4/6/25 From CFM to Diffusion

- Comment about CFM: not limited to Gaussian paths! Any path from

$$\delta(x - x_1) \quad \text{to} \quad \mathcal{N}(0,1)$$
$$p_1(x|x_1) \qquad\qquad p_0(x|x_0)$$

will work to generate $x_1 \sim p_{data}$.

- $$\begin{cases} p_t(x) = \int dz\, p_t(x|z)\, q(z) \\ u_t(x) = \dfrac{\int dz\, p_t(x|z)\, u_t(x|z)\, q(z)}{p_t(x)} \end{cases}$$

Generalizes to any $q(z)$ w/ correct bdy cond.

Tong et al
2302.00482

Ex: $q(z) = \pi(z_0, z_1)$

$$\pi(z_0) = q_0(z_0) \quad \text{base}$$
$$\pi(z_1) = q_1(z_1) \quad \text{target}$$

$$p_0(x|z_0, z_1) = \delta(x - z_0) \quad p_1(x|z_0, z_1) = \delta(x - z_1)$$

Can FM from any two cts. distributions
(just need samples)

$\pi(z_0, z_1) = \rho_0(z_0)\, \rho_1(z_1)$ reduces to prev.

$\pi(z_0, z_1) =$ OT sol'n $\rightarrow$ can try to learn

Diffusion Generative Models ~2020-2021

Consider Markov noise process

$$dx_t = f(t)\, x_t\, dt + g(t)\, dw$$

drift

"Stochastic diff eq"

Wiener process

Brownian motion

diffusion

means: in time dt

x_t shifted by Gaussian random variable

w/ mean $f(t) x_t\, dt$

& variance $\sigma^2 = g(t)^2 dt$

$dw \sim \varepsilon \sqrt{dt}$ $\quad \varepsilon \sim N(0,1)$

first time GANs surpassed in image quality, mode collapse, ...

Sol'n to this SDE is

$$x_t \sim \mathcal{N}(\gamma_t x_0, \sigma_t) \quad \text{w/} \begin{cases} \dfrac{\dot{\gamma}_t}{\gamma_t} = f(t) \\ \sqrt{2\sigma_t^2\left(\dfrac{\dot{\sigma}_t}{\sigma_t} - \dfrac{\dot{\gamma}_t}{\gamma_t}\right)} = g(t) \end{cases}$$

Deriv (SKIP)

$$x_t \sim \mathcal{N}(\gamma_t x_0, \sigma_t)$$

$$dx_t \sim \mathcal{N}(f(t) x_t \, dt, \; g(t)\sqrt{dt})$$

$$p(x_{t+dt}) = \int dx_t \, p(x_t) \, d(dx_t) \, p(dx_t) \, \delta(x_{t+dt} - x_t - dx_t)$$

$$\Rightarrow \mathcal{N}(\gamma_{t+dt} x_0, \sigma_{t+dt})$$

→ derive diffeq relating $\dot{\gamma}_t$ & $\dot{\sigma}_t$ to f & g.

Same gaussian prob path from $\mathcal{N}(0,1)$ to $\delta(x - x_0)$ as in FM! — but use SDE instead of ODE

Yang Song, Ermon & others realized:

can also reverse noise → data

"reverse diffusion" using "Langevin eqn" (from theory of SDEs & Brownian motion)

$$\boxed{dx_t = \left(f(t)\, x_t - g(t)^2 \nabla_{x_t} \log p_t(x_t) \right) dt + g(t)\, dw}$$

Solve SDE in reverse,

start w/ noise dist'n, end up w/ data dist'n!

$\nabla_{x_t} \log p_t(x_t)$ "score fn"

Note: learning $\nabla_x \log p$ should be easier than learning p → don't need normalizati

$p = e^{-E}/Z$ "energy based models"

Trick to learn $\nabla_x \log p_t$:

"conditional score matching"

(just like CFM)

• learn instead $\nabla_{x_t} \log \underbrace{p_t(x_t|x_1)}_{}$

$$\mathcal{N}(x_t \mid \gamma_t x_1, \sigma_t)$$

$$\sim \frac{x_t - \gamma_t x_1}{\sigma_t^2}$$

• Claim: $$L_{SM}(\theta) = \mathbb{E}_{\substack{t \sim U[0,1] \\ x_t \sim p_t(x_t)}} \| s_\theta(x_t, t) - \nabla_{x_t} \log p_t(x_t) \|^2$$

$$= L_{CSM}(\theta) + \text{const}$$

$$\mathbb{E}_{\substack{t \sim U[0,1] \\ x_t \sim p_t(x_t|x_1) \\ x_1 \sim p_{data}}} \| s_\theta(x_t, t) - \nabla_{x_t} \log p_t(x_t|x_1) \|^2$$

Pf: $\int dt \int dx_t \, p_t(x_t|x_1) \, dx_1 \, p_{data}(x_1)$

$$\left(s_\theta^2 - 2 s_\theta \frac{\nabla_{x_t} p_t(x_t|x_1)}{p_t(x_t|x_1)} + \text{const} \right)$$

integrate over x_1

$$p_t(x_t) s_\theta^2 - 2 s_\theta \nabla_{x_t} p_t(x_t))$$

---

as in CFM, get extremely simple objective for learning score!

s_θ can be any NN, not restricted

Very reminiscent of CFM!

No coincidence — vector field & score closely related

In fact can show:

$$\boxed{u_t(x_t) = f(t)x_t - \frac{1}{2}g(t)^2 s_t(x_t)}$$

valid only for Gaussian prob path

- score-based diffusion is special case of CFM!

$$\frac{\dot\sigma_t}{\sigma_t}x_t + \frac{(\dot\gamma_t\sigma_t - \dot\sigma_t\gamma_t)}{\sigma_t}x_1$$

Pf (can skip): $\dot\gamma_t x_1 + \dot\sigma_t \varepsilon$

$$u_t(x_t) = \frac{\int dx_1\, u_t(x_t|x_1)\, p_t(x_t|x_1)\, p_{data}(x_1)}{p_t(x_t)}$$

$$e^{-\frac{(x_t-\gamma_t x_1)^2}{2\sigma_t^2}}$$

$$= (\ )x_t + \frac{\int dx_1\, (\ )\, \underbrace{x_1\, e^{(\cdots)}}_{\sim \nabla_{x_t} p_t(x_t|x_1)}\, p_{data}(x_1)}{p_t(x_t)}$$

$$= (\ )x_t + (\ )\nabla_{x_t}\log p_t(x_t).$$

So score-matching & flow-matching

$$\| u_\theta - u_t(x_t|x_1) \|^2 \quad \text{vs}$$

$$\| s_\theta - \nabla_{x_t} \log p_t(x_t|x_1) \|^2$$

just two equiv losses!

- Diffusion sampling vs CFM sampling
  (SDE) (ODE)

  just another choice!

All part of a common framework

$\sim$ CNF.