3/6 (Thur)

$z \in \mathbb{R}^d$

How good an image is ~= How likely

Flow — Diffusion.

$X : [0,1] \to \mathbb{R}^d \ , \ t \to X_t$; Trajectory

$u : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$; Vector field.

$(x,t) \to u_t(x)$

$X_0 = x_0 \qquad \frac{d}{dt} X_t = u_t(X_t)$

Flow : ψ :

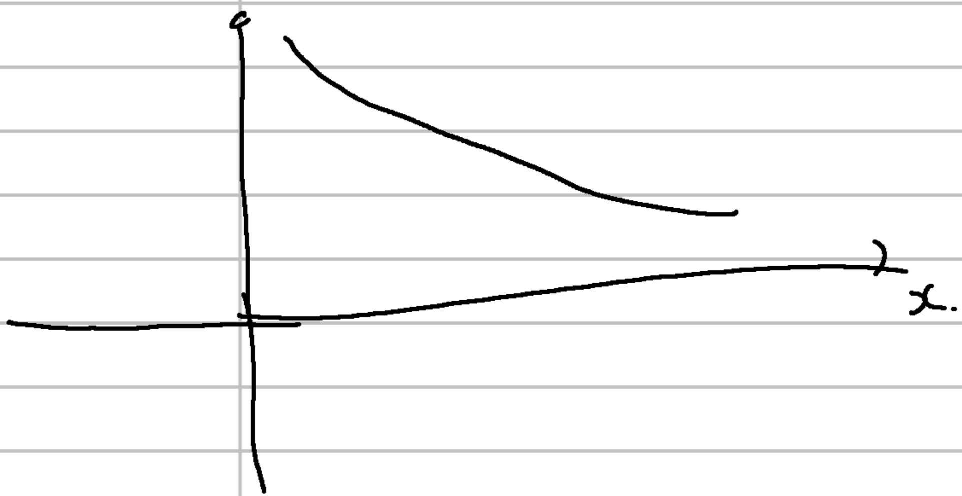
Flow

$\boxed{\text{Traj}} \xrightarrow{\text{sol.}} \boxed{\text{ODE}} \xleftarrow{\text{define.}} \boxed{\text{VF}}$

Blue - VF.

Red → Flow

$$\mu_t(x) = -\theta x.$$



$$\frac{d}{dt}\psi_t(x_0) = \mu_t(x)$$

claim $\psi_t(x_0) = \exp(-\theta t)\, x_0$

$t=0$ $\psi_0(x_0) = x_0.$

$$\frac{d}{dt}\psi_t(x_0) \overset{?}{=} \mu_t(\psi_t(x_0))$$

$$-\theta\, \underline{\exp(-\theta t)x_0} = -\theta\, \psi_t(x_0).$$

$$= \mu_t(\psi_t(x_0))$$

Numerical ODE:

$$\boxed{\text{Flow Model}}$$

$$p_{init} \xrightarrow[\text{ODE}]{} p_{data}$$

Neural network.

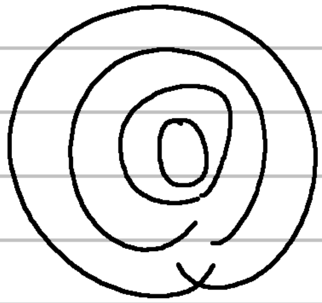
$u_t^\theta(x) : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ θ: parameters.

Random init $X_0 \sim p_{init}$

ODE: $\frac{d}{dt} X_t = u_t^\theta(X_t)$

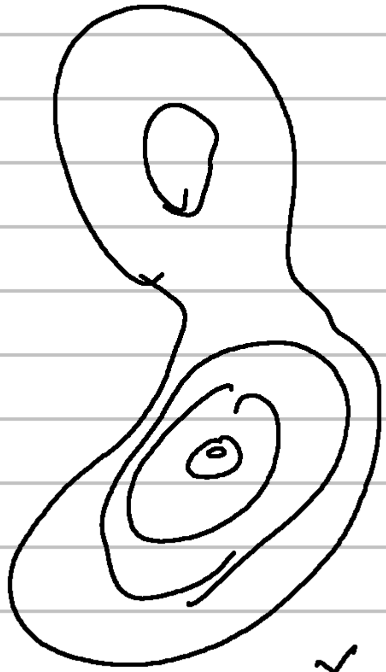
$$\boxed{\text{Goal: } X_1 \sim p_{data}}$$

