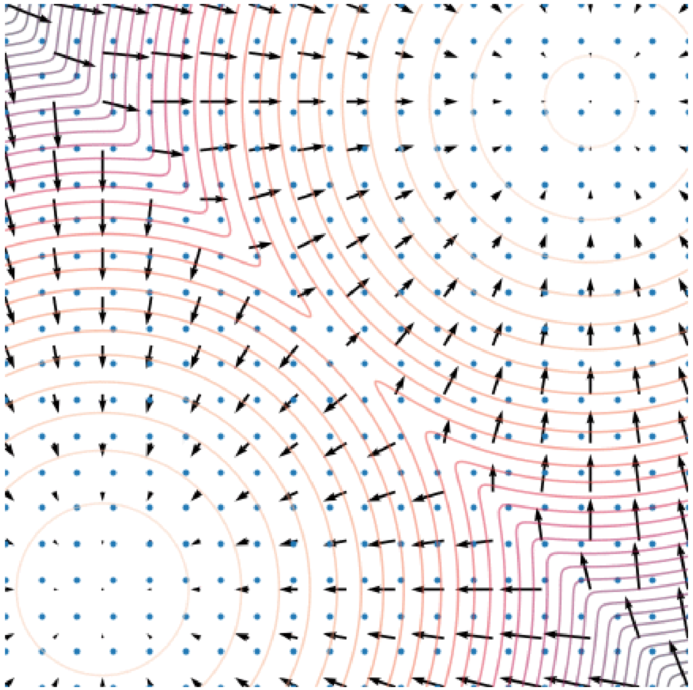
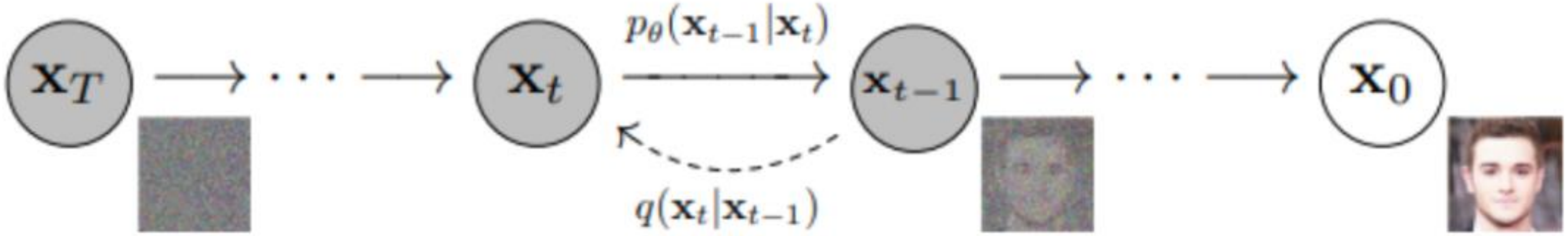
深入理解扩散模型之 From DDPM to Score and Flow Matching



Score

Review

DDPM (2020)



1. Forward:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \qquad 0 < \beta_1 < \beta_2 < ... < \beta_T < 1$$

2. Reverse：

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$$

$$\Sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2\mathbf{I} \qquad \sigma_t^2 = \beta_t$$

Score（2021, 2019）

$$p(x) = \frac{1}{Z}\exp\left(-\frac{E(x)}{kT}\right)$$

Boltzmann

Target

$$p_\theta(\mathbf{x})$$

Define

$$p_\theta(\mathbf{x}) = \frac{e^{-f_\theta(\mathbf{x})}}{Z_\theta}$$

$$Z_\theta = \int e^{-f_\theta(x)}dx > 0$$

Such that

$$\int p_\theta(\mathbf{x})\mathrm{d}\mathbf{x} = 1$$

$$\boldsymbol{p_\theta > 0}$$

To get θ^* , introduce

$$L(\theta) = \prod_{i=1}^{N} p_\theta(x_i)$$

$$\ell(\theta) = \sum_{i=1}^{n} \log p_\theta(x_i)$$

$$\boxed{\theta^* = \arg\max_\theta \sum_{i=1}^{n} \log p_\theta(x_i)}$$

$$p_\theta(\mathbf{x}) = \frac{e^{-f_\theta(\mathbf{x})}}{Z_\theta}$$

$$Z_\theta = \int e^{-f_\theta(x)} dx > 0$$

$$\theta^* = \arg\max_\theta \sum_{i=1}^{n} \log p_\theta(x_i)$$

Write out

$$\log p_\theta(\mathbf{x}) = -f_\theta(\mathbf{x}) - \log Z_\theta$$

Try

$$\nabla_\theta \log p_\theta(x) = -\nabla_\theta f_\theta(x) + \mathbb{E}_{x' \sim p_\theta}[\nabla_\theta f_\theta(x')]$$

