X_{1:N}|X_0=x)}[1]p_0(x) \\
&= \mathbf{E}_{X_{1:N}\sim p(X_{1:N}|X_0=x)}[p_0(x)] \\
&= \mathbf{E}_{X_{1:N}\sim q(X_{1:N}|X_0=x)}\left[\frac{p_0(x)p(X_{1:N}|X_0=x)}{q(X_{1:N}|X_0=x)}\right] \\
&= \mathbf{E}_{X_{1:N}\sim q(X_{1:N}|X_0=x)}\left[\frac{p(X_{1:N},X_0=x)}{q(X_{1:N}|X_0=x)}\right]
\end{aligned}
$$

Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[\frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N} | X_0 = x)} \right]$$

Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[\frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N} | X_0 = x)} \right]$$

$$p\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_N \sim N$$

$$q\text{: } \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)}\left[\frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N}|X_0 = x)}\right]$$

$$p\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_N \sim N$$

$$q\colon\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)}\left[\frac{\prod N_p(X_n|X_{n+1})}{\prod N_q(X_{n+1}|X_n)} N(X_N)\right]$$

Importance Sampling with Path RND

$$p\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_N \sim N$$

$$q\colon\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0 = x)}\left[\frac{\prod N_p(X_n|X_{n+1})}{\prod N_q(X_{n+1}|X_n)} N(X_N)\right]$$

Importance Sampling with Path RND

$$p\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_N \sim N$$

$$q\colon\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[\frac{\prod N_p(X_n | X_{n+1})}{\prod N_q(X_{n+1} | X_n)} N(X_N) \right]$$

Forward-backward RND

$$\mathbf{P}\colon\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

RECALL: Forward-backward RND

$$\mathbf{P}: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0\,(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \underbrace{\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)}_{\lim \frac{\prod N_1(X_{n+1}|X_n)}{\prod N_2(X_n|X_{n+1})}}$$

Importance Sampling with Path RND

$$p\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_N \sim N$$

$$q\colon\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[\frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)} N(X_N) \right]$$

Forward-backward RND

$$\mathbf{P}\colon\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1)\exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)\right]$$

Forward-backward RND

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}}\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot (X_{n+1}-X_n)$$

Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1)\exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)\right]$$

Forward-backward RND

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}}\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot (X_{n+1} - X_n)$$

For simplicity, we hereafter call

$$R_f^g(X) = \exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1)\exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)\right]$$

Forward-backward RND

$$\mathbf{P}: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}}\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot (X_{n+1} - X_n)$$

For simplicity, we hereafter call

$$R_f^g(X) = \lim \frac{\prod N_g(X_n|X_{n+1})}{\prod N_h(X_{n+1}|X_n)}$$

Importance Sampling with Path RND

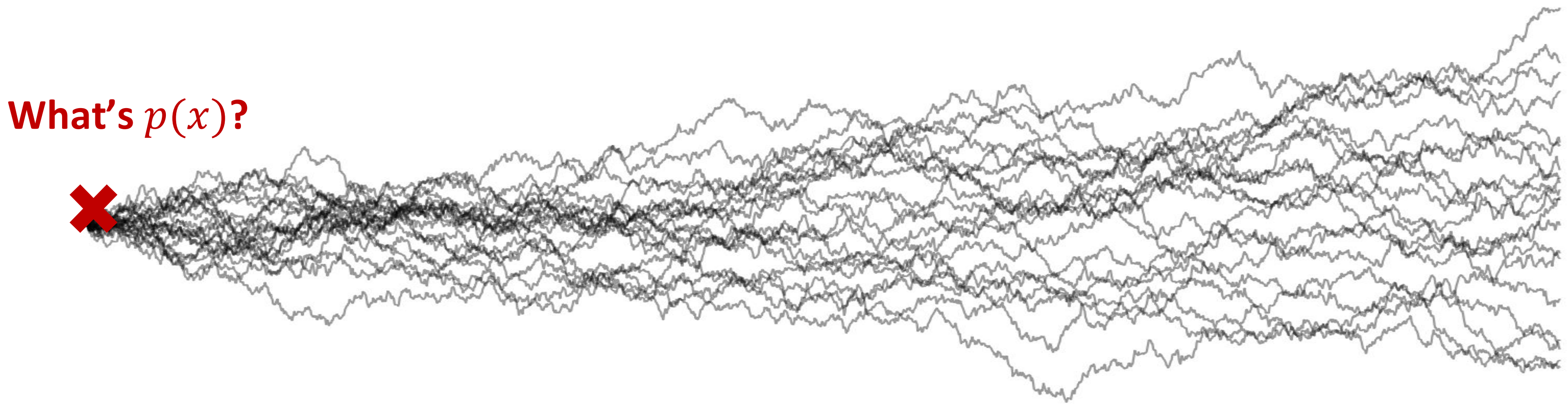
$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1) R_f^g(X_{[0,1]})\right]$$

Importance Sampling with Path RND

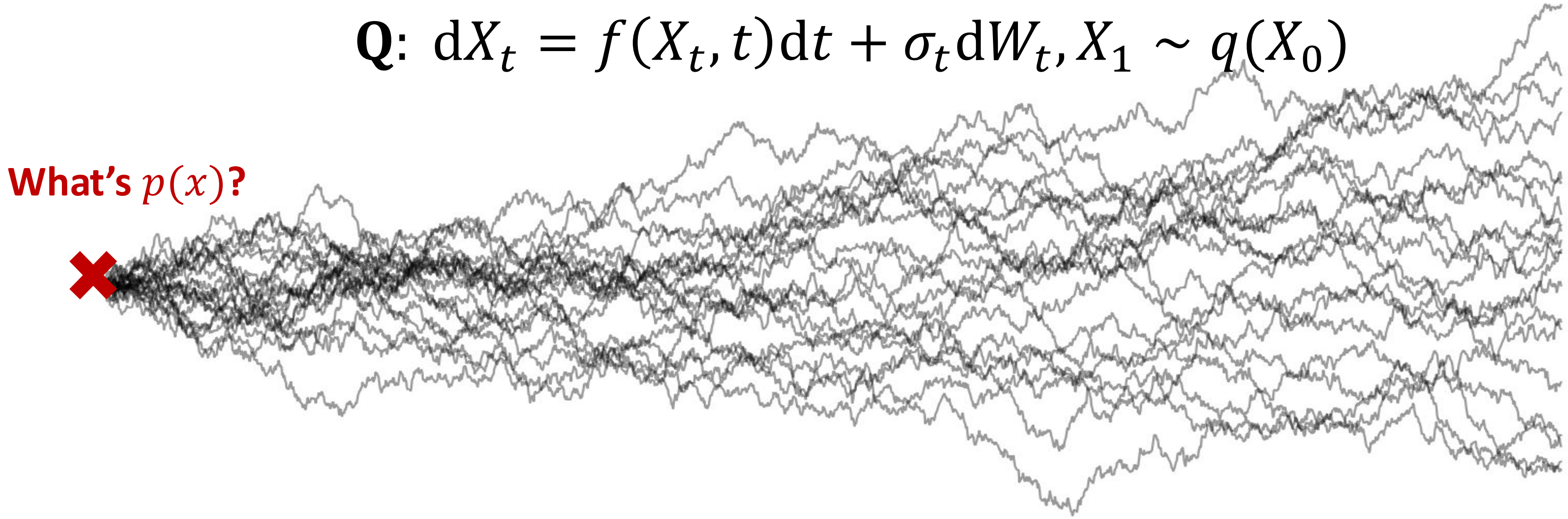
$$\overleftarrow{\mathbf{P}}\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$



