1. VAE intro
2. DM as particular case VAE
3. SDE
4. LD and FP eq.
5. DM as Score-matching
6. Classifier guidance

$X = (x_1 \dots x_n) \qquad x_i \in \mathbb{R}^D \qquad z_i \in \mathbb{R}^d$

$\log p(X|\theta)$ → $\max_\theta$

$p(x, z|\theta) = p(x|z, \theta)\, p(z) =$

$= \mathcal{N}(x \,|\, \Phi(z, \theta), \sigma^2 I)\, \mathcal{N}(z|0, I)$

↑ deep NN

ELBO

$p(X|\theta) = \int p(X, z|\theta)\, dz$

$\log p(X|\theta) \geq \mathcal{L}(q, \theta) = \int q(z|X, \varphi) \log \frac{p(X, z|\theta)}{q(z|X, \varphi)}\, dz \to \max_{\theta, \varphi}$

deep NN

$\int q(z|X, \varphi) \log \frac{p(X|z, \theta)\, p(z)}{q(z|X, \varphi)}\, dz =$

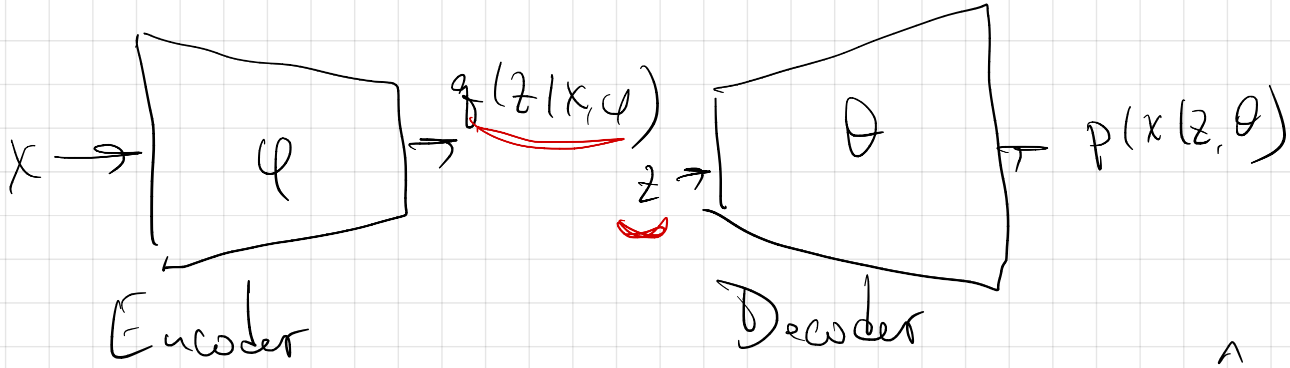
$= \int q(z|X, \varphi) \log \underline{p(X|z, \theta)}\, dz - \int q(z|X, \varphi) \log \frac{q(z|X, \varphi)}{p(z)}\, dz \to \max_{\theta, \varphi}$