It's discovered that

$$\nabla_x \log p_\theta(x) = \nabla_x[-f_\theta(x) - \log Z_\theta] = -\nabla_x f_\theta(x)$$

Design

$$\mathbf{s}_\theta(\mathbf{x})$$

and optimize

$$\mathbb{E}_{p(\mathbf{x})}[||\nabla_\mathbf{x} \log p(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x})||_2^2]$$

Score-Matching(2005)

$$\mathcal{L} = \mathbb{E}_{p(\mathbf{x})}\left[\mathrm{tr}(\nabla_x \mathbf{s}_\theta(\mathbf{x})) + \frac{1}{2}\|\mathbf{s}_\theta(\mathbf{x})\|^2\right] + \mathrm{const}$$

$$s_\theta(x) \approx \nabla_x \log p_\theta(x)$$

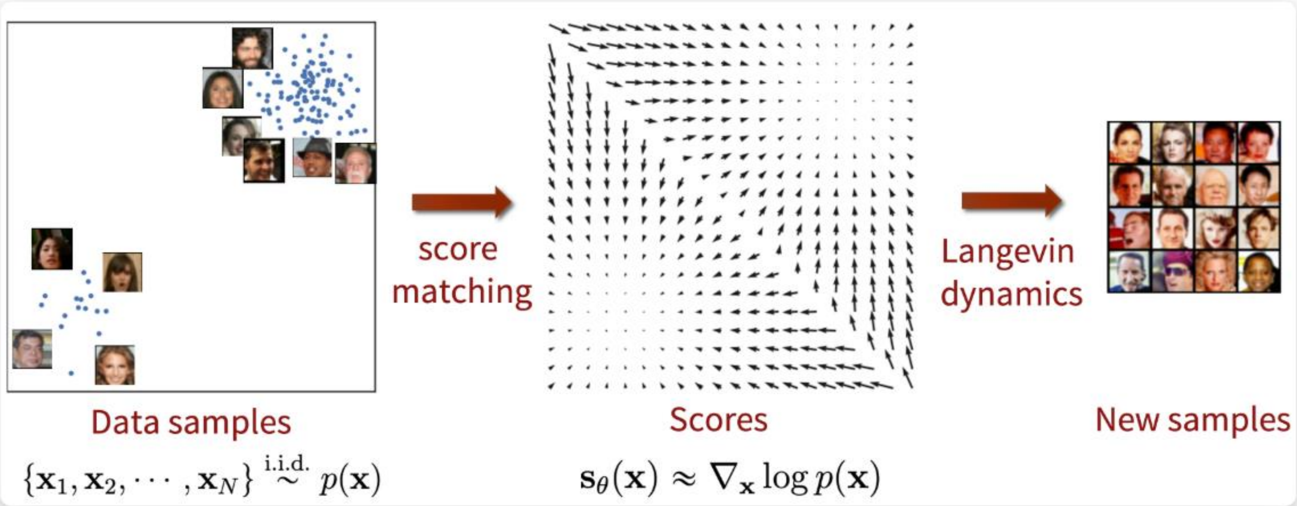
After we get

$$\mathbf{s}_\theta(\mathbf{x})$$

Langevin Dynamics:

