Introduction to Diffusion Models and Probabilistic Inference with Path Measures

Jiajun He

University of Cambridge

05/11/2025

Episode 1

Diffusion Models as Probabilistic Models in Path Space

We will discuss:

- Recap of Probability Theory and Probabilistic Models
- Generative Models and Diffusion Models
- Path Measure
- Probabilistic Inference with Path Measures: Control your Generation
- Application Demo & What’s Next?

Recap of Probability

- Random variable (RV):

  A *function* mapping from a sample space e.g., *{rain tmr, not rain tmr}* to a measurable space e.g., *{0, 1}*.

- Probability mass function (discrete RV), e.g.,

$$P(X=1)=0.7$$
$$P(X=0)=0.3$$

- Probability density function (continuous RV), e.g.,

$$N(x|0,1)\propto \exp(-x^2/2)$$

Recap of Probability

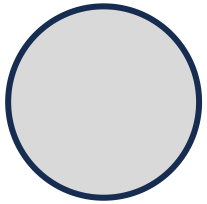
- Joint, Condition, marginal and Bayes' Rule:

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y)$$

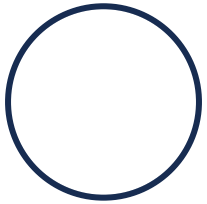
$$p(x) = \int p(x, y)\mathrm{d}x$$

Recap of Probability

- Graphical Model:



Random Variable

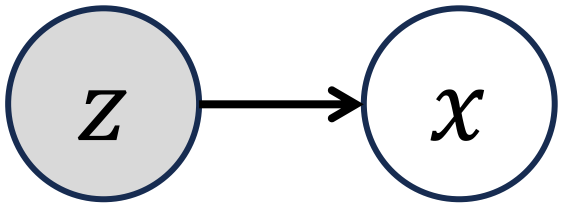


Observed Random Variable

Dependency (conditional distribution)

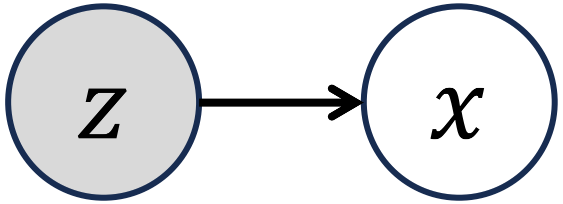
Recap of Probability

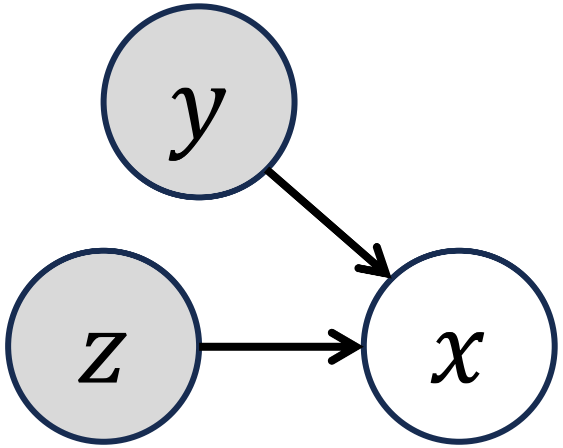
- Graphical Model:



Recap of Probability

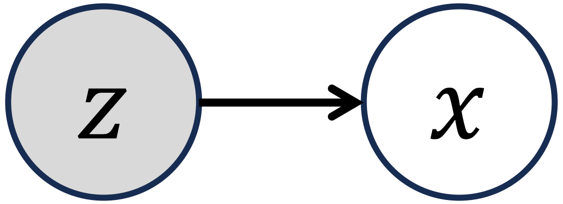
- Graphical Model:

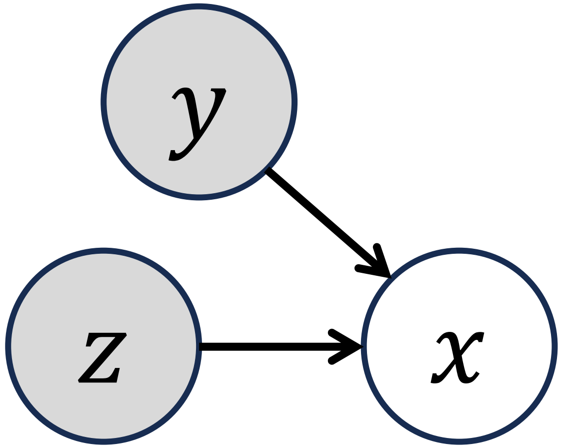


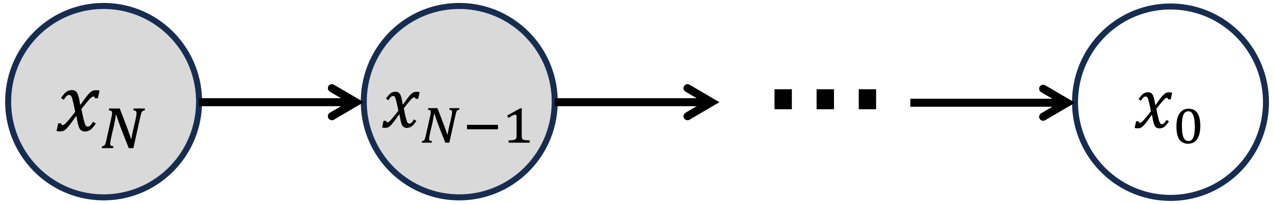


Recap of Probability

- Graphical Model:

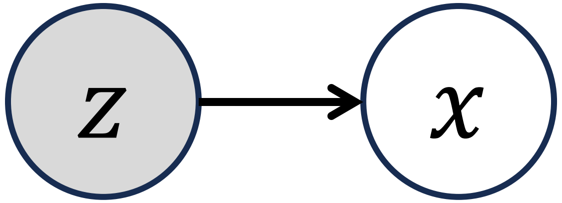






Generative Models and Diffusion Models

Generative Model:



Prior: $p(z)$

Likelihood: $p(x|z)$

Posterior: $p(z|x) \approx q(z|x)$

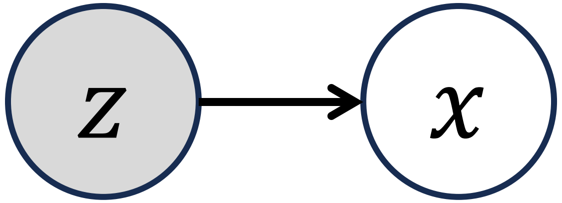
Variational Inference:

$$q(z|x)p(x) \approx \; p(z)p(x|z)$$

$$i.e., D_{\mathrm{KL}}[q(z|x)p(x) \;||\; p(z)p(x|z)]$$

Generative Models and Diffusion Models

Generative Model:



Prior: $p(z)$

Likelihood: $p(x|z)$

Posterior: $p(z|x) \approx q(z|x)$

Variational Inference:

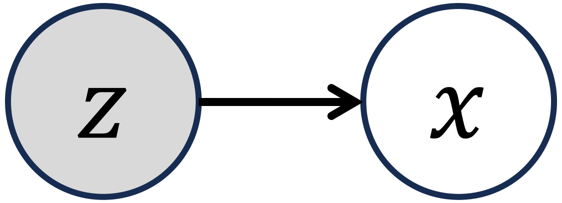
$$q(z|x)p(x) \approx p(z)p(x|z)$$

$$i.e., D_{\mathrm{KL}}[q(z|x)p(x) \,||\, p(z)p(x|z)]$$

-> Evidence Lower Bound

Generative Models and Diffusion Models

Generative Model:



Prior: $p(z)$

Likelihood: $p(x|z)$

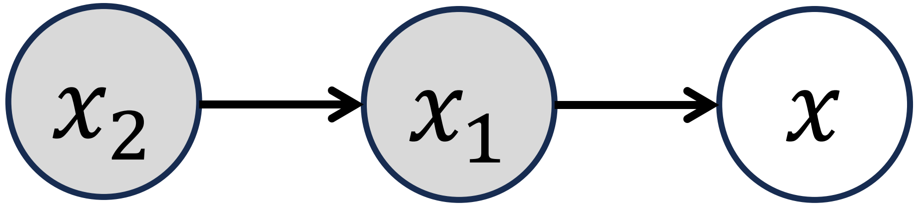
Posterior: $p(z|x) \approx q(z|x)$

Fix prior and likelihood, infer posterior

Fix prior, learn likelihood and posterior

Fix prior and posterior, learn likelihood

Generative Models and Diffusion Models



Prior: $p(x_2)p(x_1|x_2)$

Likelihood: $p(x|x_1, x_2) = p(x|x_1)$

Posterior: $p(x_1, x_2|x) \approx q(x_1|x)q(x_2|x_1)$

Variational Inference: $q(x_1|x)q(x_2|x_1)p(x) \approx p(x_2)p(x_1|x_2)p(x|x_1)$

Generative Models and Diffusion Models

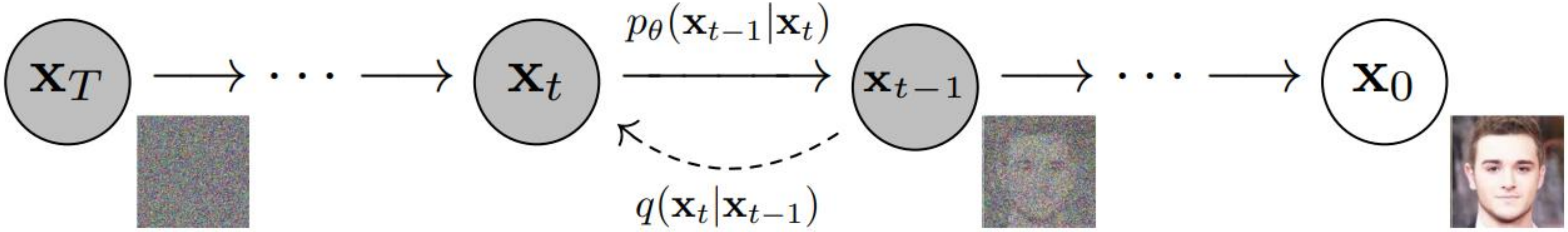


Figure 2: The directed graphical model considered in this work. Figure from [1].

[1] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *NeurIPS 2020.*

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)$

$p(x_0)$

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)$

$p(x_0)$

$q(x_1|x_0)$

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)$

$p(x_0)$

$q(x_1|x_0)$

$q(x_2|x_1)$

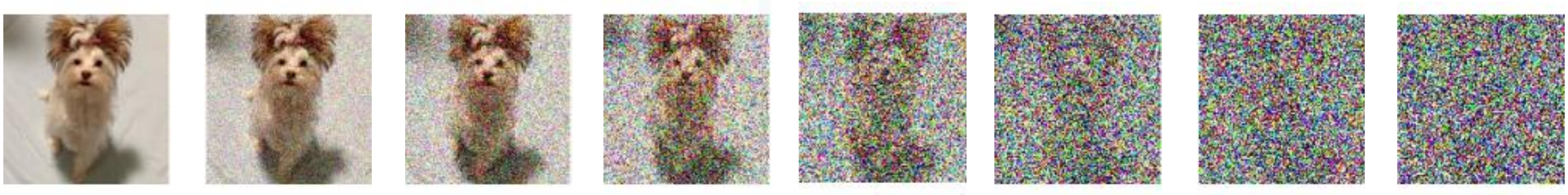
Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)$ ...

$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ ...