Reconstruction term

$KL(q(z|X, \varphi) \,\|\, p(z)) \geq 0$

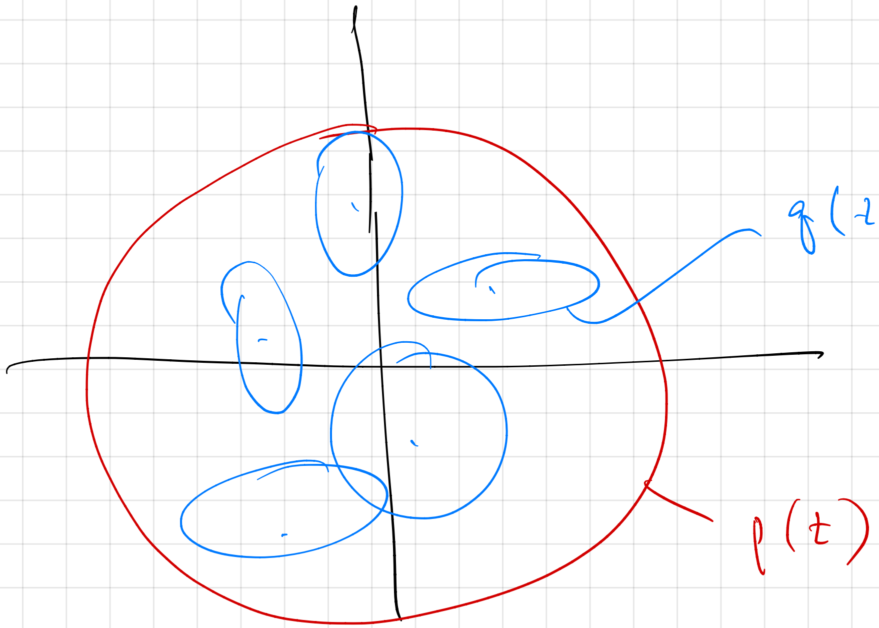
Regularizer

$$\sum_{i=1}^{n} \int q(z_i|x_i,\varphi) \log p(x_i|z_i,\theta)\, dz_i - \sum_{i=1}^{n} KL\big(q(z_i|x_i,\varphi) \,\|\, p(z_i)\big) \to \max_{\theta,\varphi}$$



$\hat{z} \sim p(z)$

$x \sim p(x|\hat{z},\theta)$



$$p(z) \neq \frac{1}{n} \sum_{i=1}^{n} q(z|x_i,\varphi)$$

Observed $x_0 \to x_1 \to \dots \to x_t \to \dots \to x_T$ Latent variables

$$\boxed{x_{t+1} = \sqrt{1-\beta}\, x_t + \sqrt{\beta}\, \varepsilon} \qquad \beta \ll 1 \qquad \varepsilon \sim \mathcal{N}(\varepsilon | 0, I)$$

$$\boxed{q(x_{t+1} | x_t)} = \mathcal{N}(x_{t+1} | \sqrt{1-\beta}\, x_t, \beta I)$$

$$x_t = \sqrt{\alpha_t}\, x_0 + \sqrt{1-\alpha_t}\, \varepsilon \qquad \text{where} \quad \alpha_t = (1-\beta)^t$$

$$\mathbb{D} x_0 = 1 \quad \Rightarrow \quad \forall t \quad \mathbb{D} x_t = 1$$

$$\alpha_T \ll 1, \quad \alpha_T \simeq 0 \qquad q(x_T | x_0) \approx \mathcal{N}(x_T | 0, I)$$

$$\boxed{q(x_t | x_0) = \mathcal{N}(x_t | \sqrt{\alpha_t}\, x_0, (1-\alpha_t))} \xrightarrow{t \to T} \mathcal{N}(x_T | 0, I)$$

$$p(\underbrace{Obs}, Hid | \theta)$$

$$\log p(Obs | \theta) \geq \int q(Hid | Obs, \varphi) \log \frac{p(Obs, Hid | \theta)}{q(Hid | Obs, \varphi)}\, d\,Hid \to \max_{\theta, \varphi}$$

$$\log p(x_0 | \theta) \geq \int q(x_1 \dots x_T | x_0) \log \frac{p(x_0, x_1, \dots x_T | \theta)}{q(x_1 \dots x_T | x_0)}\, dx_1 \dots dx_T =$$

$$= \begin{Bmatrix} p(x_0 \dots x_T \mid \theta) = p(x_0 \mid x_1, \theta)\, p(x_1 \mid x_2, \theta) \dots p(x_{T-1} \mid x_T, \theta)\, p(x_T) \\ \underline{q(x_1 \dots x_T \mid x_0)} = q(x_1 \mid x_0)\, q(x_2 \mid x_1) \dots q(x_T \mid x_{T-1}) \end{Bmatrix} =$$

$$= \int q(x_1 \dots x_T \mid x_0) \underline{\log p(x_0 \mid x_1, \theta)}\, dx_1 \dots dx_T + \boxed{\int q(x_1 \dots x_T \mid x_0) \log \frac{p(x_1 \dots x_T \mid \theta)}{q(x_1 \dots x_T \mid x_0)}\, dx_1 \dots dx_T} =$$

$$\to -KL\big(q(x_1 \dots x_T \mid x_0) \,\|\, p(x_1 \dots x_T \mid \theta)\big)$$

$$= \int q(x_1 \mid x_0) \log p(x_0 \mid x_1, \theta)\, dx_1 + \int q(x_1 \dots x_T \mid x_0) \log \frac{p(x_T)\, p(x_{T-1} \mid x_T, \theta) \dots p(x_1 \mid x_2, \theta)}{q(x_T \mid x_0)\, q(x_{T-1} \mid x_T, x_0) \dots q(x_1 \mid x_2, x_0)}\, dx_1 \dots dx_T =$$

$$= \int q(x_1 \mid x_0) \log \underline{p(x_0 \mid x_1, \theta)}\, dx_1 - \sum_{t=2}^{T} \int q(x_t \mid x_0)\, KL\big(q(x_{t-1} \mid x_t, x_0) \,\|\, p(x_{t-1} \mid x_t, \theta)\big) -$$