Importance Sampling with Path RND

$\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$

$\mathbf{Q}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$

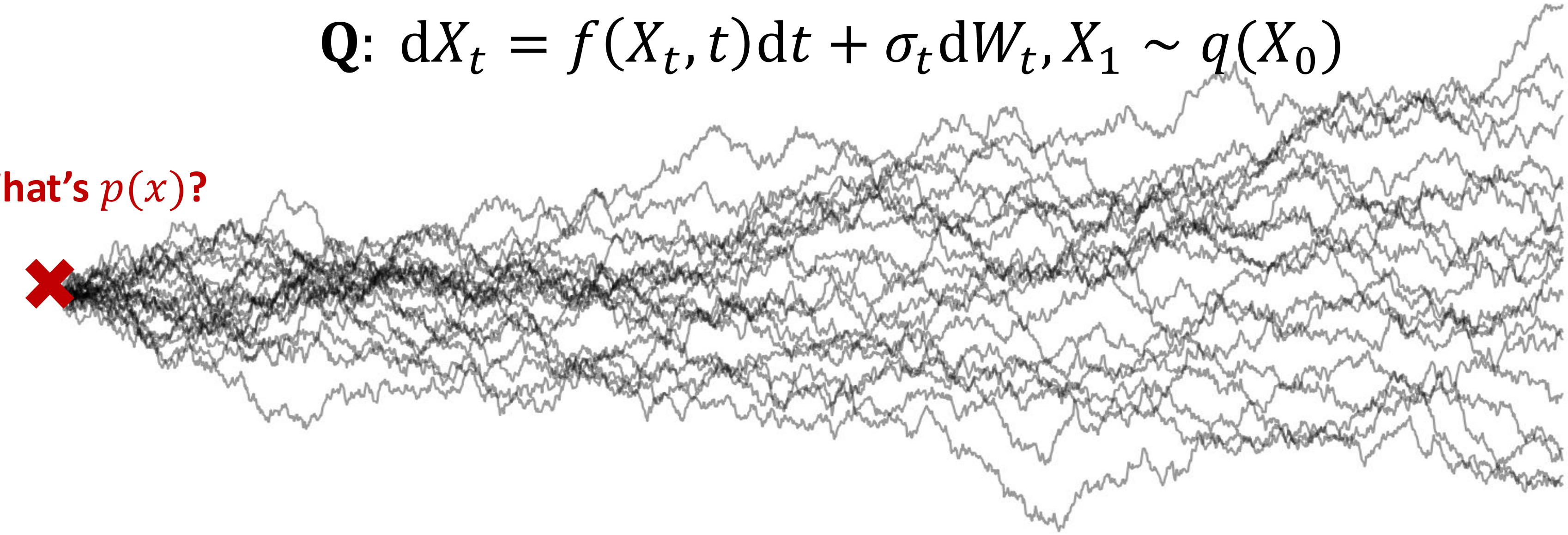


Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$

What's $p(x)$?

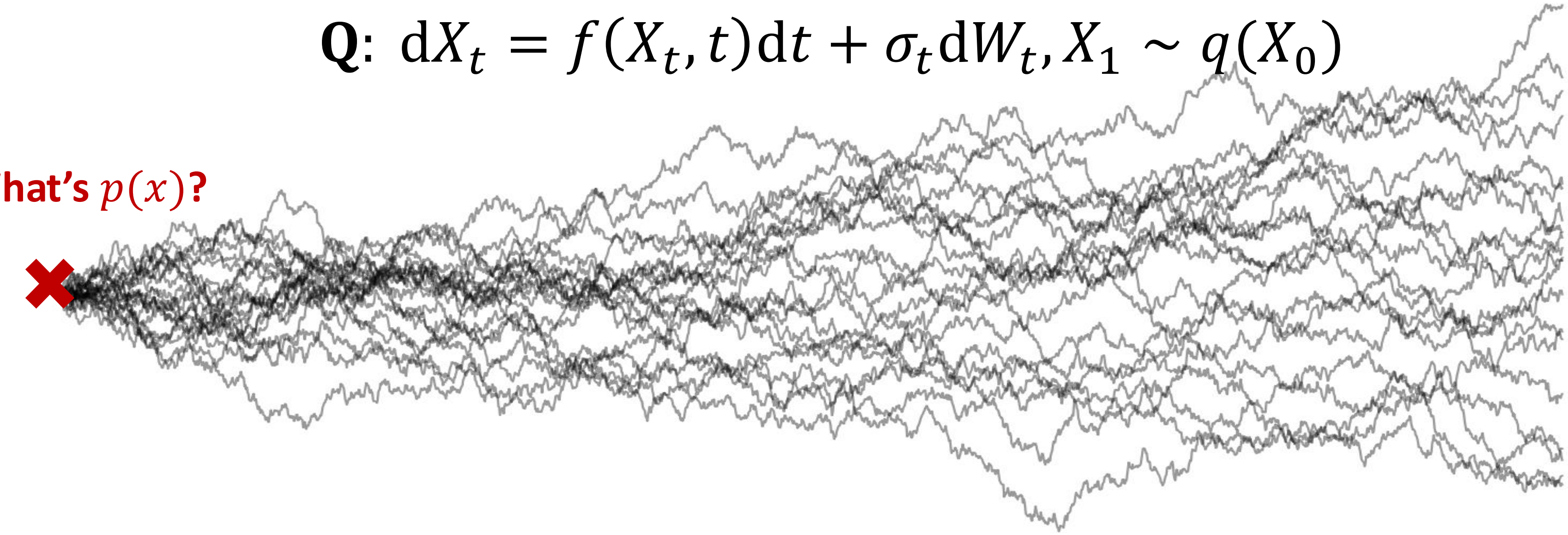


$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1) R_f^g(X_{[0,1]})\right]$$

Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{Q}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)$$



$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)}\left[N(X_1) R_f^g(X_{[0,1]})\right]$$

Can also be derived from Feynman-Kac formulation as shown in [1] and [2]:

[1] Huang, C. W., Lim, J. H., & Courville, A. C. (2021). A variational perspective on diffusion-based generative models and score matching. *NeurIPS 2021.*

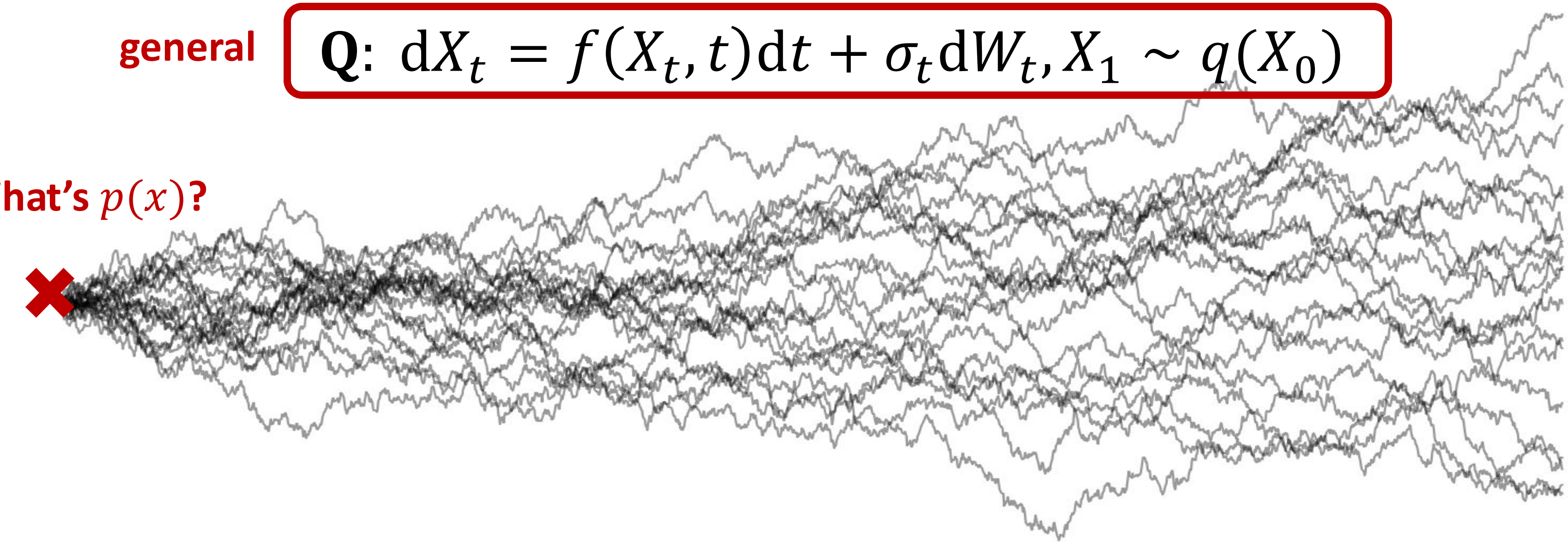
[2] Premkumar, A. (2024). Diffusion density estimators. *arXiv.*

Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

general $\boxed{\mathbf{Q}: \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_1 \sim q(X_0)}$

What's $p(x)$?



Importance Sampling with Path RND

Special Case: **time-reversal proposal**

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{P}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$
$$\mathbf{P}\colon\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_0 = x, X_{(0,1]}\right) = \mathbf{1}$$

Forward-backward RND

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot (X_{n+1} - X_n)$$

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$
$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_0 = x, X_{(0,1]}\right) = \mathbf{1}$$

$$\frac{N(X_1)}{p_0(X_0 = x)} R_f^g(X_{[0,1]}) = 1$$

Forward-backward RND

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot(X_{n+1} - X_n)$$

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$
$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_0 = x, X_{(0,1]}\right) = \mathbf{1}$$

$$N(X_1) R_f^g(X_{[0,1]}) = \color{red}{p_0(X_0 = x)}$$

Forward-backward RND

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{\mathbf{Q}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim q_1$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp\left(\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t}}_{\text{Backward Ito Integral}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

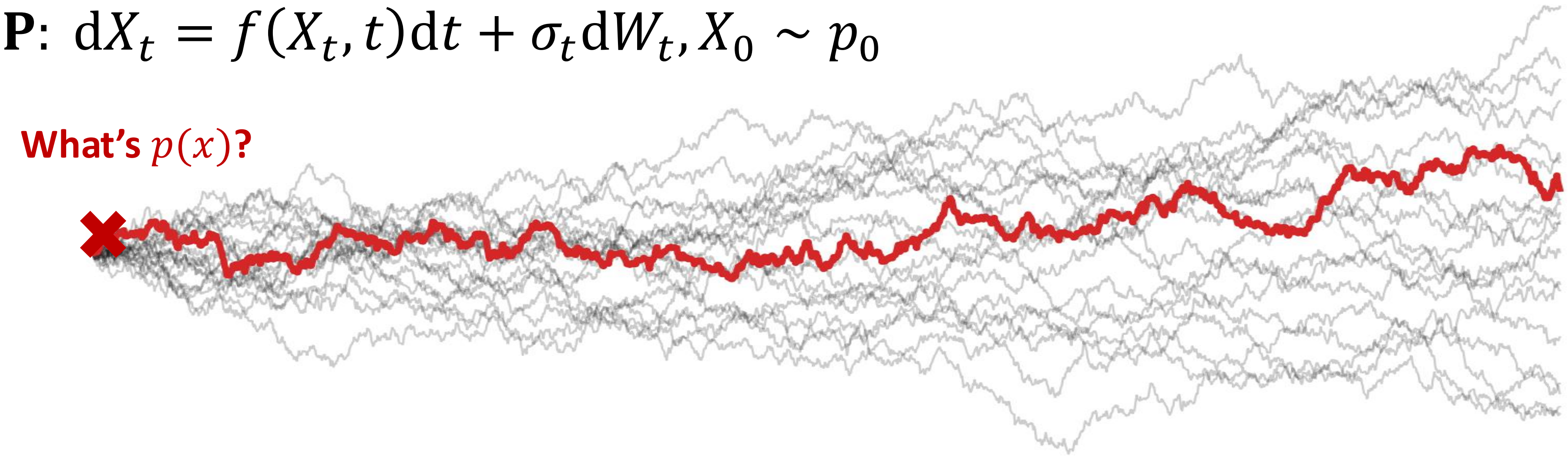
$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}\colon\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{P}\colon\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$