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ \ldots q(x_T|x_{T-1})$



$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ … $q(x_T|x_{T-1})$

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$

$p(x_0)$

$q(x_1|x_0)$

$q(x_2|x_1)$

…

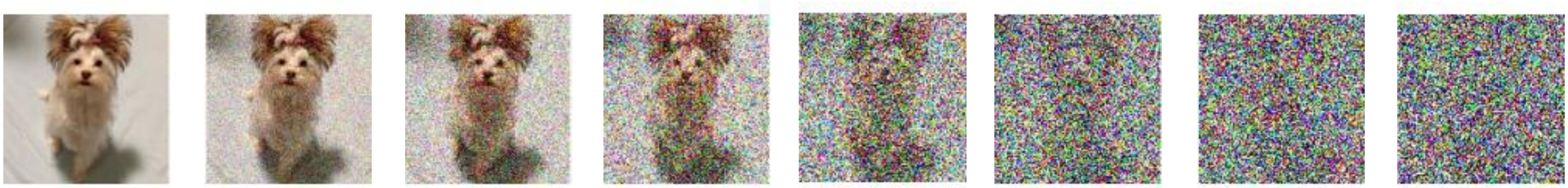
$q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)$

$p(x_T)$

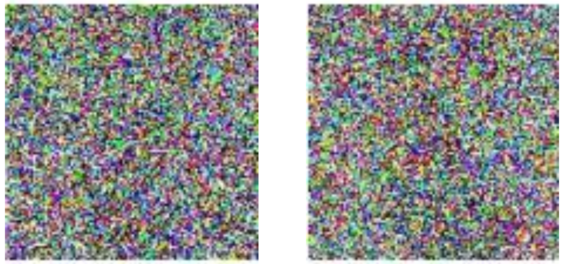
Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$



$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ … $q(x_T|x_{T-1})$

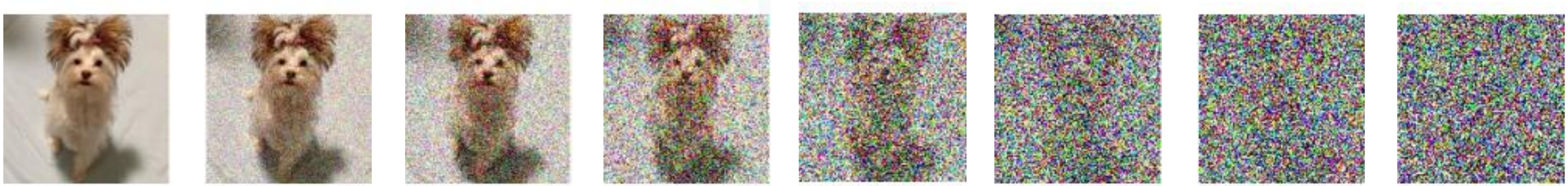
Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)$



$p(x_T)$ $p(x_{T-1}|x_T)$

Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$



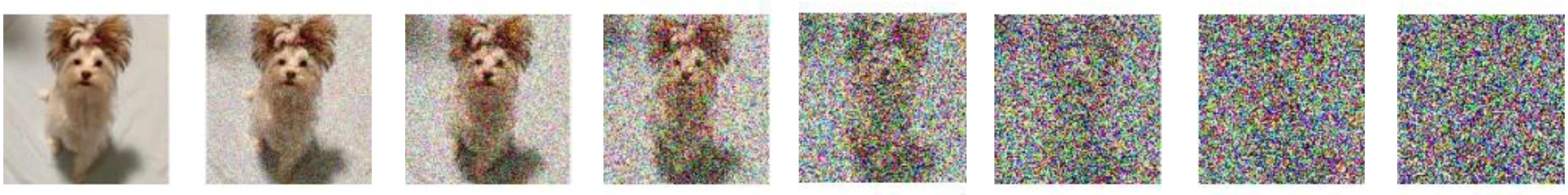
$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ … $q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\ldots$

$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ …

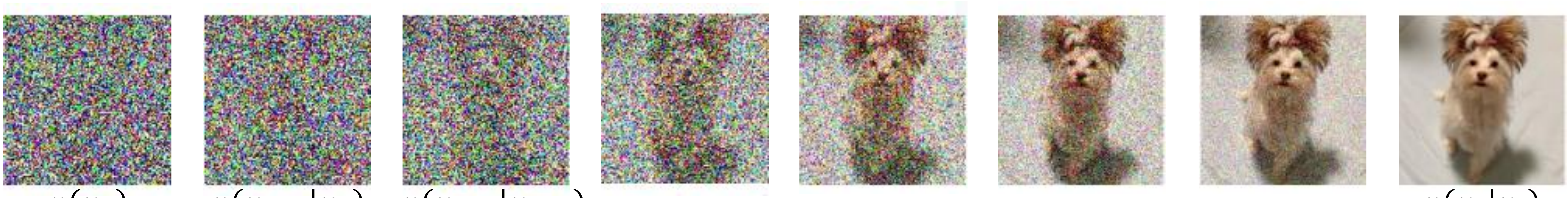
Generative Models and Diffusion Models

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$



$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ … $q(x_T|x_{T-1})$

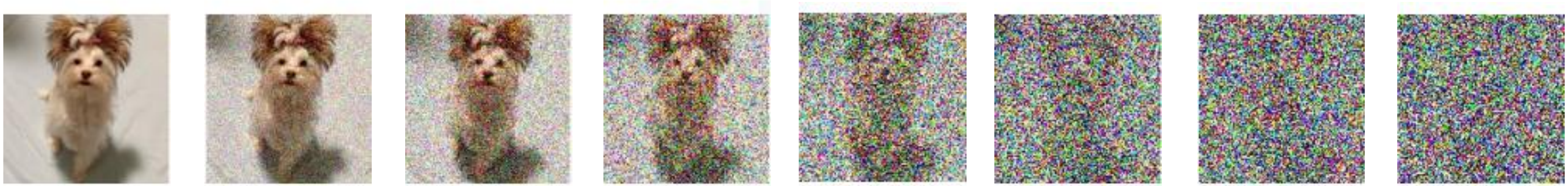
Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)$



$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ … $p(x_0|x_1)$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $x_{t'} = x_t - \beta_t x_t \mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\epsilon$



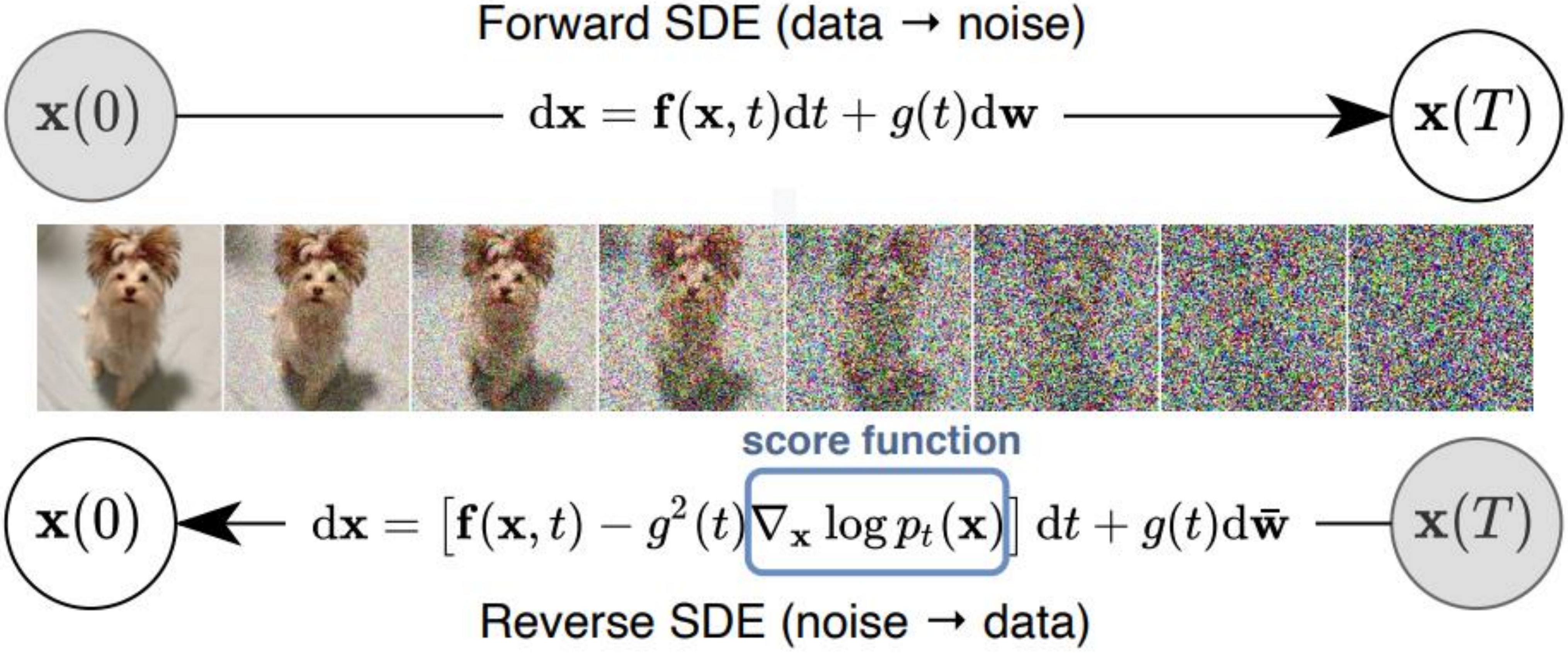
$p(x_0)$ $q(x_1|x_0)$ $q(x_2|x_1)$ … $q(x_T|x_{T-1})$

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \log p_t(x_t)]\, \mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\epsilon'$$

$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ … $p(x_0|x_1)$

Generative Models and Diffusion Models



[2] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *ICLR 2021.*

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = -\beta_t x_t \mathrm{d}t + \sqrt{2\beta_t}\, \mathrm{dW}_t$
Backward SDE (noise->data):

$$\mathrm{d}x_t = \ [-\beta_t x_t - 2\beta_t \nabla \log p_t(x_t)]\, \mathrm{d}t + \sqrt{2\beta_t} \mathrm{dW}_t^{-}$$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = -\beta_t x_t \mathrm{d}t + \sqrt{2\beta_t}\, \mathrm{dW}_t$
Backward SDE (noise->data):

$$\mathrm{d}x_t = \left[-\beta_t x_t - 2\beta_t \nabla \log p_t(x_t)\right] \mathrm{d}t + \sqrt{2\beta_t}\mathrm{dW}_t^{-}$$

The DDPM Kernel is one way to discretise the SDEs.

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = -\beta_t x_t \mathrm{d}t + \sqrt{2\beta_t}\,\mathrm{dW}_t$
Backward SDE (noise->data):

$$\mathrm{d}x_t = \left[-\beta_t x_t - 2\beta_t \nabla \log p_t(x_t)\right] \mathrm{d}t + \sqrt{2\beta_t}\mathrm{dW}_t^{-}$$

The DDPM Kernel is one way to discretise the SDEs.
Other options exist

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{d}t + \sigma_t \,\mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{d}t + \sigma_t \,\mathrm{dW}_t^-$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t^-$

Euler–Maruyama discretisation

$$x_{t+1} = x_t + f_t(x_t)\Delta\mathrm{t} + \sqrt{\sigma_t \Delta t}\epsilon$$
$$x_{t-1} = x_t - g_t(x_t)\Delta\mathrm{t} + \sqrt{\sigma_t \Delta t}\epsilon'$$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{d}t + \sigma_t \, \mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{d}t + \sigma_t \, \mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1) \ldots q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)p(x_{T-1}|x_T) \ldots p(x_0|x_1)$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{d}t + \sigma_t \,\mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{d}t + \sigma_t \,\mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\dots q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\dots p(x_0|x_1)$

$$p(x_0)q(x_1|x_0)q(x_2|x_1)\dots q(x_T|x_{T-1}) \approx p(x_T)p(x_{T-1}|x_T)\dots p(x_0|x_1)$$

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{dt} + \sigma_t\,\mathrm{dW}_t$
Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{dt} + \sigma_t\,\mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\dots q(x_T|x_{T-1})$
Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\dots p(x_0|x_1)$

$$p(x_0)q(x_1|x_0)q(x_2|x_1)\dots q(x_T|x_{T-1}) \approx p(x_T)p(x_{T-1}|x_T)\dots p(x_0|x_1)$$

“Forward SDE and backward SDE define the same joint distribution”

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{d}t + \sigma_t\, \mathrm{dW}_t$
Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{d}t + \sigma_t\, \mathrm{dW}_t^-$