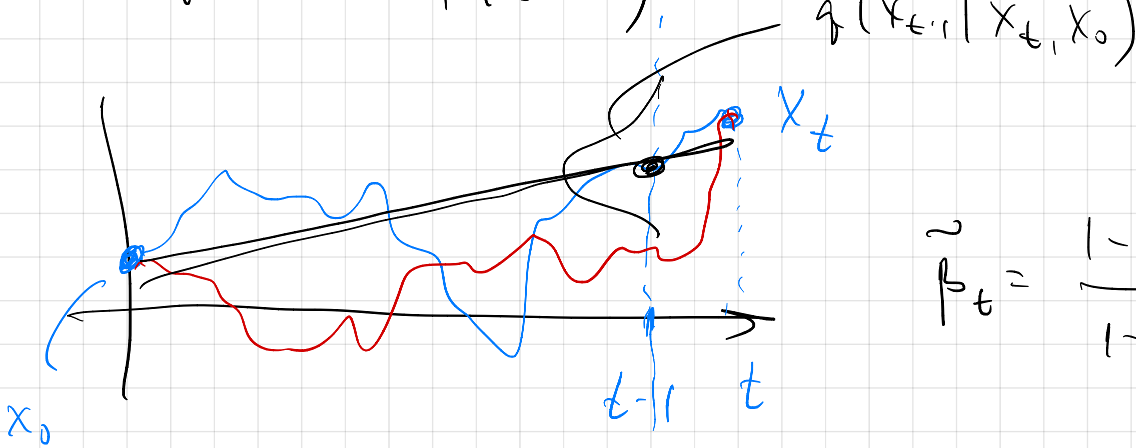
Const

$$- \underbrace{KL\big(q(x_T \mid x_0) \,\|\, p(x_T)\big)}_{0} \to \max_{\theta}$$

$$\boxed{\mathcal{L}(\theta) \approx - \sum_{t=1}^{T} \mathbb{E}_{x_t}\, KL\big(\underline{q(x_{t-1} \mid x_t, x_0)} \,\|\, \underline{p(x_{t-1} \mid x_t, \theta)}\big)}$$

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1}, \cancel{x_0})\, q(x_{t-1}|x_0)}{q(x_t|x_0)} = \frac{q(x_t|x_{t-1})\, q(x_{t-1}|x_0)}{q(x_t|x_0)} =$$

$$= \mathcal{N}\left(x_{t-1} \,\middle|\, \mu(x_0, x_t), \tilde{\beta}_t I\right)$$



$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta$$

$$\boxed{\mu(x_0, x_t) = \frac{\sqrt{\bar{\alpha}_t}\,\beta}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{1-\beta}\,(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t}$$

$$q(x_{t-1}|x_t, x_0)$$

$$p(x_{t-1}|x_t, \theta) = q(x_{t-1}|x_t, x_\theta(x_t, t))$$

deep NN

$$KL\left(q(x_{t-1}|x_t, x_0) \,\|\, p(x_{t-1}|x_t, x_0)\right) = \text{Const} \, \|x_0 - x_\theta(x_t, t)\|^2$$

Training:

1) Take arbitrary x_0 from training data
2) Take arbitrary $t \in [2, T]$
3) Generate $x_t \sim q(x_t | x_0)$ $\quad x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, \varepsilon$, $\quad \varepsilon \sim \mathcal{N}(\varepsilon | 0, I)$
4) Differentiate $\cancel{KL\left(q(x_{t-1} | x_t, x_0) \,\|\, p(x_{t-1} | x_t, \theta)\right)}$ w.r.t. θ

$\mathrm{Const}\, \| x_0 - x_\theta(x_t, t) \|^2$

$$x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\, \varepsilon$$

$$x_0 = \frac{x_t - \sqrt{1-\bar\alpha_t}\, \varepsilon}{\sqrt{\bar\alpha_t}}$$

$$x_\theta(x_t, t) = \frac{x_t - \sqrt{1-\bar\alpha_t}\, \varepsilon_\theta(x_t, t)}{\sqrt{\bar\alpha_t}}$$

$$\mathrm{Const}\, \| x_0 - x_\theta(x_t, t) \|^2 = \mathrm{Const}'\, \| \varepsilon - \varepsilon_\theta(x_t, t) \|^2$$