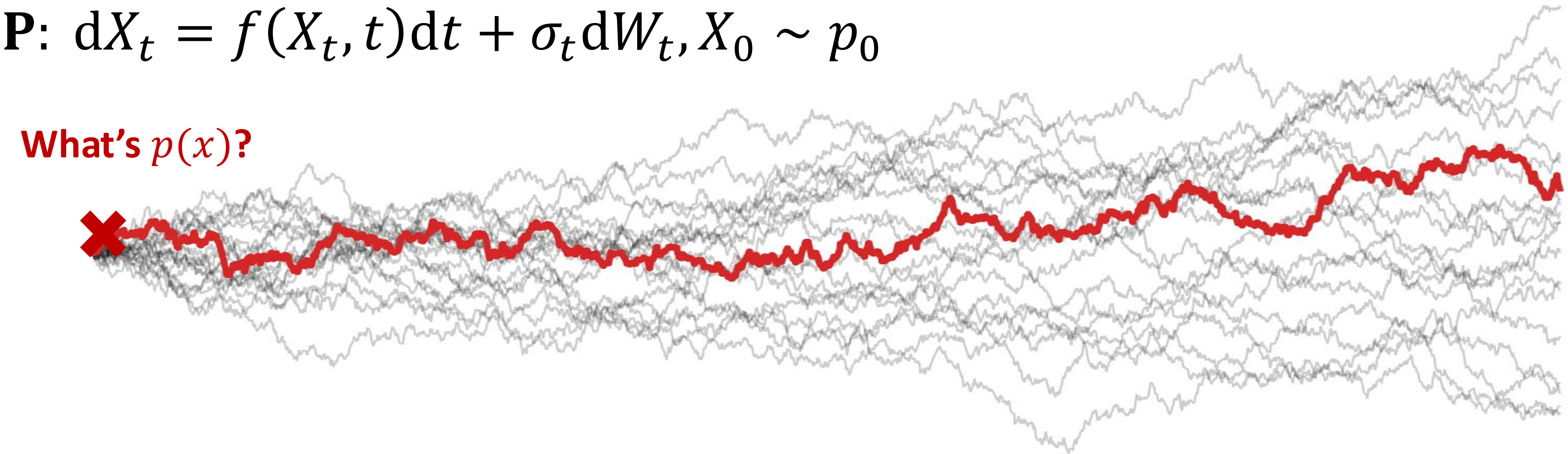


Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbf{P}: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$



Can be derived by PDEs as shown in [1] and [2]:

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025.

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$\overleftarrow{\mathbf{P}}: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$

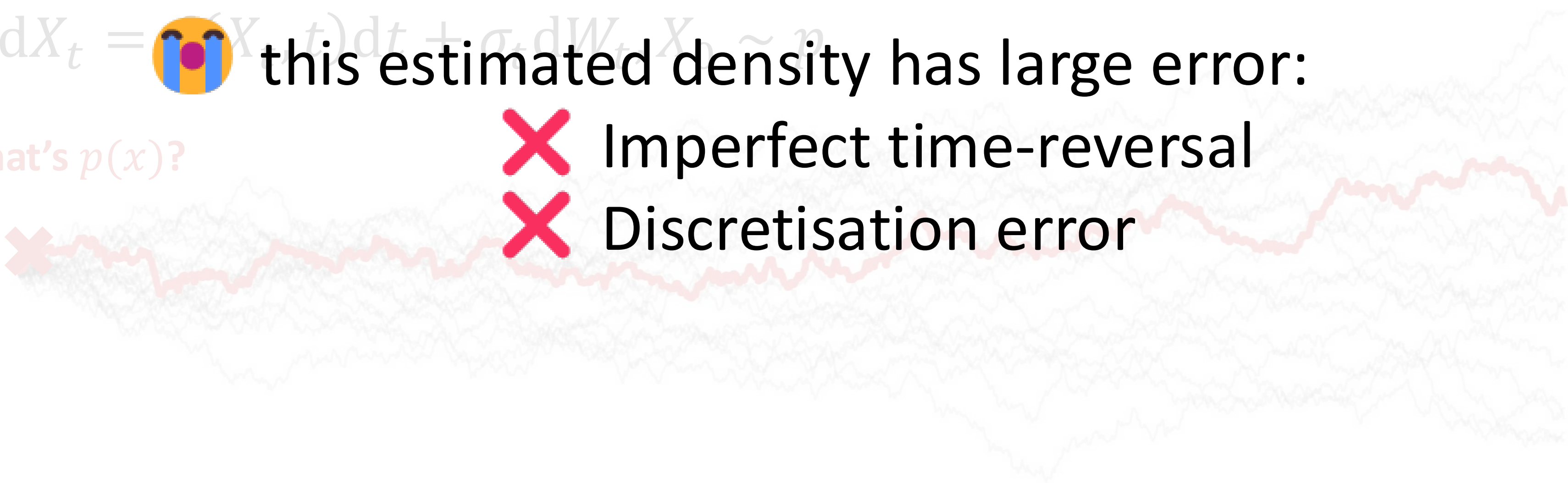
$\mathbf{P}: \; \mathrm{d}X_t = \ldots(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p$

😭 this estimated density has large error:

❌ Imperfect time-reversal

❌ Discretisation error

What's $p(x)$?



Can be derived by PDEs as shown in [1] and [2]:

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025.

Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbb{P}}: \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim N$$

$$\mathbb{P}: \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p$$

What's $p(x)$?

😭 this estimated density has large error:

❌ Imperfect time-reversal

❌ Discretisation error

✅ Only look at small intervals

Can be derived by PDEs as shown in [1] and [2]:

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025.

Importance Sampling with Path RND

- Only look at small intervals

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\Rightarrow \quad \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g\left(X_{[\tau,\tau']}\right) = 1$$

Importance Sampling with Path RND

✅ Only look at small intervals

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$
$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\Rightarrow \quad \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

👉 Equal density sampling:

for two diffusion models, find sample has **same density under both models**

👉 Calculate SMC weights:

for one (or several) DMs, **steer the distribution** by tilting/annealing/composition

👉 Energy regularisation:

Train **an energy-based diffusion** model by ensuring this relation for all time intervals

Example: Diffusion Inference-time Steering with Path RND

🛠️ Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

Example: Diffusion Inference-time Steering with Path RND

Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

Strategy:

Example: Diffusion Inference-time Steering with Path RND

Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

Strategy:

- Choose a heuristic guidance process;

Example: Diffusion Inference-time Steering with Path RND

🛠️ Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

🤖 Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;

Example: Diffusion Inference-time Steering with Path RND

🛠️ Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

🤖 Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do importance-resampling to move samples at $q_{t'}$ to q_t $(t < t')$

Example: Diffusion Inference-time Steering with Path RND

🛠️ Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

🤖 Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do importance-resampling to move samples at $q_{t'}$ to q_t $(t < t')$