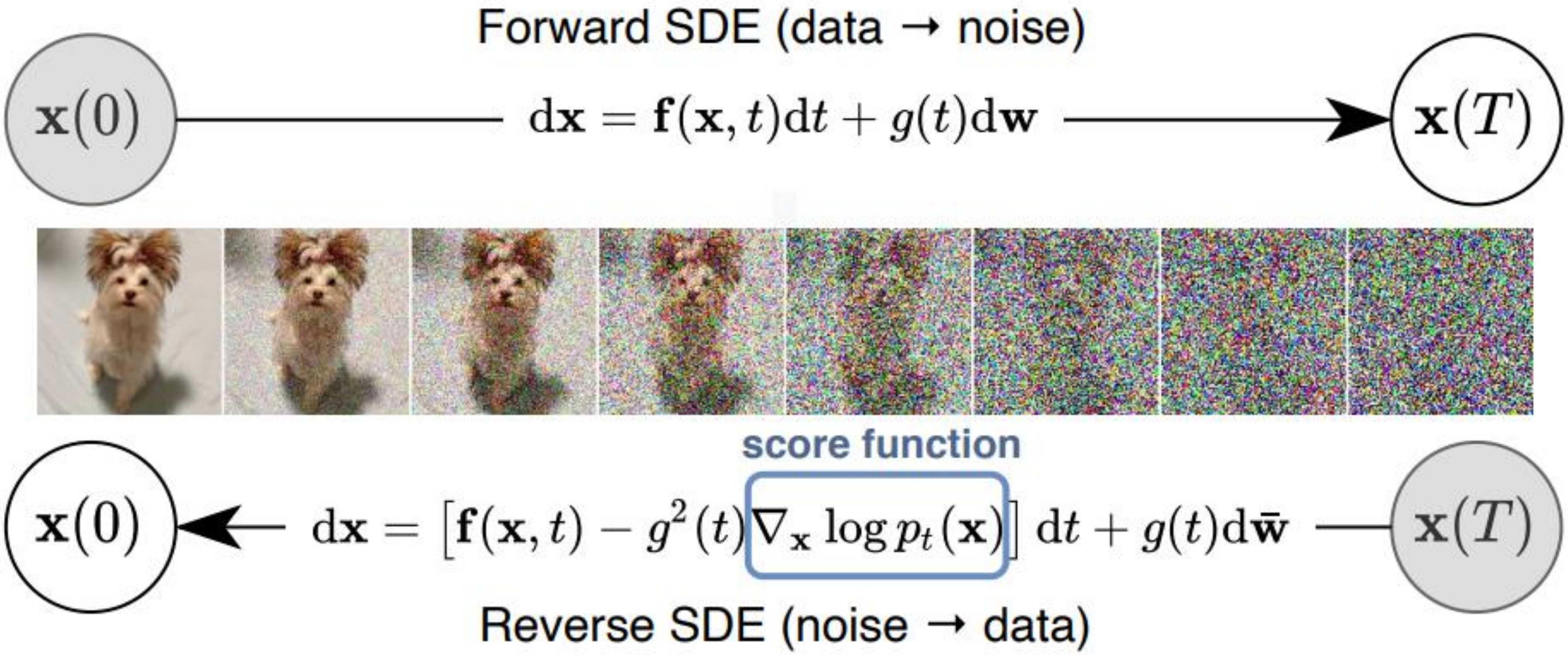
"Forward SDE and backward SDE define the same joint distribution"

iff.

$$g_t = f_t - \sigma_t^2 \nabla \log p_t$$

Nelson's relation

Generative Models and Diffusion Models



[2] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *ICLR 2021.*

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)$

$$p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1}) \approx p(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)$$

*"Forward SDE and backward SDE define the **same joint distribution**"*

Generative Models and Diffusion Models

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{dt} + \sigma_t \, \mathrm{dW}_t$
Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{dt} + \sigma_t \, \mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1) \ldots q(x_T|x_{T-1})$
Generation (noise->data): $p(x_T)p(x_{T-1}|x_T) \ldots p(x_0|x_1)$

$$p(x_0)q(x_1|x_0)q(x_2|x_1) \ldots q(x_T|x_{T-1}) \approx p(x_T)p(x_{T-1}|x_T) \ldots p(x_0|x_1)$$

*"Forward SDE and backward SDE define the **same joint distribution**"*

same joint distribution over path $x_0, x_1, \cdots, x_T$

Diffusion Models and **Path Measures**

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{dt} + \sigma_t\, \mathrm{dW}_t^-$

Posterior (data -> noise): $p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})$

Generation (noise->data): $p(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)$

$$p(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1}) \approx p(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)$$

*"Forward SDE and backward SDE define the **same joint distribution**"*

same joint distribution over path $x_0, x_1, \cdots, x_T$

Path Measures and Ito's Calculus

Forward SDE (data -> noise): $\mathrm{d}x_t = f_t(x_t)\mathrm{d}t + \sigma_t\,\mathrm{dW}_t$

Backward SDE (noise->data): $\mathrm{d}x_t = g_t(x_t)\mathrm{d}t + \sigma_t\,\mathrm{dW}_t^{-}$

$$\lim \frac{p_0(x_0)q(x_1|x_0)q(x_2|x_1)\ldots q(x_T|x_{T-1})}{p_T(x_T)p(x_{T-1}|x_T)\ldots p(x_0|x_1)}$$

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

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

Episode 1

Diffusion Models are

👉 Deep hierarchical VAEs

👉 Reverse SDEs

Path Measure is just

👉 Sequence of Gaussian densities (to the limit)

👉 something involving Ito’s integral 👻

More Math Details?

“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

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

⏮️ Backward SDE

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

💡 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 $\{X_1, X_2, \dots X_N\}$, 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:** $p(X_1, X_2, \dots X_N) = p(X_1) \prod p(X_{n+1}|X_n)$

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

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

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

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

Initial density ratio

Forward-backward RND

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

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

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

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

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

iff

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

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

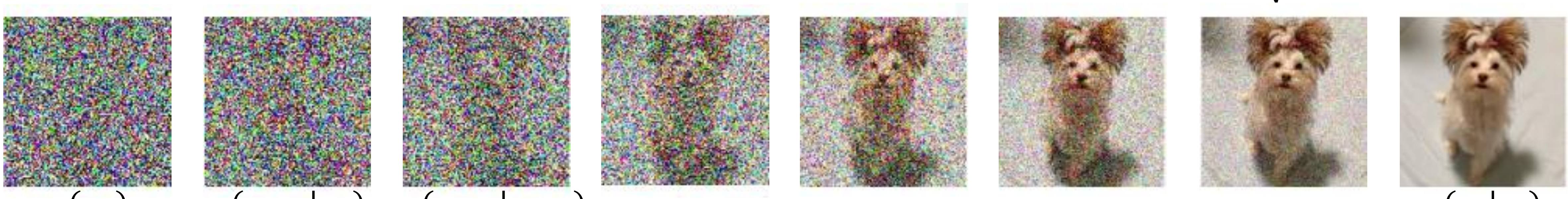
Episode 2

Control Generation with Sequential Monte Carlo in Path Space

Generation Control of Diffusion Models

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \log p_t(x_t)]\, \mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\,\epsilon'$$

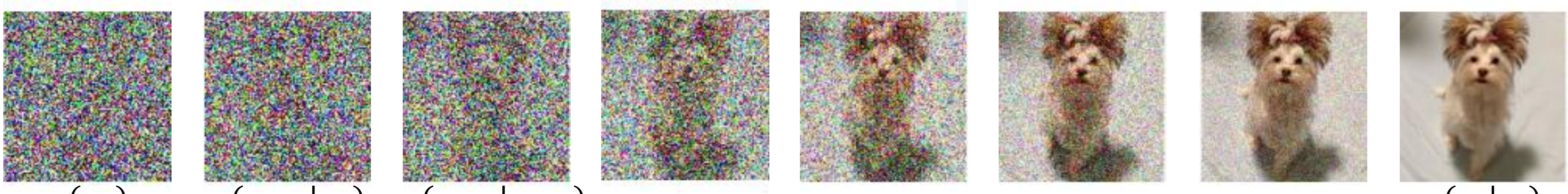


$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ … $p(x_0|x_1)$

Generation Control of Diffusion Models

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \log p_t(x_t)]\,\mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\,\epsilon'$$



$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ … $p(x_0|x_1)$

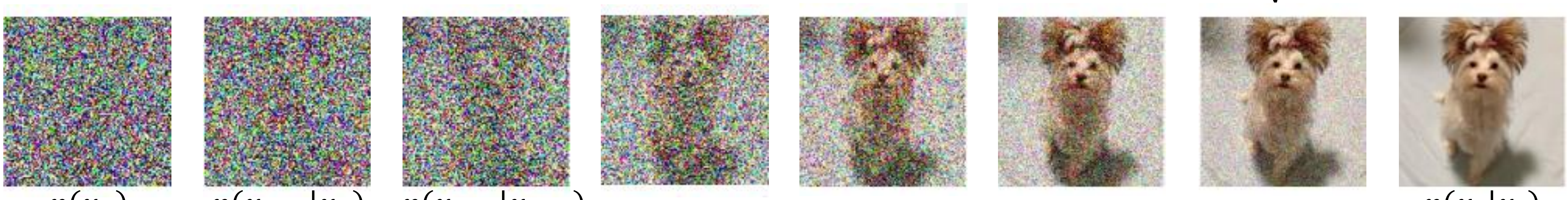
What if I want to generate samples:

👉satisfying certain constraint
👉satisfying certain reward
👉composing properties of two diffusion models
👉has sharper distribution
…

Generation Control of Diffusion Models

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \mathrm{log} p_t(x_t)]\,\mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\,\epsilon'$$



$p(x_T)$ $p(x_{T-1}|x_T)$ $p(x_{T-2}|x_{T-1})$ … $p(x_0|x_1)$

What if I want to generate samples:

👉satisfying certain constraint $q(x_0) \propto p(x_0)1\{x_0 \in \text{constraint family}\}$

👉satisfying certain reward $q(x_0) \propto p(x_0)\exp(r(x_0))$

👉composing properties of two diffusion models $q(x_0) \propto p(x_0)p'(x_0)$

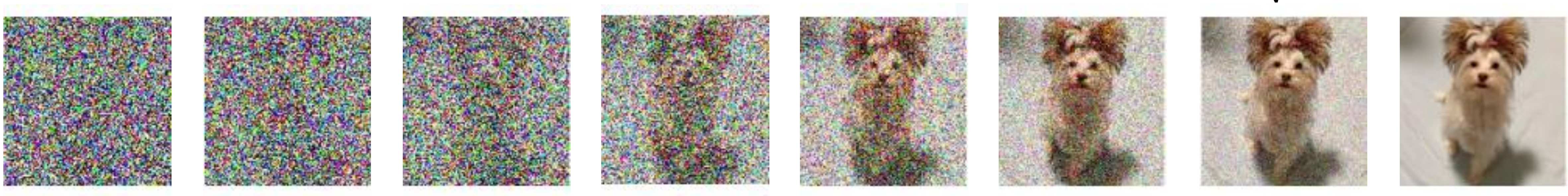
👉has sharper distribution $q(x_0) \propto p(x_0)^\alpha$

…

Generation Control of Diffusion Models

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \log p_t(x_t)]\,\mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\epsilon'$$



Options:

- Generate N samples, find the best set of samples

Generation Control of Diffusion Models

Backward SDE (noise->data):

$$x_{t'} = x_t + [\beta_t x_t + 2\beta_t \nabla \log p_t(x_t)]\,\mathrm{d}t + \sqrt{2\beta_t \mathrm{d}t}\epsilon'$$

Options:

- Generate N samples, find the best set of samples
- Generate n samples at each step, find the best set of samples for next step

