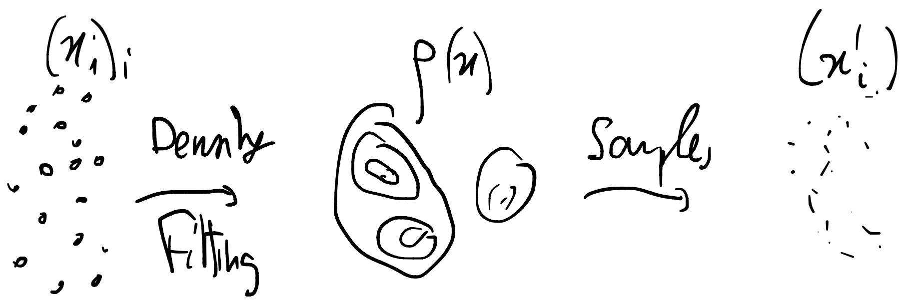
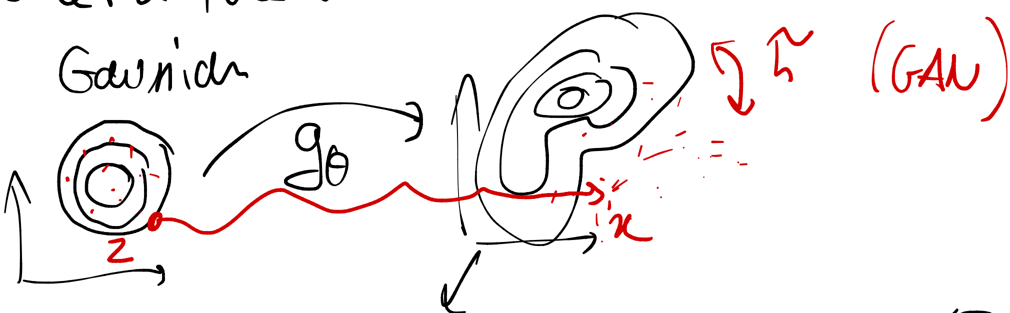
Generative modelling

$\boxed{\text{Dalle}}$ ← - - - ChatGPT



"General idea"

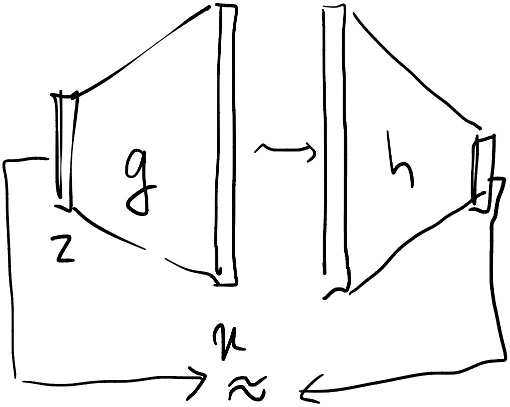


- VAE
- GAN ←
- Normalizing Flow
- Diff. Model

"Progressive" → Stable

⊝ Slow

VAE



1 Sampling / Langevin

SGD: $\min f(x) = \sum_k f_k(x)$

$$x_{t+1} = x_t - \tau_t \underbrace{\nabla f_k(x_t)}_{\nabla f(x_t) + \varepsilon_t} \qquad k = \text{rand}.$$

$$x_{t+1} = x_t - \tau_t \nabla f(x_t) + \tau_t \boxed{\varepsilon_t}$$

$$\tau_t \searrow 0 \qquad \tau_t = \frac{1}{t}$$

Decaying τ_t

$\tau_t = \tau$ constant

Langevin Monte Carlo

$$x_{t+1} = x_t - \tau_t \nabla f(x) + 2\sqrt{\tau_t} \cdot \overbrace{W_t}^{\mathcal{N}(0, I_d)}$$