Exactly that sample of noise which was used to get x_t from x_0

Generation

1) $x_T \sim \mathcal{N}(x_T \mid 0, I) = p(x_T)$

2) $x_{t-1} \sim p(x_{t-1} \mid x_t, \theta) = q(x_{t-1} \mid x_t, x_\theta(x_t, t)) =$

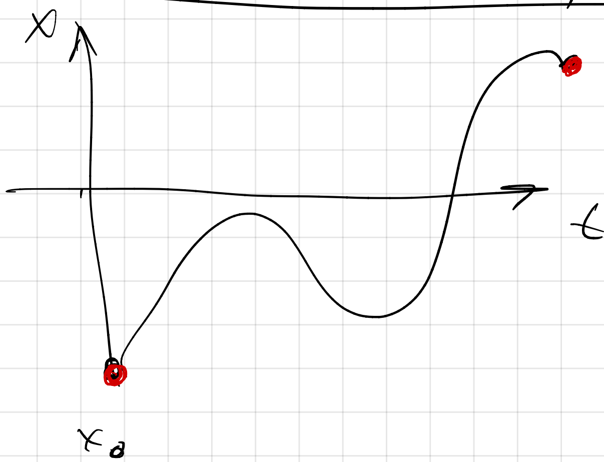
$= \mathcal{N}(x_{t-1} \mid \mu(x_t, x_\theta(x_t, t)), \tilde{\beta}_t I)$

3) Obtain $x_0 \sim p(x_0 \mid x_1) = \mathcal{N}(x_0 \mid x_1, s^2 I)$ $\quad s \ll 1$

ODE:

$dx = f(x, t)\, dt$

x_0 - initial point

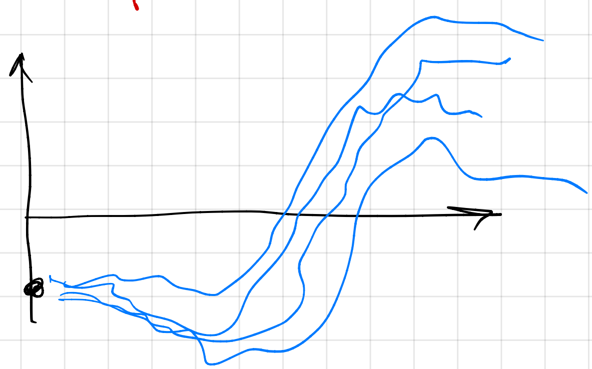


SDE:

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

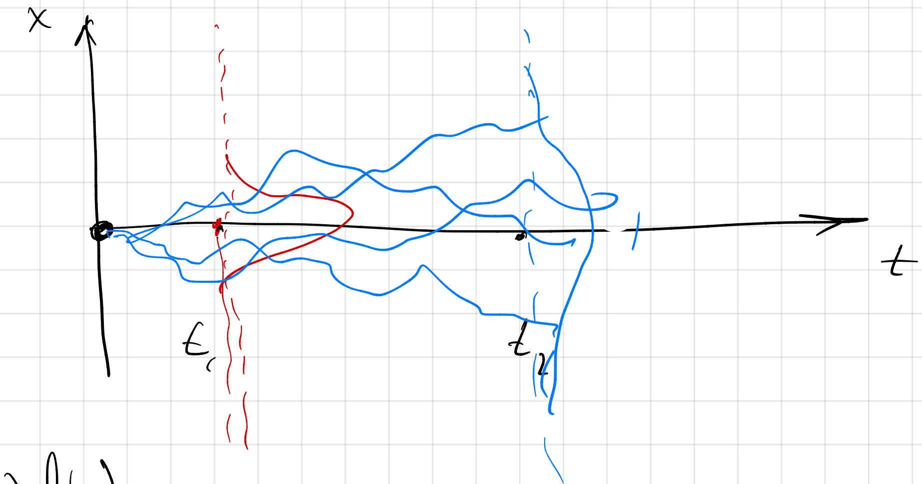
Wiener process

$dW \sim \mathcal{N}(dW \mid 0, dt)$



$$dx = dW$$

$$x_0 = 0$$



$$x_t \sim \mathcal{N}(x_t \mid 0, t)$$

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

$x_0 \sim p_0(x_0)$ $\qquad\qquad$ $p_t(x)$

Fokker-Planck equation

$$\frac{\partial p_t(x)}{\partial t} = \underbrace{-\frac{\partial}{\partial x}\left(f(x,t)\,p_t(x)\right)} + \underbrace{\frac{1}{2} g(t) \frac{\partial^2}{\partial x^2} p_t(x)} \qquad x \in \mathbb{R}$$