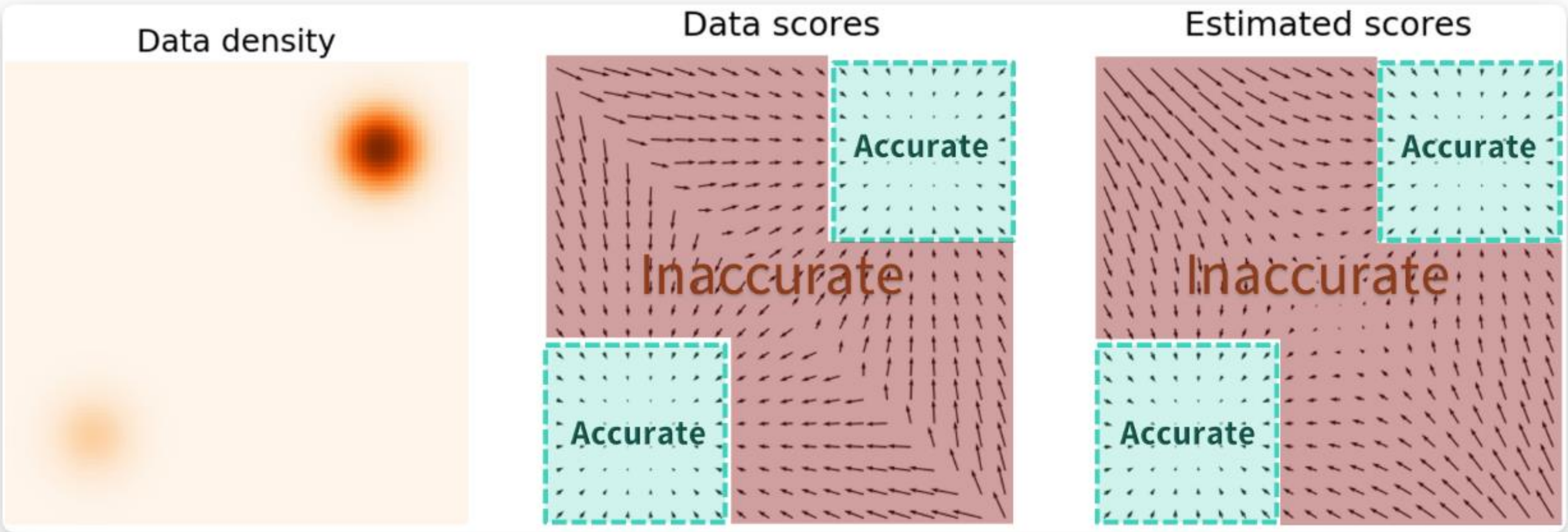
$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \epsilon \nabla_\mathbf{x} \log p(\mathbf{x}) + \sqrt{2\epsilon}\, \mathbf{z}_i, \quad i = 0, 1, \cdots, K$$

$$\mathbf{x}_0, \mathbf{z}_i \sim \mathcal{N}(0, I)$$



But Performance suffers.

$$\mathbb{E}_{p(\mathbf{x})}[||\nabla_\mathbf{x} \log p(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x})||_2^2] = \int p(\mathbf{x})||\nabla_\mathbf{x} \log p(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x})||_2^2 \mathrm{d}\mathbf{x}$$



Estimated scores are only accurate in high density regions.