Discrete evol$^{\underline{c}}$

$t \to 0$

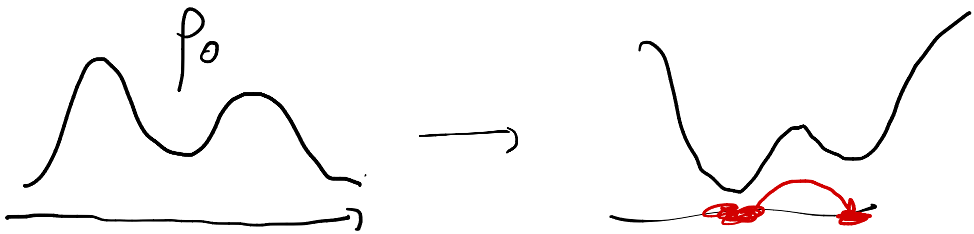
$$\frac{dx}{dt} = -\nabla f(x) + W(t)$$

$$dx = -\nabla f(x)\,dt + dW(t)$$

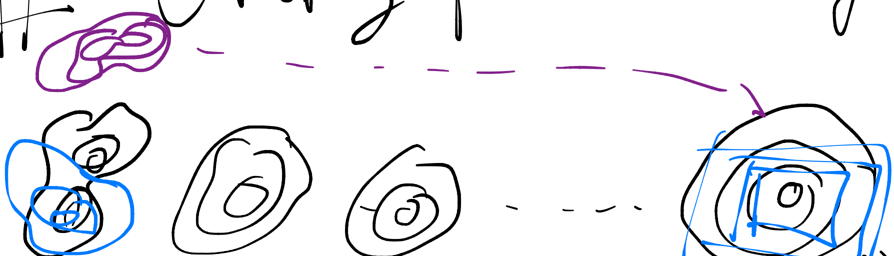
SDE

Thm: $\underbrace{x(t)}_{\text{Init.}} \xrightarrow{\text{Law}} \underbrace{\frac{1}{2} e^{-f(x)}}_{\rho_0(x)}$

$f = -\log(\rho_0) \rightarrow$ super slow



Diff: ① Noising process $\rightarrow$ Langevin



$\rho_0(x) \xleftarrow{\text{Diff Model}} N(0, I_d) = e^{-\|x\|^2/2}$

$$x_{t+1} = X_t - \tau X_t + \sqrt{2\tau}\, W_t$$

$$\hookrightarrow dx = \boxed{-\underline{x}} + dW$$

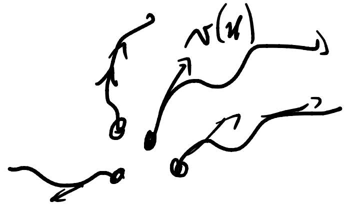
Prop: $X_t = X_0 e^{-t} + \sqrt{1-e^{-2t}}\, Z$

Evol^c of $\rho_t(x)$ the law of X_t?

Fokker_Planck

① Conservation eq^c // Advection.

$$\frac{dx}{dt} = v(x)$$



$\rho(x)$ = density.

$$\frac{\partial \rho}{\partial t} = -\mathrm{div}(\rho v)$$

scalar

vector

vector

$$\mathrm{div}(v) = \sum \frac{\partial v_i}{\partial x_i} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}$$

Fokker–Planck: $dx = v(x) + \underbrace{dW}$

$$\frac{\partial \rho}{\partial t} = \mathrm{div}(\rho v) + \underbrace{\Delta \rho}$$

$v = 0 \longrightarrow$ Heat eq.

Diffo-model: Invert $\quad t \to -t$

$$\rho_t \longrightarrow \xi_t = \rho_{T-t}$$

$$\frac{\partial \xi}{\partial t} = -\mathrm{div}(\xi v) - \Delta \xi$$

$$\Delta \rho_t = \mathrm{div}(\nabla \rho_t) = \mathrm{div}\left(\rho_t \cdot \frac{\nabla \rho_t}{\rho_t}\right)$$

$$= \mathrm{div}(\rho_t \cdot \nabla[\log(\rho)])$$

SCORE $\eta_t \equiv \nabla \log(\rho_t)$

$$\Delta \rho_t = \mathrm{div}(\rho_t \cdot \eta_t)$$

$$-\Delta \xi_t \longrightarrow +\alpha \Delta \xi_t$$

$$\Leftrightarrow \frac{\partial \xi}{\partial t} = \begin{matrix} -\mathrm{div}(\xi v) \\ -\mathrm{div}(\xi_t \eta_t)(1+\alpha) \end{matrix} + \boxed{\alpha} \Delta \xi_t$$

Reverse eq !