Toy flow Model



x_0

Yaron Lipman



x_1

Diffusion Model

Stochastic process: X_t random variable ($0 \le t \le 1$)

$$X : [0,1] \to \mathbb{R}^d, \quad t \mapsto X_t$$

Vector Field: $u : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ + Diffusion coefficient

$$\sigma : [0,1] \to \mathbb{R}$$
$$t \to \sigma_t$$

Stochastic Differential Equation (SDE)

$X_0 = x_0$ (Initial Condition) $\quad dX_t = u_t(X_t)dt + \underbrace{\sigma_t dW_t}_{\text{Stochastic noise}}$

Stochastic process $W = (W_t)_{t \ge 0}$

1. $W_0 = 0$
2. Gaussian increments:
$$W_t - W_s \sim N(0, (t-s) I_d) \quad 0 \le s < t$$
3. Independent increments:

" dX_t " -

$$\frac{d}{dt} X_t = u_t(X_t) \Longleftrightarrow X_{t+h} = X_t + h\, u_t(X_t) + h R_t(h)$$

$$dX_t = u_t(X_t)\, dt + \sigma_t\, dW_t$$

$$X_{t+h} = X_t + h \cdot u_t(X_t) + \sigma_t (W_{t+h} - W_t)$$
$$+ h \cdot R_t(h)$$
$$\lim_{h \to 0} \sqrt{\|R_t\|^2} = 0$$

$$X_{t+h} = X_t + h u_t(X_t) + \sigma_t \sqrt{h}\, \epsilon.$$
$$\epsilon \sim \mathcal{N}(0, I_d)$$

Diffusion Model

$p_{init} \xrightarrow[\text{SDE.}]{} p_{data}$

N.N. $u_t^\theta \quad \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$

Diffusion coefficient σ_t (fixed)

Random init: $X_0 \sim p_{init}$

SDE: $dX_t = u_t^\theta(X_t)dt + \sigma_t dW_t$.

```
class ODE(ABC):
class SDE(ABC):

class Simulator(ABC):


class EulerSimulator(Simulator):
```

Mar. 7th

Diffusion model 2

$u_\theta(t)$

$$L(\theta) = \| u_t^\theta(x) - u_t^{target}(x) \|^2$$

training target.

Conditional.

Marginal

Probability Path.

Vector Field

Score Function

"Conditional" = Per single data point

"Marginal" = Across distribution of data points.

Prob. path..

Prob. path.

Dirac distribution: $z \in \mathbb{R}^d$, δ_z $\quad X \sim \delta_z$

$\Rightarrow X = z$.

Conditional probability path.

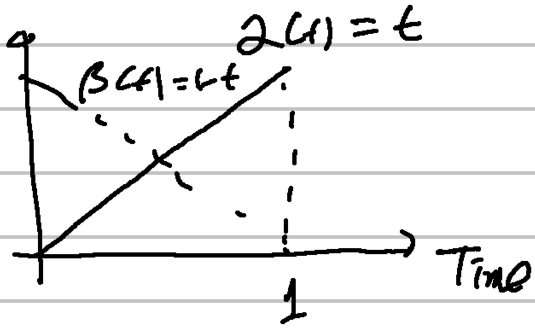
$$p_t(\cdot \mid z)$$

① $p_t(\cdot \mid z)$ distribution over $\mathbb{R}^d$

② $p_0(\cdot \mid z) = p_{init}$, $p_1(\cdot \mid z) = \delta_z$.