Example: Diffusion Inference-time Steering with Path RND

Problem Setup:

Given a pretrained model for p_0, generate samples $\sim p_0(x)\exp(r(x))$

Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do importance-resampling to move samples at $q_{t'}$ to q_t $(t < t')$

We already learned about this pipeline from Raghav (Feynman-Kac Steering); Marta (Feynman-Kac Corrector); Luhuan (RDSMC) during the talks

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

- Define a sequence of intermediate target densities $q_t \propto \ p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

$$\mathrm{d}X_t = (\text{score} + \text{guidance})\,\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

"target"? $$X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$

“target”? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \quad X_{\tau'} \sim q_{\tau'}$

"target"? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{\text{target}}{\text{proposal}}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

"target"? $$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{\text{target}}{q_{\tau'}(X_{\tau'})\prod N_a(X_n | X_{n+1})}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

"proposal" $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

"target"? $$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_{\tau} \sim q_{\tau}$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{q_\tau(X_\tau)\prod N_b(X_{n+1}|X_n)}{q_{\tau'}(X_{\tau'})\prod N_a(X_n|X_{n+1})}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

“target”? $$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{q_\tau(X_\tau)\prod N_b(X_{n+1}|X_n)}{q_{\tau'}(X_{\tau'})\prod N_a(X_n|X_{n+1})}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $\quad \mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t},$

"target"? $\quad \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$

$R_f^g(X)$

$= \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)}$

- Define a sequence of intermediate target ... $(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{q_\tau(X_\tau) \prod N_b(X_{n+1} | X_n)}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n | X_{n+1})}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t},$

“target”? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$

$1/R_b^a(X_{[\tau,\tau']})$

- Define a sequence of intermediate target $(x_t)\exp(r_t(x_t));$
- Do **importance-resampling**

$$w \propto \frac{q_\tau(X_\tau)\prod N_b(X_{n+1}|X_n)}{q_{\tau'}(X_{\tau'})\prod N_a(X_n|X_{n+1})}$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

$$\text{“proposal”} \quad \mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

$$\text{“target”?} \quad \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{q_\tau(X_\tau)}{q_{\tau'}(X_{\tau'})} 1/R_b^a(X_{[\tau,\tau']})$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

“target”? $$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{p_\tau(x_\tau)\exp(r_\tau(x_\tau))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

$$\text{“proposal”} \quad \mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

$$\text{“target”?} \quad \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{\boxed{p_\tau(x_\tau)}\exp(r_\tau(x_\tau))}{\boxed{p_{\tau'}(x_{\tau'})}\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = \text{diffusion denoising}\,\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W}_t \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

- Choose a heuristic guidance process;

"proposal" $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W}_t, \qquad X_{\tau'} \sim q_{\tau'}$

"target"? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$

$$\frac{p_\tau(X_\tau)}{p_{\tau'}(X_{\tau'})} = ?$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto \frac{p_\tau(x_\tau)\exp(r_\tau(x_\tau))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = \text{diffusion denoising}\,\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W}_t \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}:\ \mathrm{d}X_t = \text{diffusion noising}\,\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\frac{p_\tau(X_\tau)}{p_{\tau'}(X_{\tau'})} = \,?$$

$$w \propto \frac{\boxed{p_\tau(x_\tau)}\exp(r_\tau(x_\tau))}{\boxed{p_{\tau'}(x_{\tau'})}\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_{\tau'} \sim q_{\tau'}$

"target"? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t} \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\frac{p_\tau(X_\tau)}{p_{\tau'}(X_{\tau'})} = ?$$

$$w \propto \frac{\boxed{p_\tau(x_\tau)}\exp(r_\tau(x_\tau))}{\boxed{p_{\tau'}(x_{\tau'})}\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?

Example: Diffusion Inference-time Steering with Path RND

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

"proposal" $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_{\tau'} \sim q_{\tau'}$

"target"? $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$\overleftarrow{\mathbf{P}}: \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W}_t \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}: \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

$$\frac{p_\tau(X_\tau)}{p_{\tau'}(X_{\tau'})} = R_f^g(X_{[\tau,\tau']})$$

$$w \propto \frac{p_\tau(x_\tau)\exp(r_\tau(x_\tau))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?

Example: Diffusion Inference-time Steering with Path RND

$$\overleftarrow{\mathbf{P}}:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W}_t \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$$

$$\Rightarrow \quad \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}(X_{[\tau,\tau']}) = 1$$

$$\frac{p_\tau(X_\tau)}{p_{\tau'}(X_{\tau'})} = R_f^g(X_{[\tau,\tau']})$$

have $\{x\} \sim q_{\tau'}$, how to obtain exact sample $\{x\} \sim q_\tau$

- Choose a heuristic guidance process;

“proposal” $\mathrm{d}X_t = a(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”? $\mathrm{d}X_t = b(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Example: Diffusion Inference-time Steering with Path RND

- Choose a heuristic guidance process;

"proposal" $$\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, \qquad X_{\tau'} \sim q_{\tau'}$$

"target" $$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \qquad X_\tau \sim q_\tau$$

- Define a sequence of intermediate target densities $q_t \propto p_t(x_t)\exp(r_t(x_t))$;
- Do **importance-resampling**

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Summary:

👉 Define proposal and target process

👉 Define intermediate densities q_t (by steering diffusion's p_t)

👉 Replace ratio between p_t by forward-backward kernel ratio R

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Summary:

👉 Define proposal and target process

👉 Define intermediate densities q_t (by steering diffusion's p_t)

👉 Replace ratio between p_t by forward-backward kernel ratio R

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Summary:

👉 Define proposal and target process

👉 Define intermediate densities q_t (by steering diffusion's p_t)

👉 Replace ratio between p_t by forward-backward kernel ratio R

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

Summary:

👉 Define proposal and target process

👉 Define intermediate densities q_t (by steering diffusion's p_t)

👉 Replace ratio between p_t by forward-backward kernel ratio R

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

🌟 Anneal target p_t^β

🌟 Composition/CFG between 2 diffusions $\left(p_t^{(1)}\right)^\beta \left(p_t^{(2)}\right)^\alpha$

Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau,\tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

🌟 Anneal target p_t^β

$$w \propto \left[R_f^g\left(X_{[\tau,\tau']}\right)\right]^\beta \; 1/R_b^a(X_{[\tau,\tau']})$$

🌟 Composition/CFG between 2 diffusions $\left(p_t^{(1)}\right)^\beta \left(p_t^{(2)}\right)^\alpha$

$$w \propto \left[R_{f_1}^{g_1}\left(X_{[\tau,\tau']}\right)\right]^\beta \left[R_{f_2}^{g_2}\left(X_{[\tau,\tau']}\right)\right]^\alpha 1/R_b^a(X_{[\tau,\tau']})$$

Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \exp\left(- \int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2} \mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2} \mathrm{d}t \right)$$

Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \exp\left(- \int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{f_t^2(X_t)}{2\sigma_t^2} \mathrm{d}t + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{g_t^2(X_t)}{2\sigma_t^2} \mathrm{d}t \right)$$

If we (1) plug-in this R,
and (2) choose target process f to be simple noising process:

We recover a path RND perspective for FKC [1]:

[1] Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., ... & Neklyudov, K. (2025). Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *ICML 2025.*

Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)}$$

Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \lim \frac{\prod N_g(X_n|X_{n+1})}{\prod N_h(X_{n+1}|X_n)}$$

If we (1) plug-in this R,
and (2) choose target process f to be simple noising process,
and (3) focus on reward-tilting

We recover a path RND perspective for TDS [1] / FKS[2]:

[1] Wu, L., Trippe, B., Naesseth, C., Blei, D., & Cunningham, J. P. (2023). Practical and asymptotically exact conditional sampling in diffusion models. *NeurIPS 2023.*
[2] Singhal, R., Horvitz, Z., Teehan, R., Ren, M., Yu, Z., McKeown, K., & Ranganath, R. (2025). A general framework for inference-time scaling and steering of diffusion models. *ICML 2025*.

Example: Diffusion Inference-time Steering with Path RND

Inference-time control with path RND (RN Estimator) [1]:

👉 flexible! (any target process, any proposal process)

👉 plug-and-play! (as long as there is some $p_t/p_{t'}$)

👉 Unifying! (Ito density estimator/FKC/TDS/FKS…)

[1] He, J., Hernández-Lobato, J. M., Du, Y., & Vargas, F. (2025). RNE: a plug-and-play framework for diffusion density estimation and inference-time control. *arXiv.*

Example: Diffusion Inference-time Steering with Path RND

Inference-time control with path RND (RN Estimator) [1]:

👉 flexible! (any target process, any proposal process)

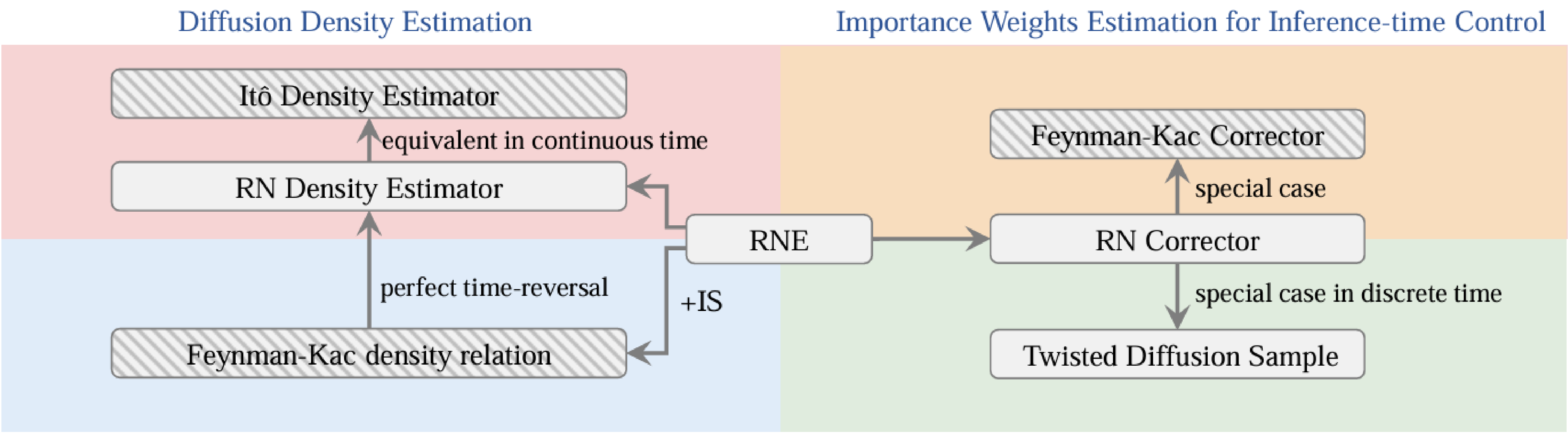
👉 plug-and-play! (as long as there is some $p_t/p_{t'}$)

👉 Unifying! (Ito density estimator/FKC/TDS/FKS...)

⚠️ **👉 Imperfect diffusion model: choice for cancellation (discrete time)** So far, we have assumed a perfectly pretrained diffusion model, giving us access to an SDE and its reversal. However, in practice, we typically encounter model imperfection and time-discretisation errors when calculating the marginal density ratio Eq. (12). Consequently, the resampled X_τ will not follow q_τ exactly, even as the number of samples $M \to \infty$. Fortunately, as we have the freedom to choose b_t, we can set it to obtain exact importance weights, despite the discretisation and score estimation errors.

[1] He, J., Hernández-Lobato, J. M., Du, Y., & Vargas, F. (2025). RNE: a plug-and-play framework for diffusion density estimation and inference-time control. *arXiv.*

Example: Diffusion Inference-time Steering with Path RND



Example: Diffusion Inference-time Steering with Path RND

Inference-time Annealing: (Alanine Dipeptide 800K to 300K)

	Metric	Energy TV(↓)	Distance TV(↓)	Sample W_2(↓)
	Anneal score (wo SMC)	0.794	0.023	**0.173***
	FKC	0.338	0.022	0.289
In theory = FKC 👉	RNC ($c_a = 1, c_b = 0$)	0.386	0.017	0.282
Flexible choices 👉	RNC ($c_a = 0.6, c_b = 0.4$)	**0.034**	**0.011**	0.253

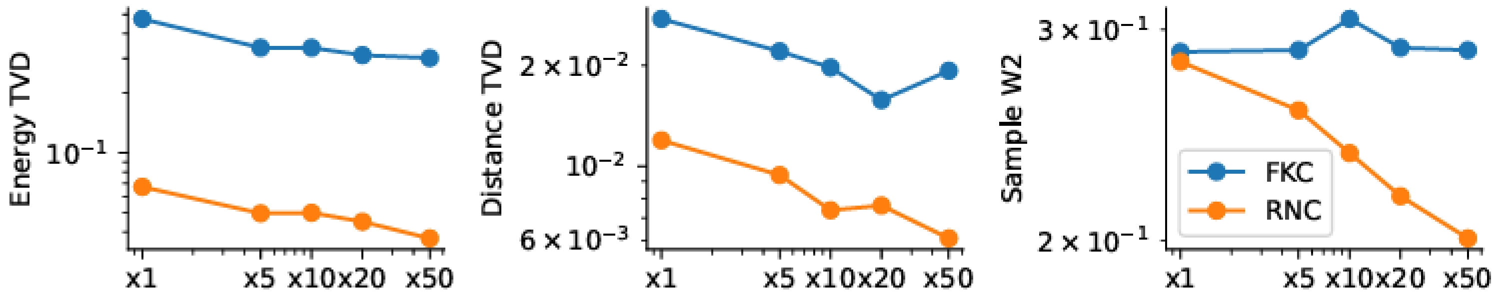


Figure 6: Performance scaling with particle numbers in SMC.