(1) $g(t) = 1$ $\qquad$ Let $f(x,t) = \frac{1}{2}\frac{\partial}{\partial x} \log p_t(x)$

$$\boxed{dx = \frac{1}{2}\frac{\partial}{\partial x} \log p_t(x) + dW}$$ — Langevin equation

$x_0 \sim p_0(x)$

$$\frac{\partial p_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial}{\partial x} \log p_t(x) \cdot p_t(x)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x) =$$

$$= -\frac{1}{2}\frac{\partial}{\partial x}\left(\frac{1}{\cancel{p_t(x)}}\frac{\partial}{\partial x} p_t(x) \cdot \cancel{p_t(x)}\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x) = 0$$

$$dx = \frac{1}{2}\frac{\partial}{\partial x}\log p_0(x)\,dt + dW \quad \text{- Forward LE}$$

$x_0 \sim p_0(x)$

tries to find mode of $p_0(x)$

destroys the information

Backward LE

$$dx = dW - \frac{1}{2}\frac{\partial}{\partial x}\log p_0(x)\,dt$$

1. $dx = h(x,t)\,dt$ where $h(x,t) = f(x,t) - \frac{g^2(t)}{2}\frac{\partial}{\partial x}\log q_t(x)$

$x_0 \sim q_0(x)$

$$\frac{\partial q_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left(h(x,t)q_t(x)\right) = -\frac{\partial}{\partial x}\left(\left(f(x,t) - \frac{g^2(t)}{2}\frac{\partial}{\partial x}\log q_t(x)\right)q_t(x)\right) =$$

$$= -\frac{\partial}{\partial x}\left(f(x,t)q_t(x)\right) + \frac{1}{2}g^2(t)\frac{\partial}{\partial x}\left(q_t(x)\frac{\partial}{\partial x}\log q_t(x)\right) =$$

$$= \boxed{-\frac{\partial}{\partial x}\left(f(x,t)q_t(x)\right) + \frac{1}{2}g^2(t)\frac{\partial^2}{\partial x^2}q_t(x)}$$

Remember that $dx = f(x,t)\,dt + g(t)\,dW$

$x_0 \sim p_0(x)$

$$\boxed{\frac{\partial p_t}{\partial x} = -\frac{\partial}{\partial x}\left(f(x,t)p_t(x)\right) + \frac{1}{2}g^2(t)\frac{\partial^2}{\partial x^2}p_t(x)}$$

Exactly the same dynamics if $p_0(x) = q_0(x)$

③ Consider ODE

$$dx = \left(f(x,t) - \frac{g^2(t)}{2}\frac{\partial}{\partial x}\log p_t(x)\right)dt$$

$$x_0 \sim p_0(x)$$

Add forward LE:

does not affect the evolution of $p_t(x)$

$$dx = \left(f(x,t) - \cancel{\frac{g^2(t)}{2}\frac{\partial}{\partial x}\log p_t(x)}\right)dt + \boxed{\cancel{\frac{g^2(t)}{2}\frac{\partial}{\partial x}\log p_t(x)\,dt} + g(t)\,dW} =$$

$$= \boxed{f(x,t)\,dt + g(t)\,dW} \quad \text{- Forward SDE}$$

Add backward LE

$$dx = \left(f(x,t) - \frac{g^2(t)}{2}\frac{\partial}{\partial x}\log p_t(x)\right)dt \boxed{- \frac{g^2(t)}{2}\frac{\partial}{\partial x}\log p_t(x)\,dt + g(t)\,dW} =$$

$$= \boxed{\left(f(x,t) - g^2(t)\frac{\partial}{\partial x}\log p_t(x)\right)dt + g(t)\,dW} \quad \text{- backward SDE}$$

To go backward in time we need to know $\frac{\partial}{\partial x}\log p_t(x)$

also known as score function

$$x_{t+1} = \sqrt{1-\beta}\, x_t + \sqrt{\beta}\, \varepsilon \qquad \beta \ll 1$$

let us $\beta = \hat{\beta}\, dt$

$$x_{t+1} = \sqrt{1-\hat{\beta}\, dt}\, x_t + \sqrt{\hat{\beta}\, dt}\, \varepsilon = \left(1 - \frac{\hat{\beta}}{2} dt\right) x_t + \sqrt{\hat{\beta}}\, \varepsilon \sqrt{dt} + \bar{o}(dt)$$