Generation Control of Diffusion Models

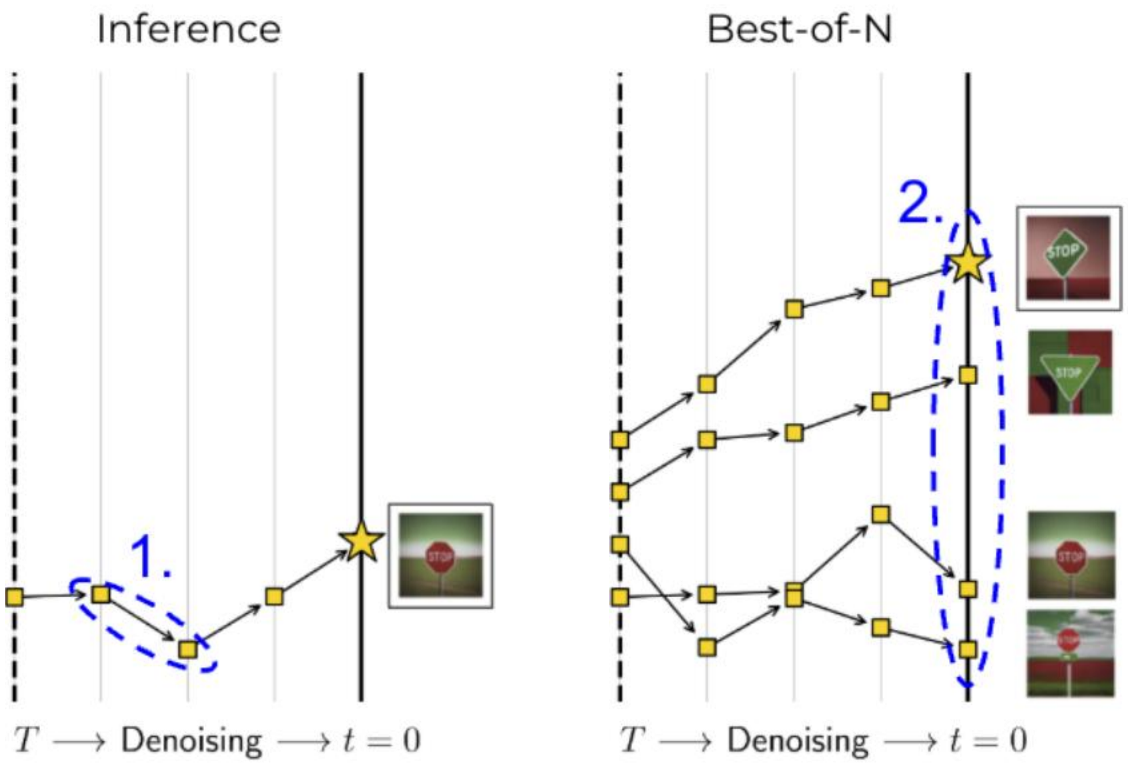


Figure taken from
Singhal, Raghav, et al. "A general framework for inference-time scaling and steering of diffusion models." *ICML 2025*

Generation Control of Diffusion Models

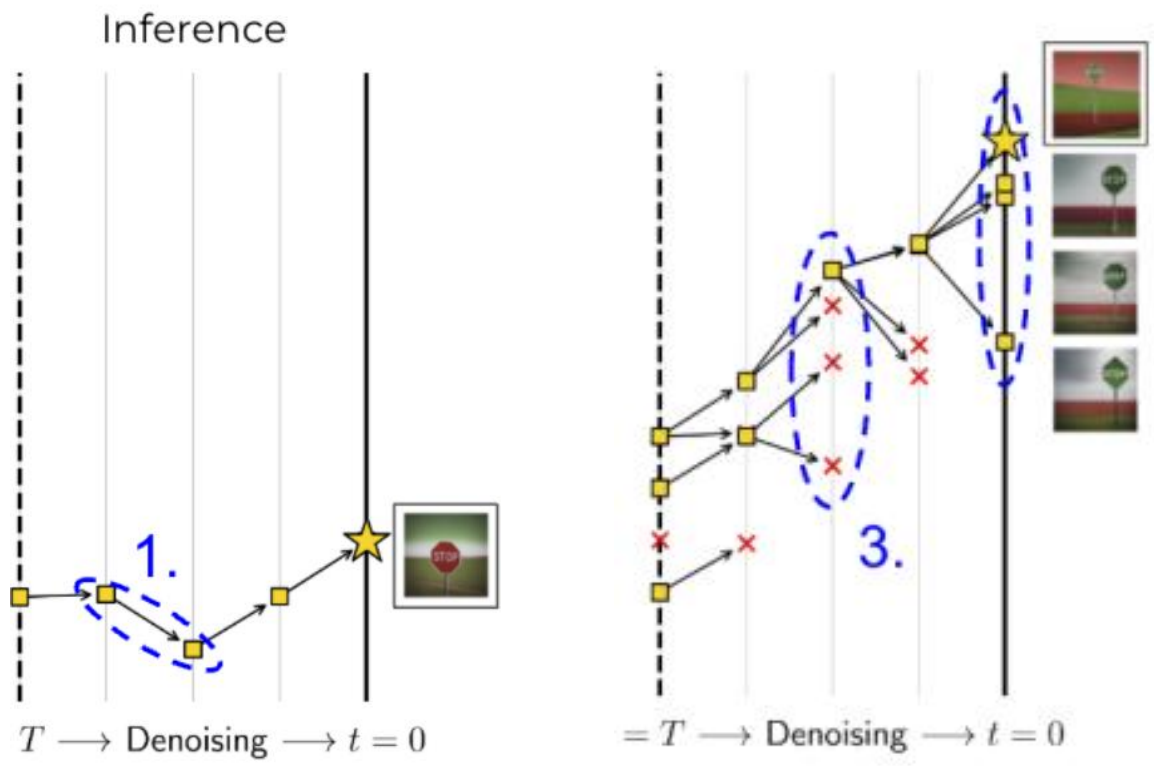
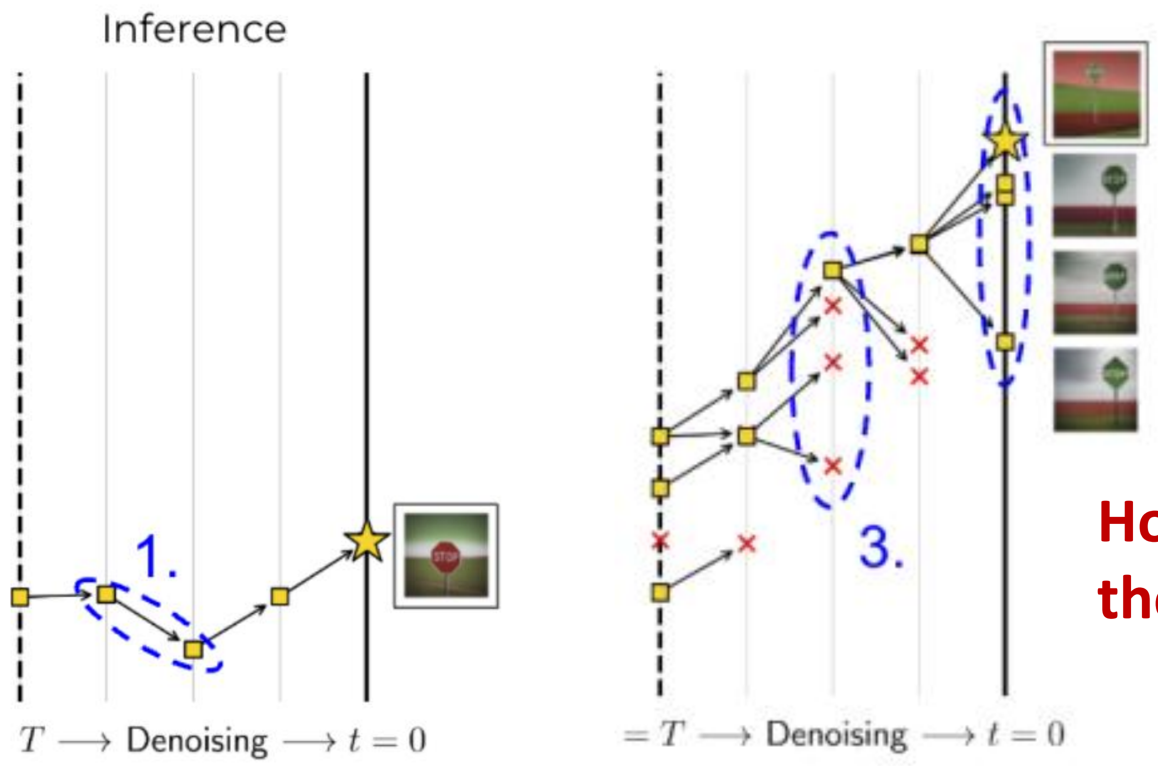


Figure taken from
Singhal, Raghav, et al. "A general framework for inference-time scaling and steering of diffusion models." *ICML 2025*

Generation Control of Diffusion Models



How to pick the "best" set of samples?

Figure taken from
Singhal, Raghav, et al. "A general framework for inference-time scaling and steering of diffusion models." *ICML 2025*

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$

But we **want to draw** sample from $p(x)$

HOW?

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

HOW?

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

HOW?

$$E_p[f(x)] = E_q\left[f(x)\frac{p(x)}{q(x)}\right]$$

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

HOW?

$$E_p[f(x)] = E_q\left[f(x)\boxed{\frac{p(x)}{q(x)}}\right]$$

Importance Weight

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

HOW?

$$E_p[f(x)] = E_q\left[f(x)\boxed{\frac{p(x)}{q(x)}}\right]$$

Importance Weight

💡 Requirements on $q(x)$: can sample & eval density

💡 Requirements on p(x): can eval density

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

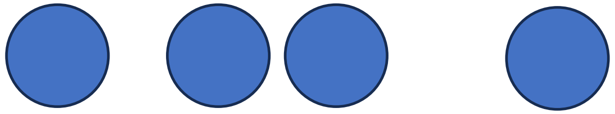
1. Draw $x_1, x_2, \cdots, x_N \sim q$
2. Calculate $\frac{p(x)}{q(x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim \mathrm{Cat}\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)$
4. Return $x_{i_1}, x_{i_2}, \cdots, x_{i_M}$

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

1. **Draw $\boldsymbol{x_1, x_2, \cdots, x_N \sim q}$**
2. Calculate $\frac{p(x)}{q(x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim \mathrm{Cat}\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)$
4. Return $x_{i_1}, x_{i_2}, \cdots, x_{i_M}$

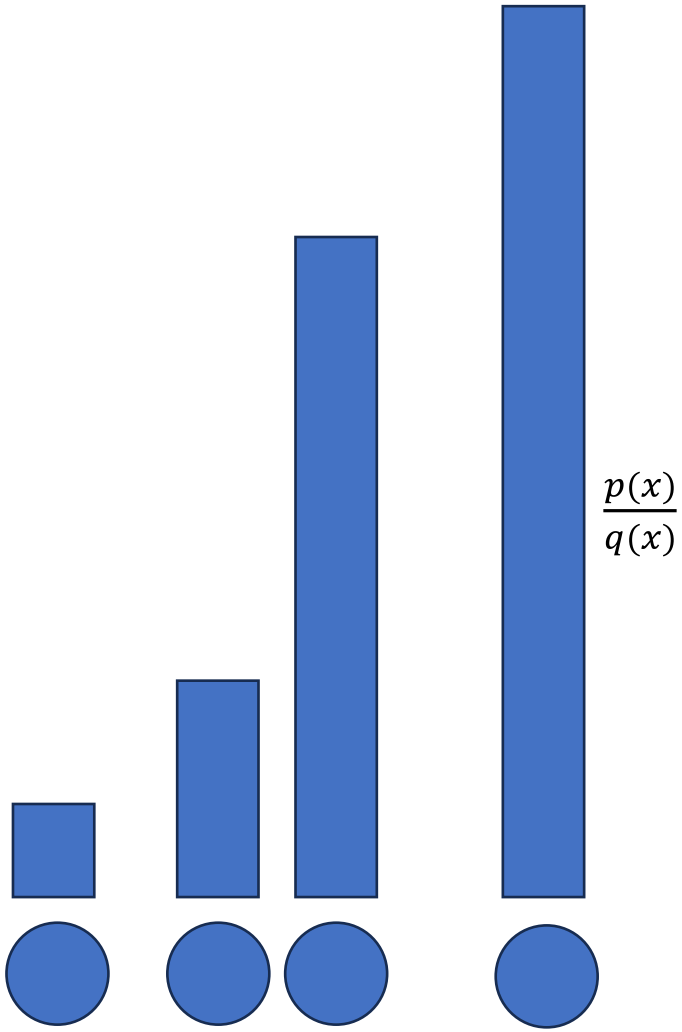


Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

1. Draw $x_1, x_2, \cdots, x_N \sim q$
2. **Calculate $\frac{p(x)}{q(x)}$ for all of the N samples**
3. Draw $i_1, i_2, \cdots, i_M \sim \mathrm{Cat}\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)$
4. Return $x_{i_1}, x_{i_2}, \cdots, x_{i_M}$