$$\frac{\partial \xi_t}{\partial t} = -\mathrm{div}\big(\xi_t \times (v + (1+\alpha)\eta_t)\big) + \alpha \Delta \xi_t$$

usualley $\rightarrow$ $\alpha = 1$

$\xi_t(y)$ $\quad Y_t$

$$dY_t = -\left(Y_t + \underbrace{(1+\alpha)\eta_t(Y)}\right) + \alpha\, dW_t$$

τ step size

$$Y_{t+1} = Y_t - \tau\left(Y_t + (1+\alpha)\eta_t(Y_t)\right)$$
$$+ \sqrt{2\tau}\; W_t$$

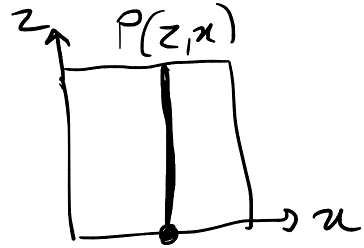
Compute/estimate $\cancel{\rho_0(x)}$

$$\eta_t = \nabla \log(\rho_t) : \mathbb{R}^d \to \mathbb{R}^d$$

Tweedie Formula η_t as an average

$X_0 \to X$ $\qquad \rho_0$

$X_t \to Z$ $\qquad \rho_t$



$$\mathbb{P}(z|x) = \frac{P(z,x)}{P(x)}$$

known

$$\log(\mathbb{P}(z|x)) = \frac{\|z - e^{-t}x\|^2}{2(1-e^{-t})}$$

$$\nabla[\quad] = \frac{z - e^{-t}x}{(1-e^{-t})}$$

Tweedy!

$$\eta_t(z) = \nabla \log \mathbb{P}(z) = \int_x \nabla_z \log(\mathbb{P}(z|x))\, d\mathbb{P}(x|z)$$

$$\Updownarrow$$

$$\eta_t(z) = \underset{\boxed{\varphi_t(z)}}{\operatorname{argmin}} \int \|\nabla_z \log(\mathbb{P}(z|x)) - \varphi_t(z)\|^2\, d\mathbb{P}(x|z)$$

Key step: impose $\varphi_t(z) = \psi_\theta(z,t)$

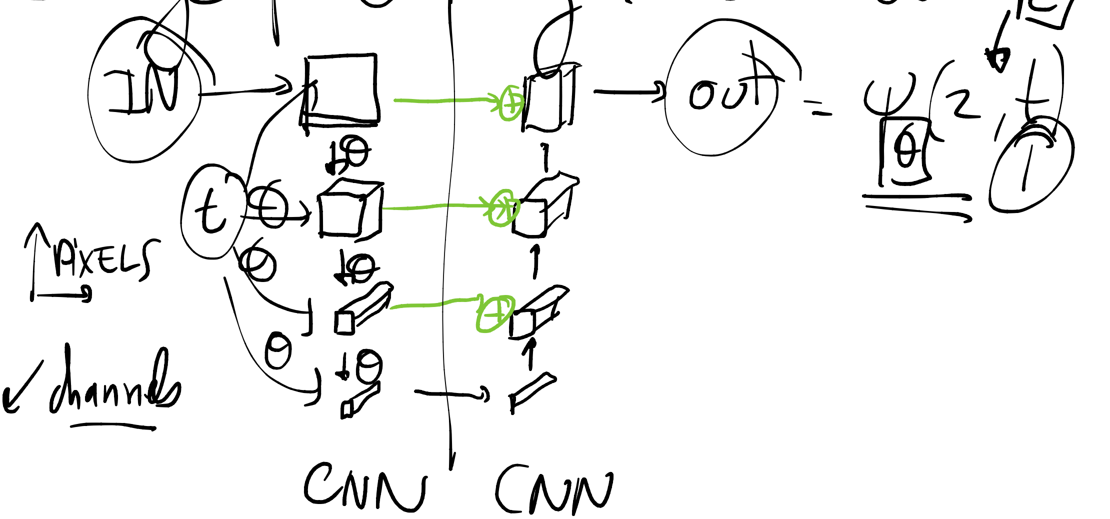
Neural Networks

In practice:

$$\min_\theta \int_C \int_0^T \int_x \int_z \Big\| \underbrace{\overbrace{\nabla_z \log(\mathbb{P}(z|x))}^{\text{easy}}}_{\text{above}} - \psi_\theta(z,t)\Big\|^2 \, d\mathbb{P}(x,z)\, dt$$

denoising Score Matching

Image processing: U-Net



Gen$^{\underline{c}}$

$$\psi_\theta(z, t_0)$$

$$\psi_\theta(z, t_1)$$

$$\vdots$$