Example - Gaussian prob. path.

$$p_t(\cdot \mid z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$



$$p_0(\cdot \mid z) = \mathcal{N}(0, I_d)$$
$$p_1(\cdot \mid z) = \mathcal{N}(\alpha_1, 0)$$

Marginal prob. path

$z \sim p_{data}$, $\quad x \sim p_t(\cdot|z) \Rightarrow x \sim p_t$.

① $p_t(x) = \int p_t(x|z)\, p_{data}(z)\, dz$

② $p_0 = p_{init}$, $\quad p_1 = p_{data}$.

- - - - - - - - - - - -

$\left\{ \begin{array}{l} \text{Conditional} \\ \text{Marginal} \end{array} \right.$ — vector field.

Conditional vector field

$u_t^{target}(x|z) \qquad \left(\begin{array}{c} 0 \le t \le 1 \\ x, z \in \mathbb{R}^d \end{array} \right)$

$p_{init} \quad p_t(\cdot|z) \quad \delta_z$

$\xrightarrow{\text{ODE}}$

Such that $X_0 \sim p_{init}$, $\quad \dfrac{dX_t}{dt} = u_t^{target}(X_t|z)$

$p_{init} = p_0(\cdot|z)$

$\Rightarrow X_t \sim p_t(\cdot|z)$ $\quad (0 \le t \le 1)$

Conditional Gaussian Vector Filed.

$$\boxed{u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right)z + \frac{\dot{\beta}_t}{\beta_t}x}$$

$\dot{\square}$:= time derivative

Thm (Marginalization Trick):

The marginal vector field by

$$\boxed{u_t^{target}(x) = \int u_t^{target}(x|z)\frac{p_t(x|z)\,p_{data}(z)}{p_t(x)}dz.}$$

??

$$\left(p_t(x|z) = \frac{p_t(x,z)}{p_t(z)}.\right)$$

satisfies

$$\boxed{x_0 \sim p_{init},\ \frac{d}{dt}x_t = u_t^{target}(x_t) \Rightarrow x_t \sim p_t \ (0 \le t \le 1)}$$

Continuity Equation.

Proof

$$\frac{d}{dt} p_t(x) = \frac{d}{dt} \int p_t(x|z)\, p_{data}(z)\, dz.$$

$$= \int \frac{d}{dt} p_t(x|z)\, p_{data}(z)\, dz.$$

$$p(z|x) = \frac{p(x|z) \cdot p_{data}(z)}{p_t(x)}$$

Conditional
Marginal) Score function.

Conditional Score.

$$\nabla_x \log p_t(x|z)$$

conditional probability path.

Marginal Score

$$\nabla_x \log p_t(x)$$

Formula
$$\nabla \log p_t(x) = \frac{\nabla p_t(x)}{p_t(x)} \overset{①}{=} \frac{\nabla \int p_t(x|z)\, p_{data}(z)\, dz}{p_t(x)}$$

① $$\frac{\int \nabla p_t(x|z)\, p_{data}(z)\, dz}{p_t(x)} = \int \frac{\nabla \log p_t(x|z) \cdot p_t(x|z)\, p_{data}(z)\, dz}{p_t(x)}$$

Gaussian Score: $$\nabla_x \log p_t(x|z) = -\frac{(x-\alpha_t z)}{\beta_t^2}$$

$$p_t(x|z) \sim \exp\left[-\frac{(x-\alpha_t z)^2}{2\beta_t^2}\right]$$

Stochastic Differential Equation (SDE)

Let $u_t^{target}(x)$

Then, for arbitrary $\sigma_t \geq 0$

$$X_0 \sim p_{init}, \quad dX_t = \left[u_t^{target}(X_t) + \frac{\sigma_t^2}{2} \nabla \log p_t(X_t) \right] dt + \sigma_t dW_t.$$

$$\Rightarrow X_t \sim p_t \quad (0 \leq t \leq 1)$$

$$\underline{X_1 \sim p_{data}.}$$

March. 8th — Note

Flow and Diffusion Models

<Training Flow and Diffusion Models>

Conditional {Probability Path, Vector field, Score function}.