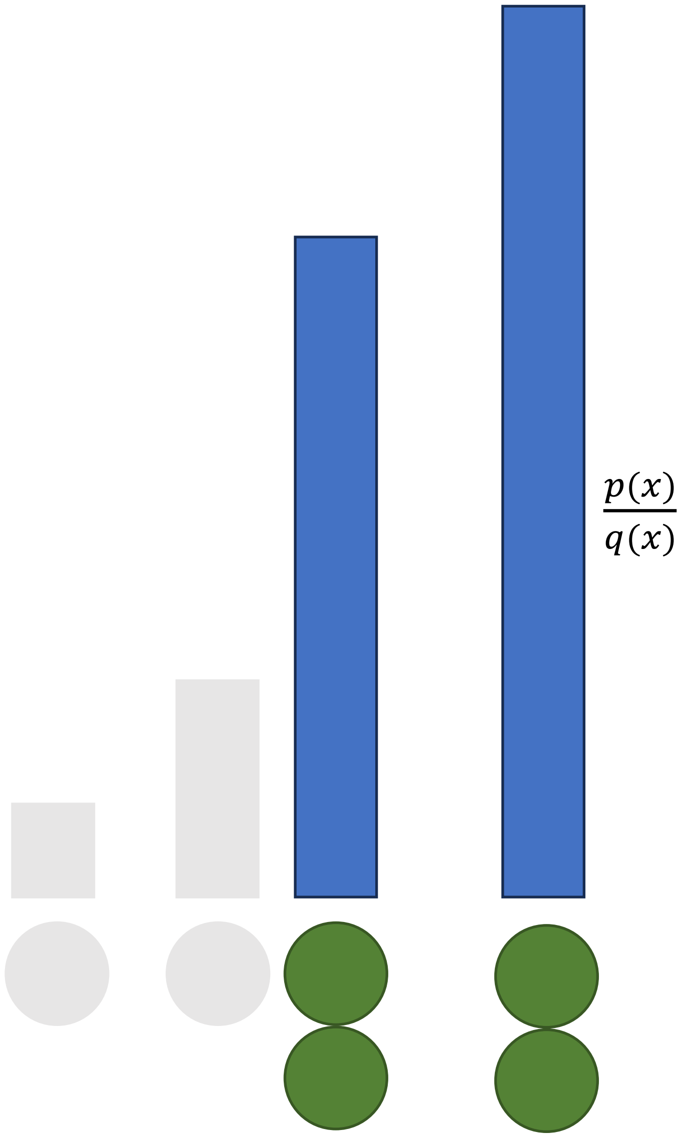


Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ *--- proposal*

But we **want to draw** sample from $p(x)$ *--- target*

1. Draw $x_1, x_2, \cdots, x_N \sim q$
2. Calculate $\frac{p(x)}{q(x)}$ for all of the N samples
3. **Draw $\boldsymbol{i_1, i_2, \cdots, i_M \sim Cat\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)}$**
4. **Return $\boldsymbol{x_{i_1}, x_{i_2}, \cdots, x_{i_M}}$**

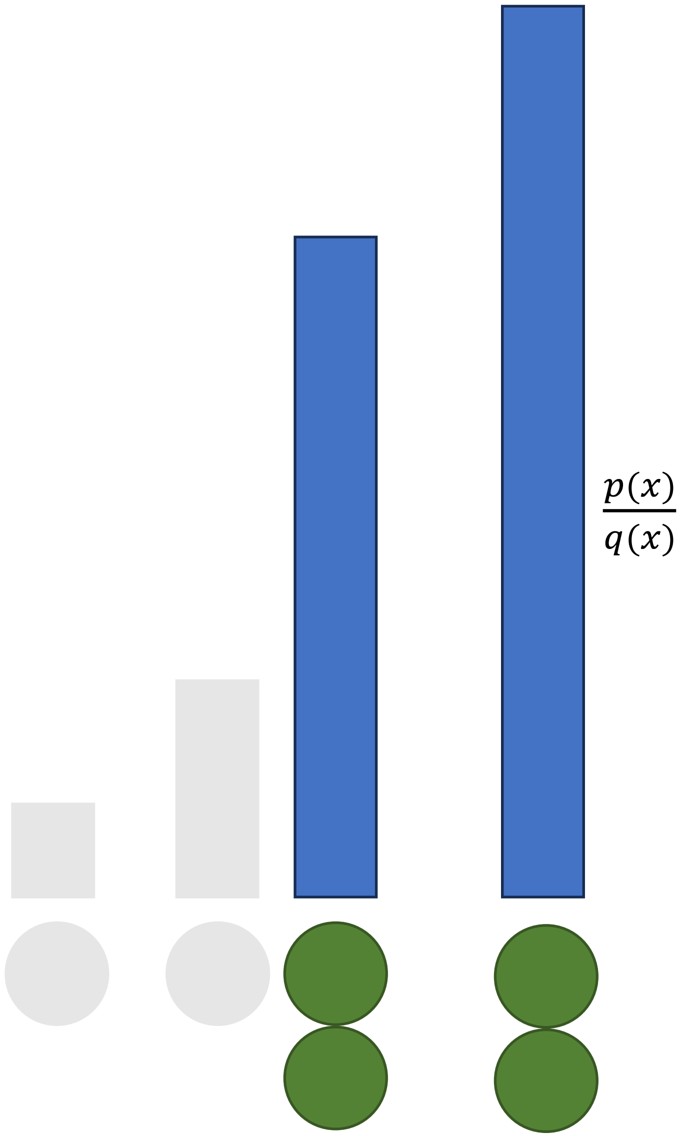


Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ *--- proposal*

But we **want to draw** sample from $p(x)$ *--- target*

1. Draw $x_1, x_2, \cdots, x_N \sim q$
2. Calculate $\frac{p(x)}{q(x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim \mathrm{Cat}\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)$
4. Return $x_{i_1}, x_{i_2}, \cdots, x_{i_M}$



Exact when $N \to \infty$

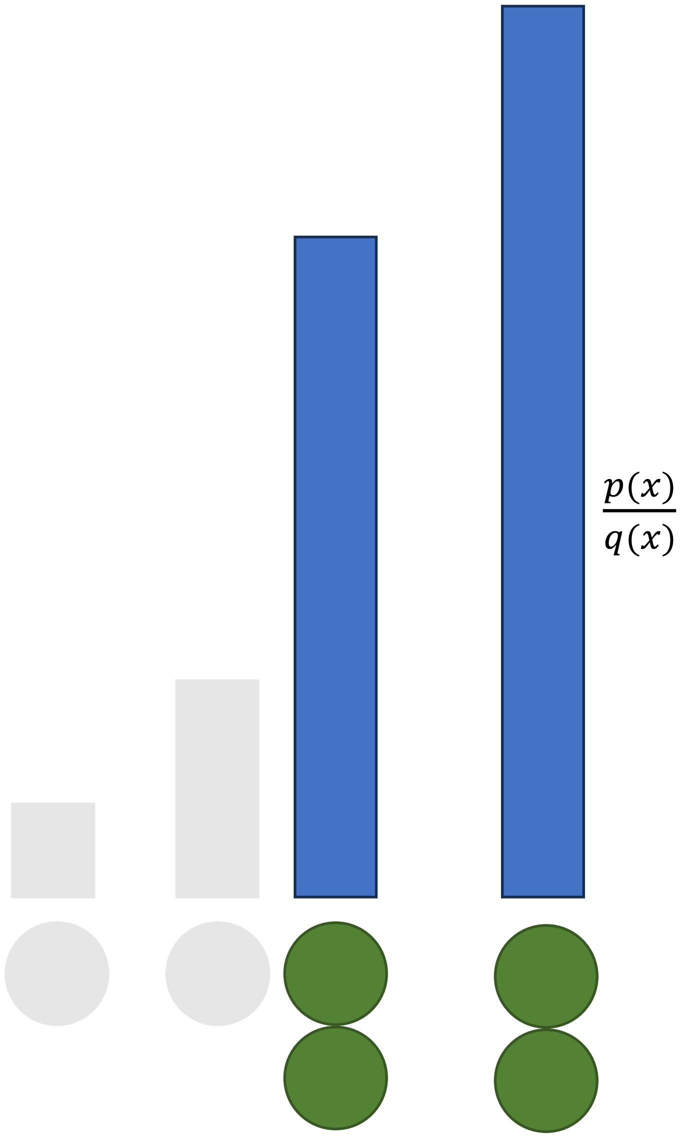
Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ *--- proposal*

But we **want to draw** sample from $p(x)$ *--- target*

1. Draw $x_1, x_2, \cdots, x_N \sim q$
2. Calculate $\frac{p(x)}{q(x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim \mathrm{Cat}\left(\frac{p(x_1)}{q(x_1)}, \frac{p(x_2)}{q(x_2)}, \cdots, \frac{p(x_N)}{q(x_N)}\right)$
4. Return $x_{i_1}, x_{i_2}, \cdots, x_{i_M}$

Importance Re-sampling



Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

Then…

We **are able to draw** sample from $q(y|x)$ --- *proposal*

But we **want to draw** sample from $p(y)$ --- *target*

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x)$ --- *proposal*

But we **want to draw** sample from $p(x)$ --- *target*

Then…

We **are able to draw** sample from $q(y|x)$ --- *proposal*

But we **want to draw** sample from $p(y)$ --- *target*

Apply the previous procedure again!

Importance Sampling and Sequential Monte Carlo

1. Draw $y_1, y_2, \cdots, y_N \sim q(y|x)$
2. Calculate $\frac{p(y)p(x|y)}{p(x)q(y|x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim Cat\left(\frac{p(y_1)p(x_1|y_1)}{p(x_1)q(y_1|x_1)}, \frac{p(y_2)p(x_2|y_2)}{p(x_2)q(y_2|x_2)}, \cdots, \frac{p(y_N)p(x_N|y_N)}{p(x_N)q(y_N|x_N)}\right)$
4. Return $y_{i_1}, y_{i_2}, y_{i_M}$.

We **are able to draw** sample from $q(y|x)$ --- *proposal*

But we **want to draw** sample from $p(y)$ --- *target*

Apply the previous procedure again!

Importance Sampling and Sequential Monte Carlo

1. Draw $y_1, y_2, \cdots, y_N \sim q(y|x)$
2. Calculate $\frac{p(y)\boldsymbol{\color{red}p(x|y)}}{p(x)q(y|x)}$ for all of the N samples
3. Draw $i_1, i_2, \cdots, i_M \sim Cat\left(\frac{p(y_1)p(x_1|y_1)}{p(x_1)q(y_1|x_1)}, \frac{p(y_2)p(x_2|y_2)}{p(x_2)q(y_2|x_2)}, \cdots, \frac{p(y_N)p(x_N|y_N)}{p(x_N)q(y_N|x_N)}\right)$
4. Return $y_{i_1}, y_{i_2}, y_{i_M}$.

We **are able to draw** sample from $q(y|x)$ --- *proposal*

But we **want to draw** sample from $p(y)$ --- *target*

Apply the previous procedure again!

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x_N), q(x_{N-1}|x_N)$, … --- *proposals*

But we **want to draw** sample from $p(x_N), p(x_{N-1})$, … --- *targets*

Apply the previous procedure again and again and again!

Sequential Monte Carlo / Particle Filtering

Modified Diffusion process

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x_N), q(x_{N-1}|x_N), \ldots$ --- *proposals*

But we **want to draw** sample from $p(x_N), p(x_{N-1}), \ldots$ --- *targets*

Apply the previous procedure again and again and again!

Sequential Monte Carlo / Particle Filtering

Importance Sampling and Sequential Monte Carlo

We **are able to draw** sample from $q(x_N), q(x_{N-1}|x_N), \ldots$ *--- proposals*

But we **want to draw** sample from $p(x_N), p(x_{N-1}), \ldots$ *--- targets*

Apply the previous procedure again and ag and again!

Modified Diffusion Marginal

Sequential Monte Carlo / Particle Filtering

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

What does this mean in terms of density functions?

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

What does this mean in terms of density functions?

Diffusion Model p_0

👉 Tempering: $p_0{}' \propto p_0^\alpha$

💡 Molecular simulation

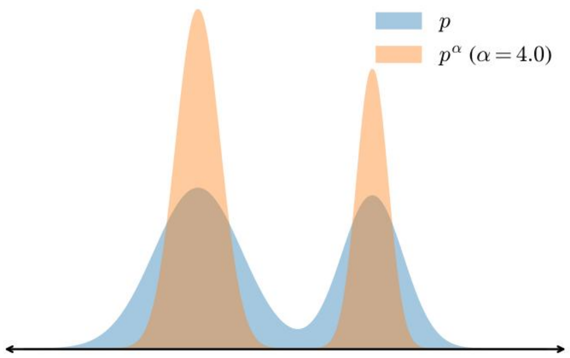


Figure taken from: Karan, Aayush, and Yilun Du. "Reasoning with Sampling: Your Base Model is Smarter Than You Think." *arXiv*.