Takeaways So Far…

Path RND connects **transition kernels** with **marginal densities**

(known) (unknown)

👉 Density estimation
👉 SMC corrector

Takeaways So Far…

Path RND connects **transition kernels** with **marginal densities**

(known) (unknown)

👉 Density estimation
👉 SMC corrector
❓ What else? Where do we also need density?

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Example: Energy-parameterised Diffusion Training

Problem Setup:

train a diffusion model, parameterised by energy instead of score

$$\overleftarrow{\mathbf{P}}\text{: } \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

$$\mathbf{P}\text{: } \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$$

$$\Rightarrow \quad \frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g\left(X_{[\tau,\tau']}\right) = 1$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t} \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

$\mathbf{P}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$$\Rightarrow \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}(X_{[\tau,\tau']}) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau,\tau']}) = 1$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t}$ $\quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process 👉 $\mathbf{P}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$ $\quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$

$$\Rightarrow \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}(X_{[\tau,\tau']}) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g(X_{[\tau,\tau']}) = 1$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}$ $\quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process 👉 $\mathbf{P}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$ $\quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$$\Rightarrow \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

Learned network 👉 (numerator)

Learned network 👉 (denominator)

$$\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g\left(X_{[\tau,\tau']}\right) = 1$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}\colon \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t} \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process 👉 $\mathbf{P}\colon \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$

$$\Rightarrow \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

Learned network 👉

Learned network 👉

$$\left\| \frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g\left(X_{[\tau,\tau']}\right) - 1 \right\|^2$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}$ $\quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process 👉 $\mathbf{P}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$ $\quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$$\Rightarrow \frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

Learned network 👉

Learned network 👉

$$\left\|\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g\left(X_{[\tau,\tau']}\right) - 1\right\|^2 \qquad \tau' = \tau + \Delta t$$

Example: Energy-parameterised Diffusion Training

🛠️ Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network 👉 $\overleftarrow{\mathbf{P}}$: $\mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}$ $\quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process 👉 $\mathbf{P}$: $\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$ $\quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$

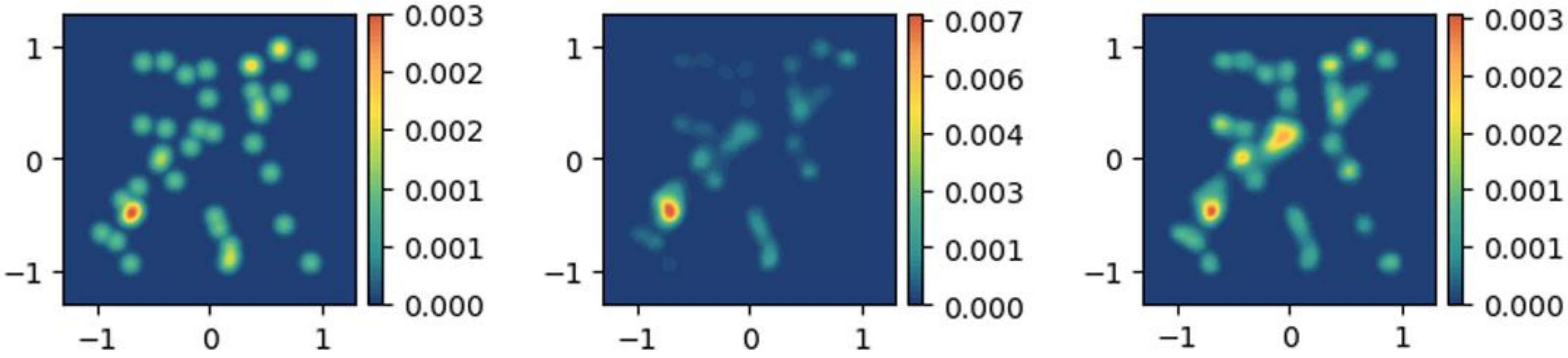
$$\Rightarrow \frac{\mathrm{d}\overleftarrow{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\left(X_{[\tau,\tau']}\right) = 1$$

Learned network 👉

Learned network 👉

$$\left\|\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g\left(X_{[\tau,\tau']}\right) - 1\right\|^2_{\tau' = \tau + \Delta t} + \mathrm{DSM}$$