Flow Matching

u_t^θ (θ: parameters)

Goal: $u_t^\theta \simeq u_t^{target}$

Flow matching loss: $\mathcal{L}_{fm}(\theta) = \mathbb{E}\left[\|u_t^\theta(x) - u_t^{target}(x)\|^2\right]$

$t \sim$ uniform $\leftarrow$ Uniform in $[0, 1]$

$z \sim p_{data} \leftarrow$ draw data point

! (minimizer is Not tractable) $x \sim p_t(\cdot|z) \leftarrow$ draw from cond prob. path.

Conditional Flow matching loss:

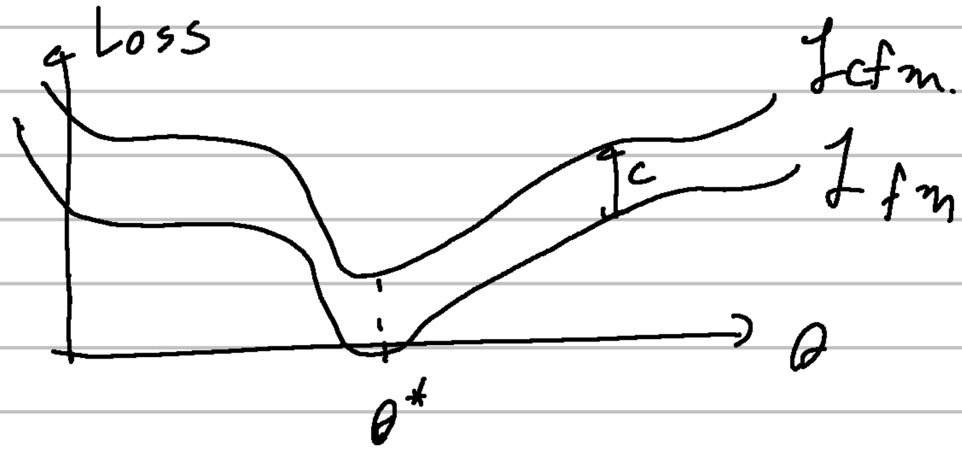
$$\mathcal{L}_{cfm}(\theta) = \mathbb{E}\left[\|u_t^\theta(x) - u_t^{target}(x|z)\|^2\right]$$

✓. Tractable.

? Minimizer

Thm.

$$\boxed{\begin{aligned} &\mathcal{L}_{fm}(\theta) = \mathcal{L}_{cfm}(\theta) + C \\ &\text{for } C < 0, \text{ independent of } \theta. \end{aligned}}$$



$\Rightarrow$ ① For minimizer θ^* of $\mathcal{L}_{cfm}$, $u_t^{\theta^*} = u_t^{target}$

② $\nabla_\theta \mathcal{L}_{cfm} = \nabla_\theta \mathcal{L}_{fm}(\theta)$

$\mathcal{L}_{cfm}$ for Gaussian cond. path

$$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$

$$u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right)\cdot z + \frac{\dot{\beta}_t}{\beta_t}\cdot x$$

$$\epsilon \sim \mathcal{N}(0, I_d) \Rightarrow \alpha_t z + \beta_t \epsilon \overset{def}{=} x \sim p_t(\cdot|z)$$

$$\mathcal{L}_{cfm}(\theta) = \mathbb{E}_{\substack{t \sim \text{Unif.} \\ z \sim p_{data} \\ x \sim \mathcal{N}(\alpha_t z,\, \beta_t^2 I_d)}} \left[\left\| u_t^\theta(x) - \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z - \frac{\dot{\beta}_t}{\beta_t}(\alpha_t z + \beta_t \varepsilon) \right\|^2 \right]$$

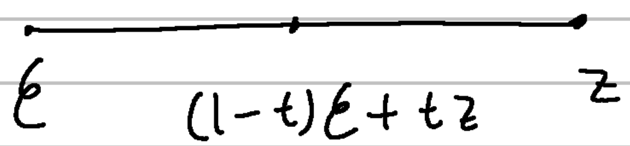
$$= \quad \left\| u_t^\theta(x) - (\dot{\alpha}_t z + \dot{\beta}_t \varepsilon) \right\|^2$$

Conditinal Optimal path.