Generation Control of Diffusion Models

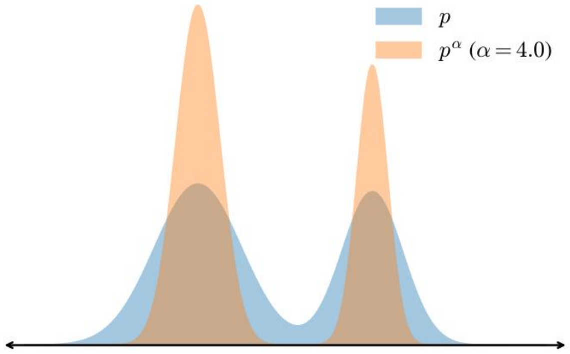
What kind of control we want to impose to our diffusion model?

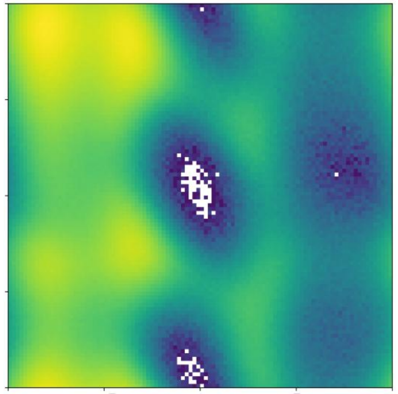
What does this mean in terms of density functions?

Diffusion Model p_0

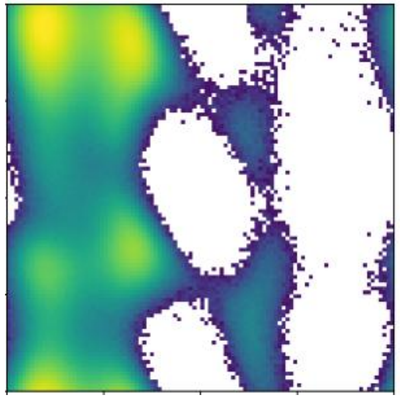
👉 Tempering: $p_0{}' \propto p_0^\alpha$

💡 Molecular simulation





Alanine at 800K



Alanine at 300K

Figure taken from: Karan, Aayush, and Yilun Du. "Reasoning with Sampling: Your Base Model is Smarter Than You Think." *arXiv*.

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

What does this mean in terms of density functions?

Diffusion Model p_0

👉 Tempering: $p_0' \propto p_0^\alpha$

- 💡 Molecular simulation

👉 Tilting: $p_0' \propto p_0 \exp(r_0(x_0))$

- 💡 Inpainting, infilling (motif-scaffolding), reward-tilting, etc

(a) **class:** *balloon*;
Reward prompt: *a blue balloon.*

Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXi*

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

What does this mean in terms of density functions?

Diffusion Model p_0

👉 Tempering: ${p_0}' \propto p_0^{\alpha}$

💡 Molecular simulation

👉 Tilting: ${p_0}' \propto p_0 \exp(r_0(x_0))$

💡 Inpainting, infilling (motif-scaffolding), reward-tilting, etc

👉 Composition: ${p_0}' \propto \left(p_0^{(1)}\right)^{\alpha} \left(p_0^{(2)}\right)^{\beta}$

💡 Stitching / composing model properties e.g., ligand binding to two protein pockets

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

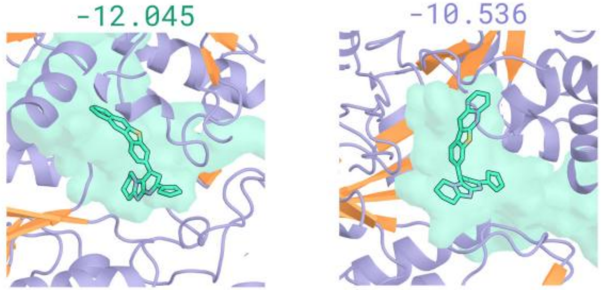
What does this mean in terms of density functions?

	Better than known. (↑)	(P_1 * P_2) (↑)	max(P_1, P_2) (↓)
Sum score	$0.345_{\pm 0.288}$	$65.110_{\pm 17.802}$	$-7.222_{\pm 1.348}$
FKC	$0.608_{\pm 0.390}$	$\mathbf{82.371}_{\pm \mathbf{24.928}}$	$\mathbf{-8.296}_{\pm \mathbf{1.450}}$
RNC ($c_a = 1, c_b = 0.0$)	$0.589_{\pm 0.413}$	$81.186_{\pm 26.158}$	$-8.122_{\pm 1.588}$
RNC ($c_a = 1, c_b = 0.2$)	$\mathbf{0.649}_{\pm \mathbf{0.356}}$	$81.771_{\pm 24.673}$	$-8.112_{\pm 1.660}$

P_1 top-1 (↓)	P_2 top-1 (↓)	Div. (↑)	Val. & Uniq. (↑)	Qual. (↑)
$-9.411_{\pm 1.574}$	$-9.769_{\pm 1.758}$	$0.881_{\pm 0.010}$	$0.927_{\pm 0.147}$	$0.134_{\pm 0.087}$
$-9.437_{\pm 1.733}$	$-10.035_{\pm 1.601}$	$0.814_{\pm 0.043}$	$0.925_{\pm 0.113}$	$0.192_{\pm 0.191}$
$\mathbf{-9.650}_{\pm \mathbf{1.608}}$	$-10.075_{\pm 1.663}$	$0.823_{\pm 0.027}$	$0.942_{\pm 0.069}$	$0.222_{\pm 0.173}$
$-9.585_{\pm 1.885}$	$\mathbf{-10.102}_{\pm \mathbf{1.525}}$	$\mathbf{0.836}_{\pm \mathbf{0.025}}$	$\mathbf{0.950}_{\pm \mathbf{0.066}}$	$\mathbf{0.223}_{\pm \mathbf{0.202}}$

👉 Composition: $p_0{}' \propto \left(p_0^{(1)}\right)^{\alpha} \left(p_0^{(2)}\right)^{\beta}$

💡 Stitching / composing model properties
e.g., ligand binding to two protein pockets



He, Jiajun, et al. "RNE: a plug-and-play framework for diffusion density estimation and inference-time control." *arXiv*.
Skreta, Marta, et al. "Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts." *ICML 2025*.

Generation Control of Diffusion Models

What kind of control we want to impose to our diffusion model?

What does this mean in terms of density functions?

Diffusion Model p_0

👉 Tempering: ${p_0}' \propto p_0^{\alpha}$

💡 Molecular simulation

👉 Tilting: ${p_0}' \propto p_0 \exp(r_0(x_0))$

💡 Inpainting, infilling (motif-scaffolding), reward-tilting, etc

👉 Composition: ${p_0}' \propto \left(p_0^{(1)}\right)^{\alpha} \left(p_0^{(2)}\right)^{\beta}$

💡 Stitching / composing model properties e.g., ligand binding to two protein pockets

Sequential Monte Carlo Weight Calculation

Diffusion model generates $p_T(x_T), p_{T-1}(x_{T-1}), \dots, p_0(x_0)$
with denoising kernels $p(x_{T-1}|x_T)$, $p(x_{T-2}|x_{T-1}), \dots, p(x_0|x_1)$

We want to generate from $p_T{}'(x_T), p_{T-1}{}'(x_{T-1}), \dots, p_0{}'(x_0)$
with proposal kernels $p'(x_{T-1}|x_T)$, $p'(x_{T-2}|x_{T-1}), \dots, p'(x_0|x_1)$

$$\frac{p'(x_{t-1})p(x_t|x_{t-1})}{p'(x_t)p'(x_{t-1}|x_t)}$$

More Math?

Forward-backward RND

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

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

For simplicity, we hereafter call

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

Forward-backward RND

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

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

For simplicity, we hereafter call

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

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

"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**

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

“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**

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

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;

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

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

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

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

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

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

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

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

- 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

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

- Choose a heuristic guidance process;

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

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

"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**

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

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

“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**

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

?

Example: Diffusion Inference-time Steering with Path RND

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

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

- Choose a heuristic guidance process;

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

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

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

- 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**

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

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

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

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

?

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

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

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

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

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

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

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

?

Example: Diffusion Inference-time Steering with Path RND

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

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

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

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

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

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

- 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 $\{x\} \sim q_\tau$

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

- Choose a heuristic guidance process;

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

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

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

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

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

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

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 $\left(p_t^{(1)}\right)^\beta \left(p_t^{(2)}\right)^\alpha$

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

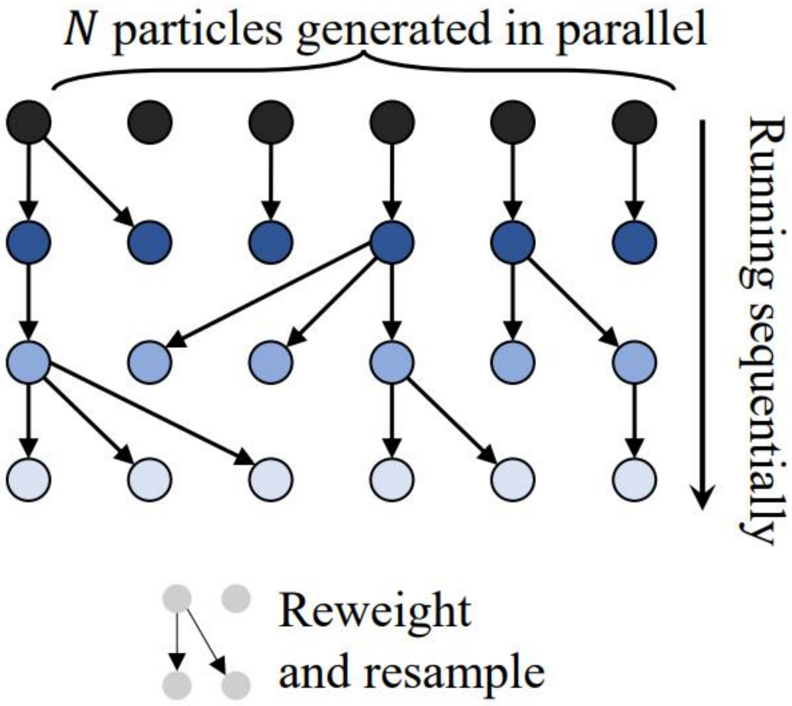
Episode 2

Control Generation with Sequential Monte Carlo in Path Space

💡 **Sequence of Importance Resampling (SMC)** along the denoising path

💡 Flexible Control of diffusion generation process

Sequel Episode: Curse of Diversity



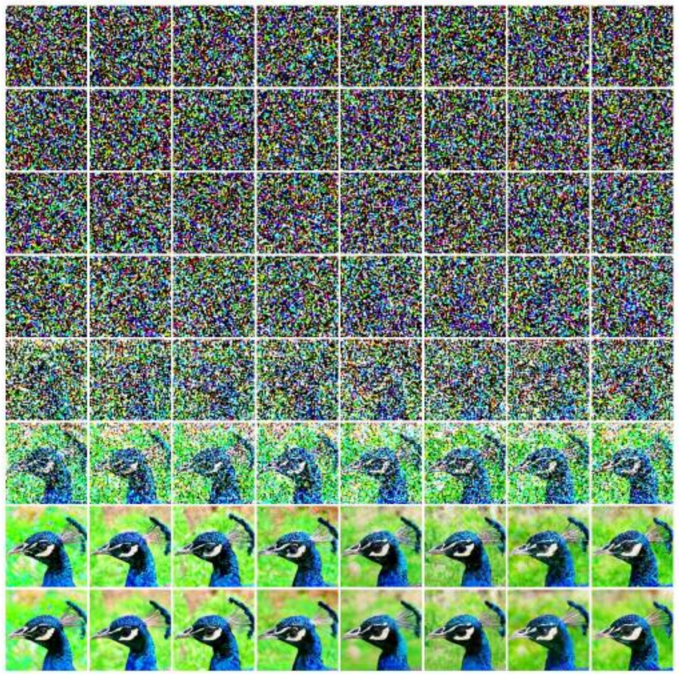


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Sequel Episode: Curse of Diversity

?