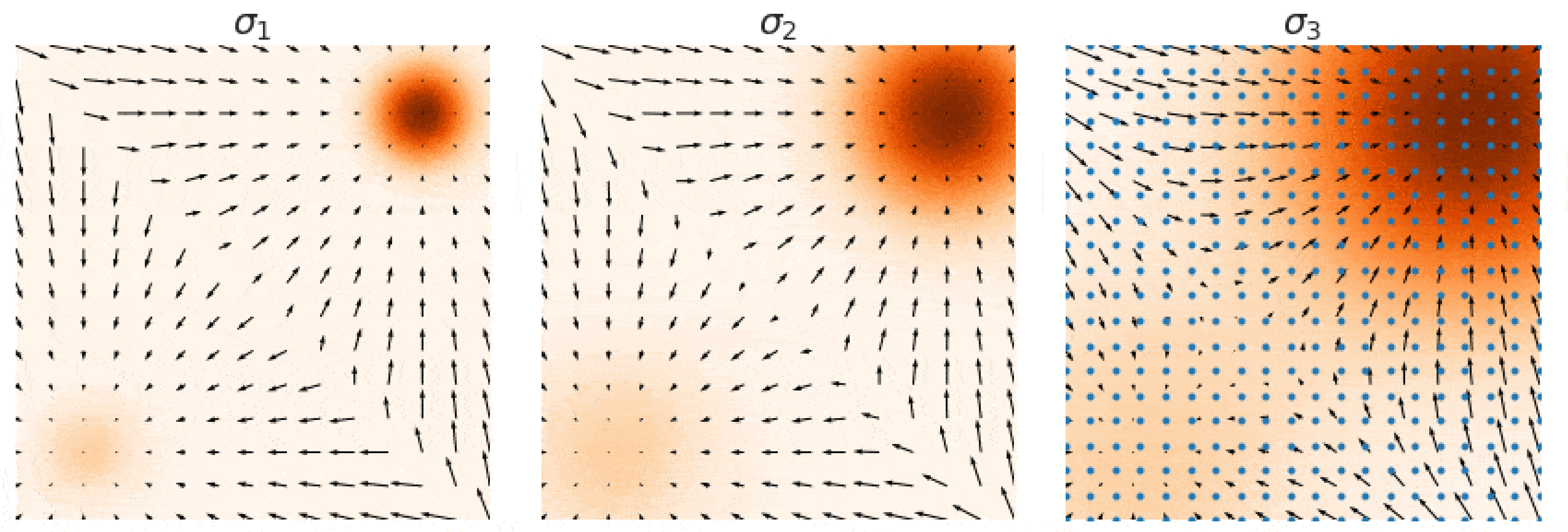
Multi-scale perturbation

$$p_{\sigma_i}(\mathbf{x}) = \int p(\mathbf{y})\mathcal{N}(\mathbf{x};\mathbf{y},\sigma_i^2 I)\mathrm{d}\mathbf{y} \qquad \sigma_1 < \sigma_2 < \cdots < \sigma_L \qquad \mathbf{y} \sim p_{\mathrm{data}}(\mathbf{y})$$

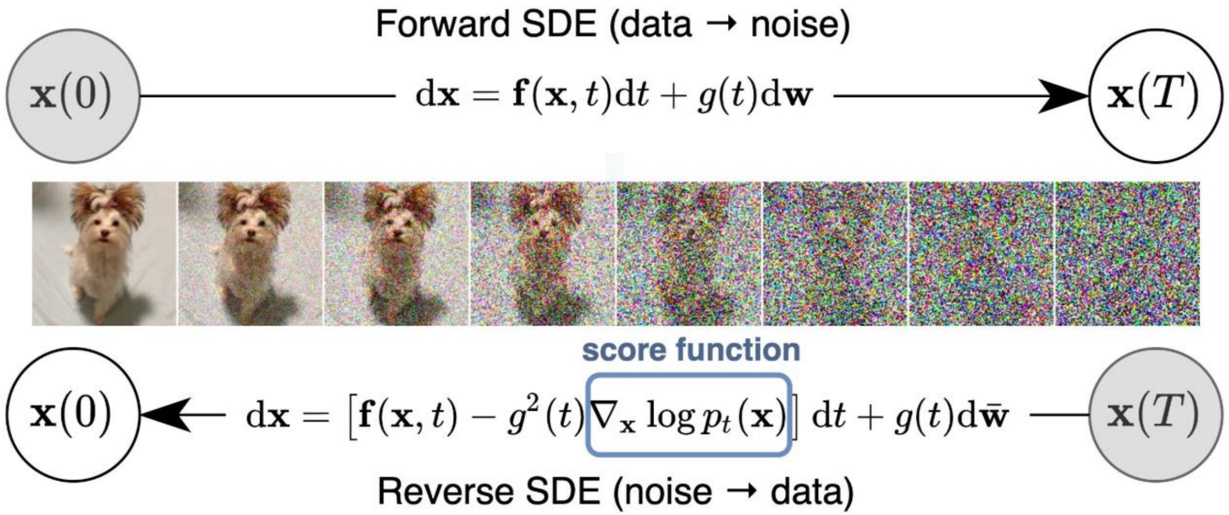
Optimize

$$\sum_{i=1}^{L} \lambda(i)\mathbb{E}_{p_{\sigma_i}(\mathbf{x})}[\|\nabla_{\mathbf{x}} \log p_{\sigma_i}(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x}, i)\|_2^2]$$

Annealed Langevin Dynamics



When L goes infinite, the Langevin process can be described as a SDE



optimize

$$\mathbb{E}_{t\in\mathcal{U}(0,T)}\mathbb{E}_{p_t(\mathbf{x})}[\lambda(t)\|\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x}, t)\|_2^2]$$

sample

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g^2(t)\mathbf{s}_\theta(\mathbf{x}, t)]\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

probability flow ODE

$$\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right]\mathrm{d}t$$

Flow Matching

Mapping a random q0 to data distribution q1

$$x \sim q_0$$
$$y = \phi(x)$$

How to calculate p1(y) using q0(x)?

$$p_1(y) = q_0(\phi^{-1}(y)) \left| \det \left[\frac{\partial \phi^{-1}}{\partial y}(y) \right] \right|$$
$$= \frac{q_0(x)}{\left| \det \left[\frac{\partial \phi}{\partial x}(x) \right] \right|} \quad \text{with } x = \phi^{-1}(y)$$