Example: Energy-parameterised Diffusion Training

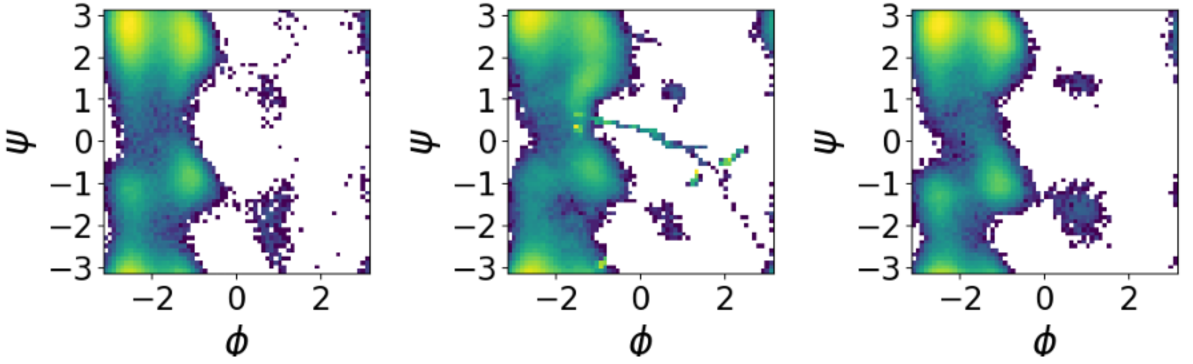


(a) Ground truth (b) DSM (c) RNE regularise

Example: Energy-parameterised Diffusion Training

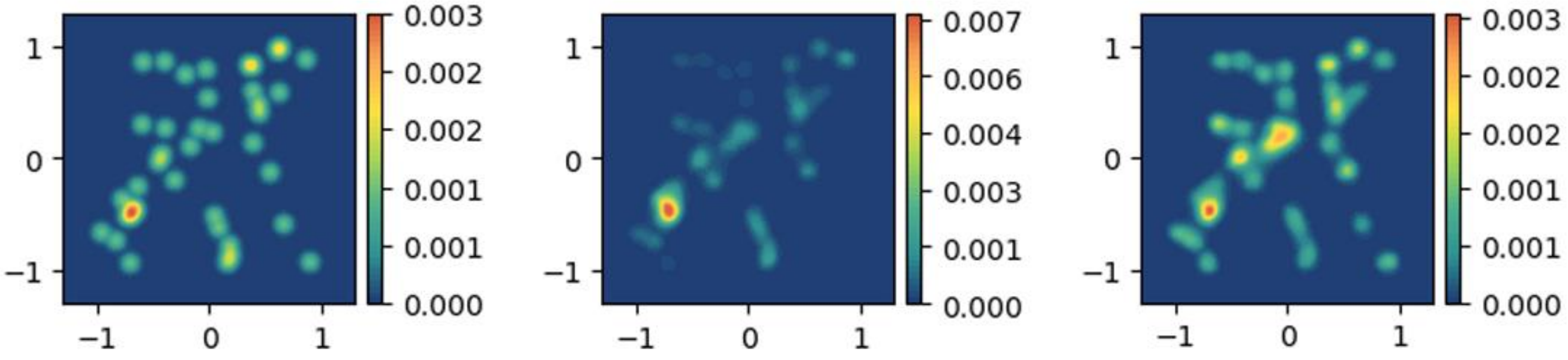
Table 4: Quality of samples obtained by running denoising process (denoted as DM) and running MCMC on learned energy at $t = 0$.

Training method	Sample Method	Sample $W2$
DSM	**DM**	0.1811
	MCMC	0.9472
RNE Reg	**DM**	0.1809
	MCMC	0.1836



(a) Ground truth (b) DSM (c) RNE regularise

Figure 8: Ramachandran plot of samples by MCMC (with Metropolis–Hastings) on learned energy.



(a) Ground truth (b) DSM (c) RNE regularise

Importance Sampling with Path RND

Importance Sampling:

$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)} f(x)\right]$$

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) f(X)\right]$$

Importance Sampling with Path RND

Importance Sampling:

$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)} f(x)\right]$$

❓ What is the optimal proposal for $f > 0$?

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) f(X)\right]$$

Importance Sampling with Path RND

Importance Sampling:

$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)}f(x)\right]$$

❓ What is the optimal proposal for $f > 0$?

$$q \propto pf$$

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \propto f$$

Importance Sampling with Path RND

Likelihood

Importance Sampling:

$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)}f(x)\right]$$

❓ What is the optimal proposal for $L > 0$?

$$q \propto pL$$

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X)f(X)\right]$$

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \propto L$$

Variational Inference with Path RND

Importance Sampling:

$$\mathbf{E}_{x\sim p}[f(x)] = \mathbf{E}_{x\sim q}\left[\frac{p(x)}{q(x)} f(x)\right]$$

Importance Sampling in Path space:

$$\mathbf{E}_{X\sim \mathbf{P}}[f(X)] = \mathbf{E}_{X\sim \mathbf{Q}}\left[\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) f(X)\right]$$

Variational Inference:

$$q^* \propto pL$$

$$q = \min D[q, q^*]$$

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min D[\mathbf{Q}, \mathbf{Q}^*]$$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min D[\mathbf{Q}, \mathbf{Q}^*]$$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min D[\mathbf{Q}, \mathbf{Q}^*]$$

$$\mathbf{P}: \ \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

$$\mathbf{Q}: \ \mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathrm{KL}[\mathbf{Q}||\mathbf{Q}^*]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] \quad = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} + \log \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}^*}\right]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] \quad = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \log L\right]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] \quad = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \log L\right]$$

$$\mathbf{P}\colon\ \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

$$\mathbf{Q}\colon\ \mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

Forward-forward RND and Girsanov

$$\mathbf{P}:\ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\mathbf{Q}:\ \mathrm{d}X_t = h(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}(X) = \underbrace{\frac{p(X_0)}{q(X_0)}}_{\text{Initial density ratio}} \exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t}_{\text{Forward Ito Integral } \int a_t(X_t)\cdot \mathrm{d}X_t = \lim \sum a_n(X_n)\cdot(X_{n+1}-X_n)} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] = \min \mathbf{E}_{\mathbf{Q}}\left[\int \frac{1}{2}||u_t||^2 \, \mathrm{d}t - \log L\right]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] = \min \mathbf{E}_{\mathbf{Q}}\left[\int \frac{1}{2}\left|\left|u_t\right|\right|^2 \mathrm{d}t - \log L\right]$$

$\mathbf{P}$: $\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

$\mathbf{Q}$: $\mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$

stochastic optimal control

Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{\mathrm{d}\mathbf{Q}^*}{\mathrm{d}\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}}\left[\log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{Q}^*}\right] = \min \mathbf{E}_{\mathbf{Q}}\left[\int \frac{1}{2}\left|\left|u_t\right|\right|^2 \mathrm{d}t - \log L\right]$$

$$\mathbf{P}:\ \mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

$$\mathbf{Q}:\ \mathrm{d}X_t = (b + \sigma_t u)(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_0$$

stochastic optimal control

👉 **fine-tune diffusion models [1]**

👉 **neural samplers [2,3,4…]**

[1] Domingo-Enrich, C., Drozdzal, M., Karrer, B., & Chen, R. T. (2024). Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. ICLR 2025

[2] Havens, A., Miller, B. K., Yan, B., Domingo-Enrich, C., Sriram, A., Wood, B., ... & Chen, R. T. (2025). Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. ICML 2025.

[3] Liu, G. H., Choi, J., Chen, Y., Miller, B. K., & Chen, R. T. (2025). Adjoint Schrodinger Bridge Sampler. arXiv.

[4] Zhu, Y., Guo, W., Choi, J., Liu, G. H., Chen, Y., & Tao, M. (2025). MDNS: Masked Diffusion Neural Sampler via Stochastic Optimal Control. arXiv.

Collaborators

Fantastic Path RNDs

- 🪄*intuitive understand from sequence of Gaussian kernels*
- 🪄*Forward-forward RND with Girsanov theorem*
- 🪄*Forward-backward RND*

Where to Find Them?

- *Importance Sampling with Path RND*
  - *AIS, Jarzynski and Crook's Fluctuation Theorem*
  - *Free-energy estimation*
  - *Density estimation*
  - *Generation control*
  - *Energy regularisation*

- *Variational Inference with Path RND*
  - *Neural samplers*
  - *Diffusion model fine-tuning*

Thank You!

Jiajun He (jh2383@cam.ac.uk)
Francisco Vargas (vargfran@gmail.com)