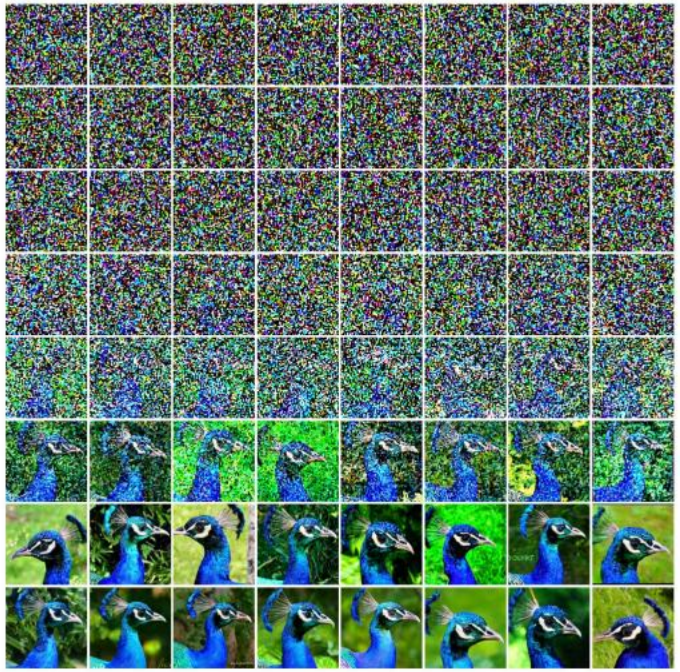


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Sequel Episode: Curse of Diversity

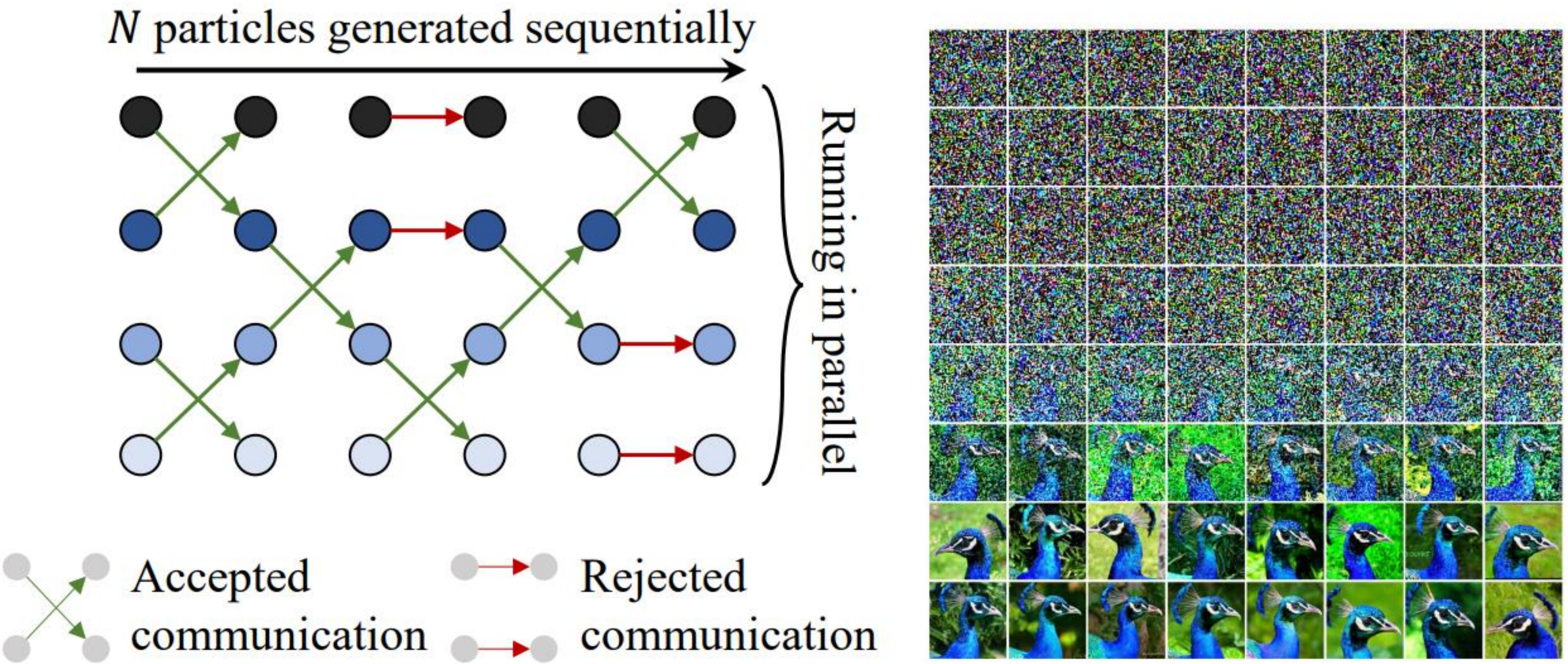


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Replica Exchange: Intuition

Sequential Monte Carlo:

Generate N samples at each step, select the “best” set, go to next step

Replica Exchange (parallel Tempering):

Generate initial guess at all steps,

attempt to exchange guesses at adjacent steps,

accept exchange if the change makes the guess “better”,

otherwise reject

Sequel Episode: Curse of Diversity

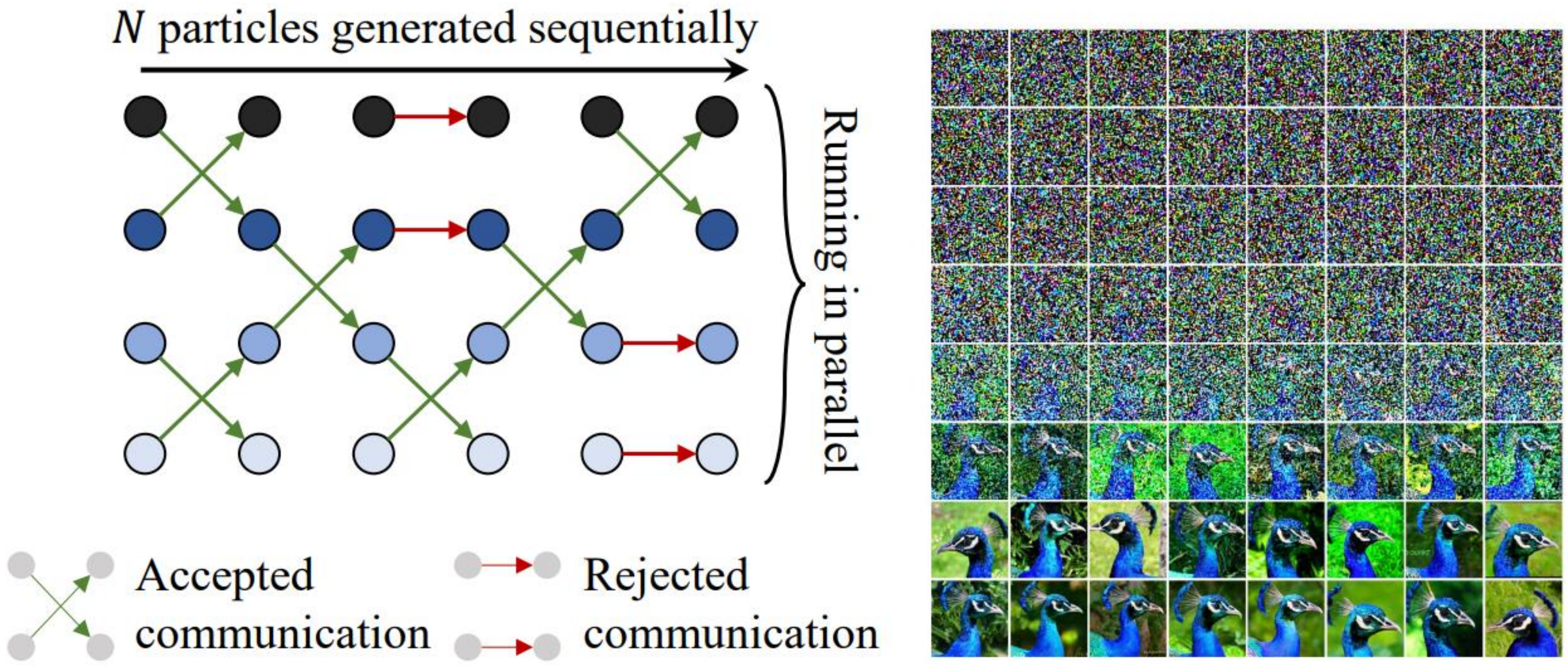


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Sequel Episode: Curse of Diversity

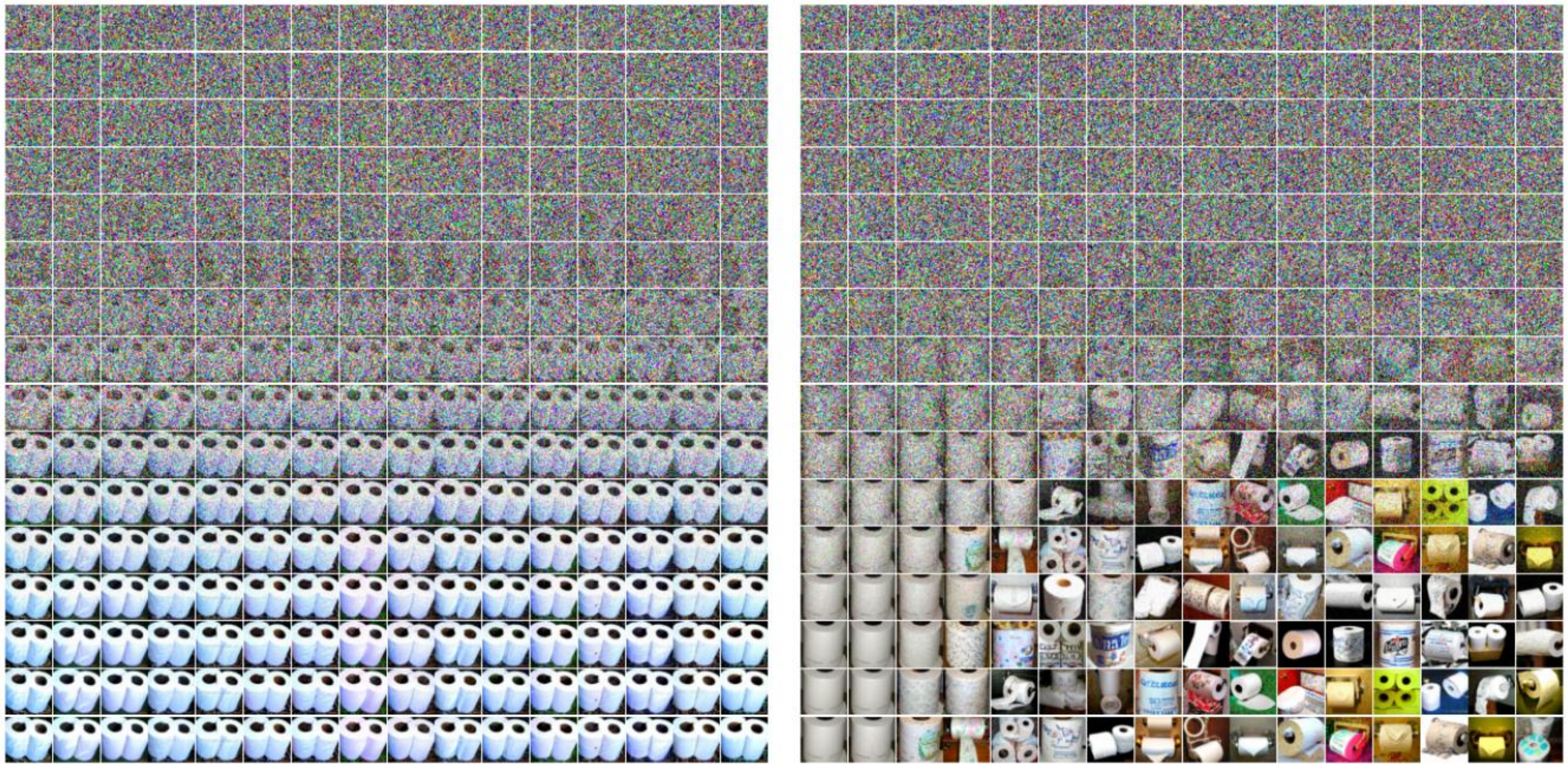


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Sequel Episode: Curse of Diversity