Aiming at

$$\text{argmax}_\theta \ \mathbb{E}_{x \sim \mathcal{D}}[\log p_1(x)]$$

$$\log p_1^\theta(x) = \log p_0(\phi_\theta^{-1}(x)) + \log |\det J_{\phi_\theta^{-1}}(x)|$$

$$\log p_1^{\theta}(x) = \log p_0(\phi_\theta^{-1}(x)) + \log|\det J_{\phi_\theta^{-1}}(x)|$$

Introduce Residual Flow

$$\phi_k(x) = x + \delta\, u_k(x) \qquad (1)$$

Compose such flows

$$\phi = \phi_K \circ \ldots \circ \phi_2 \circ \phi_1$$

$$K \to \infty \qquad\qquad \delta = 1/K$$

Write (1) as another form

$$\frac{\phi(x) - x}{\delta} = u(x)$$

Derivation?

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \lim_{\delta\to 0} \frac{x_{t+\delta} - x_t}{\delta} = \frac{\phi_t(x_t) - x_t}{\delta} = u_t(x_t)$$

Such that

$$x_t \triangleq \phi_t(x_0) = x_0 + \int_0^t u_s(x_s)\,\mathrm{d}s$$

$$\log p_\theta(x) \triangleq \log p_1(x) = \log p_0(x_0) - \int_0^1 (\nabla \cdot u_\theta)(x_t)\mathrm{d}t$$

$$x_t \triangleq \phi_t(x_0) = x_0 + \int_0^t u_s(x_s)\,\mathrm{d}s$$

To Optimize?

$$\mathcal{L}(\theta) = \mathbb{E}_{x\sim q_1}[\log p_1(x)]$$

That's suboptimal. Flow Matching!

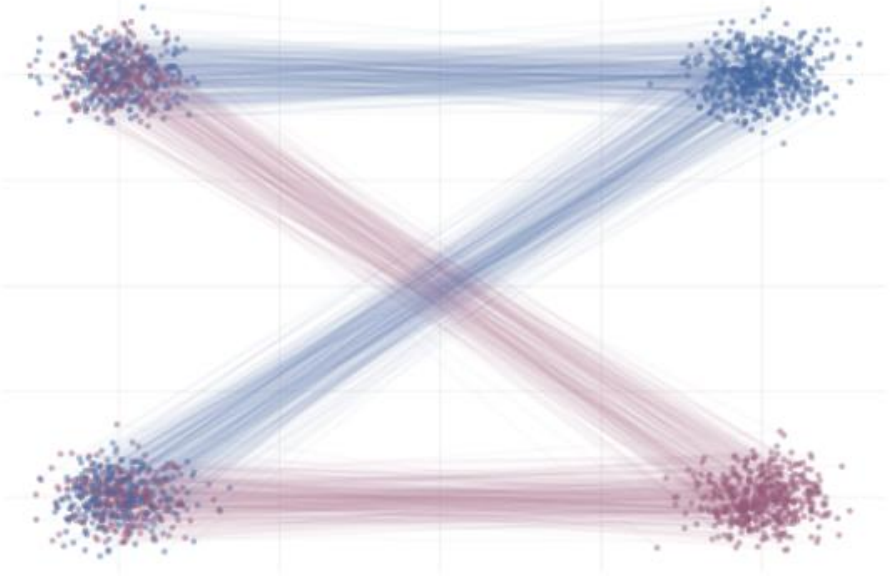
$$\mathcal{L}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1]}\mathbb{E}_{x\sim p_t}\left[\|u_\theta(t,x) - u(t,x)\|^2\right]$$

But we don't know u (t, x). Conditional Flow Matching!

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1],x_1\sim q,x_t\sim p_t(x|x_1)}\left[\|u_\theta(t,x) - u_t(x \mid x_1)\|^2\right]$$

But is CFM really all rainbows and unicorns?

crossing conditional paths!



Coupling

$$p_t(x_t) = \int p_t(x_t \mid z) q(z) \, \mathrm{d}z = \int p_t(x_t \mid x_1, x_0) q(x_1, x_0) \, \mathrm{d}x_1 \, \mathrm{d}x_0$$

Optimal Transport Coupling

$$q(x_1, x_0) = \pi(x_1, x_0) \in \arg \inf_{\pi \in \Pi} \int \|x_1 - x_0\|_2^2 \mathrm{d}\pi(x_1, x_0)$$

Mini-Batch!

Thanks for your Listening!