Cond. path $\begin{cases} \alpha_t = t \\ \beta_t = 1 - t \end{cases}$ $\qquad \dot{\alpha} = 1$, $\dot{\beta} = -1$

$$\mathcal{L}_{cfm}(\theta) = \mathbb{E}\left[\| u_t^\theta(\alpha_t z + \beta_t \varepsilon) - (z - \varepsilon)\|^2\right]$$

$$= \mathbb{E}\left[\| u_t^\theta(t z + (1-t)\varepsilon) - (z - \varepsilon)\|^2\right]$$



$$\|a-b\|^2 = \|a\|^2 - 2a^T b + \|b\|^2 \qquad (a, b) \in \mathbb{R}^d.$$

Proof.

$$L_{fm}(\theta) = \mathbb{E}_x \left[\| u_t^\theta(x) - u_t^{target}(x) \|^2 \right]$$

$$\underline{-2 u_t^\theta(x)\, u_t^{target}(x)}$$

$$L_{cfm}(\theta) = \mathbb{E}_x \left[\| u_t^\theta(x) - u_t^{target}(x|z) \|^2 \right]$$

$$\underline{-2 u_t^\theta(x)\, u_t^{target}(x|z)}$$

4.2. Score matching.

Score Network: s_t^θ (θ: parameters)

Goal $s_t^\theta \approx \nabla \log p_t$

Score matching loss $\leftarrow$ Minimizer is Not tractable

$$L_{sm}(\theta) = \mathbb{E}_{t, x, z}\left[\| s_t^\theta(x) - \nabla \log p_t(x) \|^2 \right]$$

Conditional Score matching loss $\leftarrow$ Minimizer is tractable.
(aka denoising score matching loss)

$$L_{dsm}(\theta) = \mathbb{E}_{t, z, x}\left[\| s_t^\theta(x) - \nabla \log p_t(x|z) \|^2 \right]$$

Thm. 4.2 $\boxed{L_{sm}(\theta) = L_{dsm}(\theta) + C}$

① For θ^* of L_{dsm} $s_t^{\theta^*} = \nabla \log p_t$ $c < 0$

② $\nabla_\theta L_{sm}(\theta) = \nabla_\theta L_{dsm}(\theta)$

Proof :

Reminder

$$\nabla \log p_t(x) = \int \underbrace{\nabla \log p_t(x|z)}_{\text{Conditional score function}} \frac{p_t(x|z)\, p_{data}(z)}{p_t(x)} \, dz.$$

Denoising Score Matching for Gaussian Prob Path

$$\nabla \log p_t(x|z) = -\frac{x-\alpha_t z}{\beta_t^2}$$

$$\epsilon \sim \mathcal{N}(0, I_d) \Rightarrow \quad x = \alpha_t z + \beta_t \epsilon \sim \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$

$$\mathcal{L}_{dsm}(\theta) = \mathbb{E}_{t \sim Unif,\, z \sim p_{data},\, x \sim p_t(\cdot|z)} \left[\left\| s_t^\theta(x) + \frac{x-\alpha_t z}{\beta_t^2} \right\|^2 \right]$$

$$= \quad '' \quad \left[\left\| s_t^\theta(\alpha_t z + \beta_t \epsilon) + \frac{\epsilon}{\beta_t} \right\|^2 \right]$$

Numerically unstable for low β_t.

Stochastic Sampling

Denoising Diffusion Models (DDMs).

Special property about DDMs:

Next steps:

- N.N. architectures
- Conditiong on a prompt
- Image generators
- Applications

```
class ODE:
    def drift_coefficient.

class SDE:
    def drift_coeff.

    def diffusion_coefficient.

class Simulator:
    def step
    def simulate
    def simulate_trajectory

class EulerSimulator(Simulator):
```