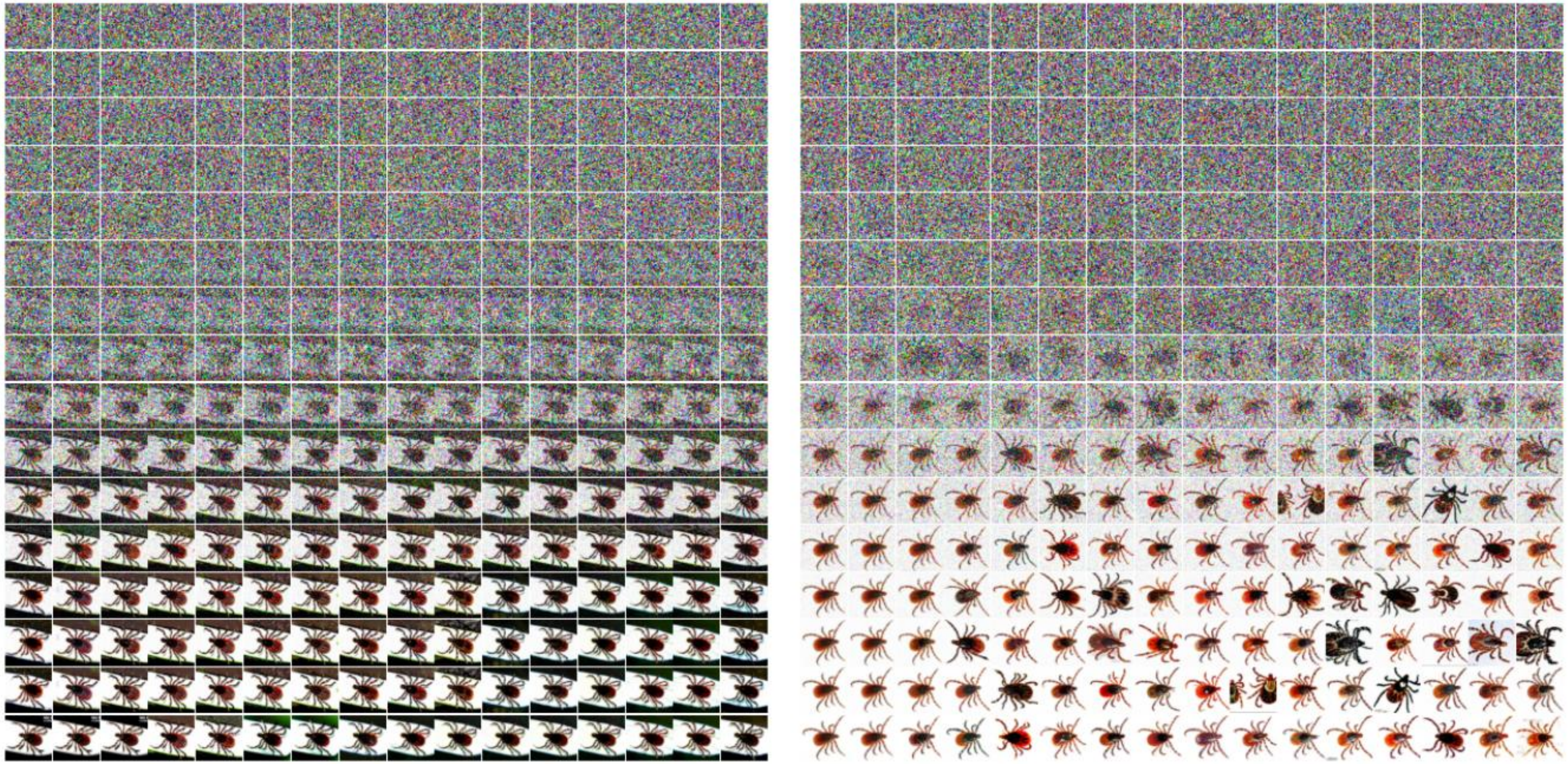
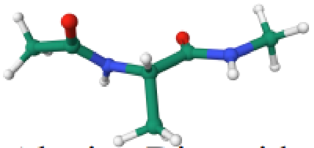


Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

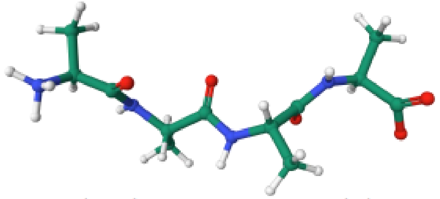
Control Diffusion with Replica Exchange

👉 Tempering: $p_0{}' \propto p_0^{\alpha}$

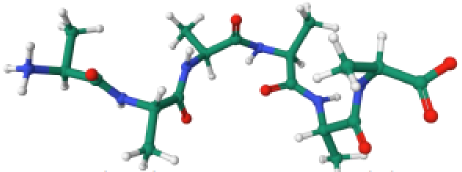
Table 1: Inference-time tempering performance for Alanine Dipeptide, Tetrapeptide and Hexapeptide.



Alanine Dipeptide



Alanine Tetrapeptide



Alanine Hexapeptide

		FKC		**RNE**	**CREPE** (Ours)
		Anneal Score	Anneal Noise		
ALA Dipeptide (800K → 300K)	Energy TVD	0.345 $\pm$ 0.010	0.894 $\pm$ 0.002	0.391 $\pm$ 0.006	**0.224** $\pm$ 0.005
	Distance TVD	0.023 $\pm$ 0.001	0.036 $\pm$ 0.001	0.024 $\pm$ 0.001	**0.019** $\pm$ 0.000
	Sample W2	0.293 $\pm$ 0.001	0.282 $\pm$ 0.001	0.282 $\pm$ 0.001	**0.264** $\pm$ 0.001
	TICA MMD	0.116 $\pm$ 0.003	0.108 $\pm$ 0.004	0.168 $\pm$ 0.007	**0.096** $\pm$ 0.014
ALA Tetrapeptide (800K → 500K)	Energy TVD	**0.122** $\pm$ 0.012	0.436 $\pm$ 0.007	0.154 $\pm$ 0.006	**0.122** $\pm$ 0.004
	Distance TVD	**0.014** $\pm$ 0.000	0.015 $\pm$ 0.000	**0.013** $\pm$ 0.001	**0.013** $\pm$ 0.001
	Sample W2	0.923 $\pm$ 0.008	0.892 $\pm$ 0.001	0.893 $\pm$ 0.005	**0.856** $\pm$ 0.004
	TICA MMD	0.183 $\pm$ 0.020	0.138 $\pm$ 0.017	0.155 $\pm$ 0.009	**0.035** $\pm$ 0.002
ALA Hexapeptide (800K → 600K)	Energy TVD	**0.091** $\pm$ 0.006	0.206 $\pm$ 0.005	**0.087** $\pm$ 0.003	0.398 $\pm$ 0.001
	Distance TVD	0.018 $\pm$ 0.000	0.020 $\pm$ 0.001	**0.010** $\pm$ 0.001	**0.009** $\pm$ 0.001
	Sample W2	1.585 $\pm$ 0.001	1.652 $\pm$ 0.012	1.618 $\pm$ 0.001	**1.299** $\pm$ 0.004
	TICA MMD	0.088 $\pm$ 0.004	0.068 $\pm$ 0.010	0.042 $\pm$ 0.004	**0.009** $\pm$ 0.001

Control Diffusion with Replica Exchange

👉 reward-tilting: $p_0{}' \propto p_0 \exp(r_0)$

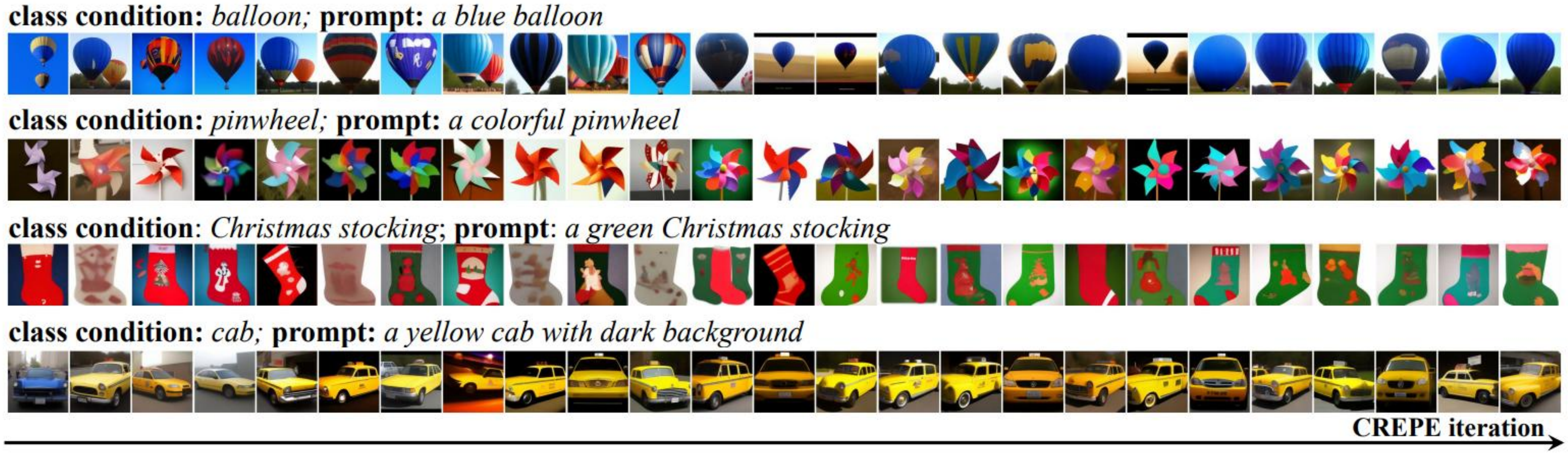
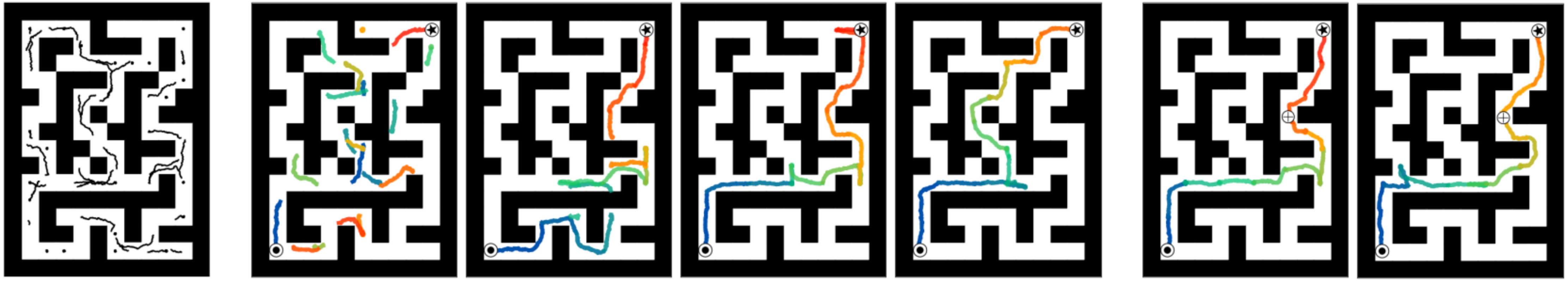


Figure 1: Trajectory of images generated using CREPE for prompted reward-tilting on ImageNet-512, thinned every 8 iterations. After burn-in, the samples align closely with the prompt.

Control Diffusion with Replica Exchange

👉 Composition + reward-tilting: $p_0{}' \propto \prod p_0^{(i)} \ \exp(r_0)$



Example of training trajectories. | Trajectory after 1 PT iteration. | Trajectory after 10k PT iterations. | Trajectory after 50k PT iterations. | Trajectory after 100k PT iterations. | Trajectory after 101k PT iteration. | Trajectory after 150k PT iterations.

Figure taken from: He, Jiajun, et al. "CREPE: Controlling Diffusion with Replica Exchange." *arXiv.*

Summary

- Diffusion Model
- Path Measure
- Importance Sampling and SMC / Replica Exchange with Path Measures
- Control your Diffusion Model

What’s next?