$$dx = x_{t+1} - x_t \qquad \varepsilon \sqrt{dt} = \underline{dW} \sim \mathcal{N}(dW \mid 0, dt)$$

$$dx = -\frac{\hat{\beta}}{2} x_t\, dt + \sqrt{\hat{\beta}}\, dW \quad \text{- Forward SDE}$$

$$p_0(x) \longrightarrow p_T(x) \approx \mathcal{N}(x \mid 0, I)$$

Backward SDE:

$$dx = \Big(-\frac{\hat{\beta}}{2} x_t - \hat{\beta} \underbrace{\frac{\partial}{\partial x} \log p_t(x)}_{\text{score function}}\Big) dt + \sqrt{\hat{\beta}}\, dW$$

$$s_\theta(x,t) \approx \frac{\partial}{\partial x} \log p_t(x)$$

$$\int p_t(x_t) \left\| s_\theta(x_t, t) - \frac{\partial}{\partial x} \log p_t(x_t) \right\|^2 dx_t \rightarrow \min_\theta$$

$$= \text{Const} + \int p_0(x_0) \int q(x_t \mid x_0) \left\| s_\theta(x_t, t) - \frac{\partial}{\partial x} \log q(x_t \mid x_0) \right\|^2 dx_t\, dx_0$$

we know how to compute it for DM

Remember that $q(x_t|x_0) = \mathcal{N}(x_t|\sqrt{\alpha_t}\, x_0, (1-\alpha_t))$

$$\frac{\partial}{\partial x_t} \log q(x_t|x_0) = -\frac{(x_t - \sqrt{\alpha_t}\, x_0)}{(1-\alpha_t)} = -\frac{\varepsilon}{\sqrt{1-\alpha_t}}$$

$$x_t = \sqrt{\alpha_t}\, x_0 + \sqrt{1-\alpha_t}\cdot\varepsilon \qquad \varepsilon = \frac{x_t - \sqrt{\alpha_t}\, x_0}{\sqrt{1-\alpha_t}}$$

$$\int q(x_t|x_0) \| \varepsilon_\theta(x_t, t) - \varepsilon \|^2 dx_t \iff \int q(x_t|x_0) \| s_\theta(x_t, t) - \frac{\partial}{\partial x} \log q(x_t|x_0) \|^2 dx_t$$

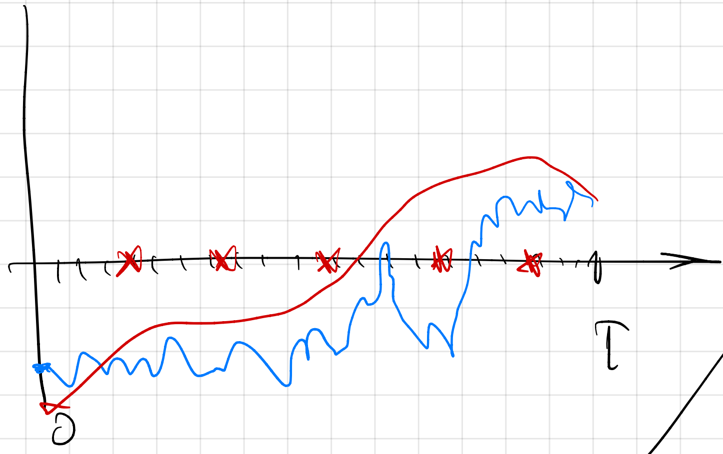
$$s_\theta(x_t, t) = -\frac{\varepsilon_\theta(x_t, t)}{\sqrt{1-\alpha_t}}$$

Equivalent ODE

$$\boxed{dx = \left(f(x,t) - \frac{g^2(t)}{2} \frac{\partial}{\partial x} \log p_t(x)\right) dt}$$

From seminal paper on Neural ODE we know how to efficiently the value of $p_0(x)$

DDIM



Conditional Generation

Assume that we have unconditional Generative model $p_\theta(x)$

We would like to generate objects from $p_\theta(x|y)$

$$p_\theta(x|y) = \frac{p(y|x)\,p_\theta(x)}{p(y)}$$

$$\frac{\partial \log p_\theta(x|y)}{\partial x} = \frac{\partial}{\partial x} \log p(y|x) + \frac{\partial}{\partial x} \log p_\theta(x)$$

$$\frac{\partial \log p_t(x_t|y)}{\partial x_t} = \underbrace{\frac{\partial}{\partial x} \log p(y|x_t)} + \underbrace{\frac{\partial}{\partial x} \log p_t(x_t)}_{\text{We know how to compute it}}$$

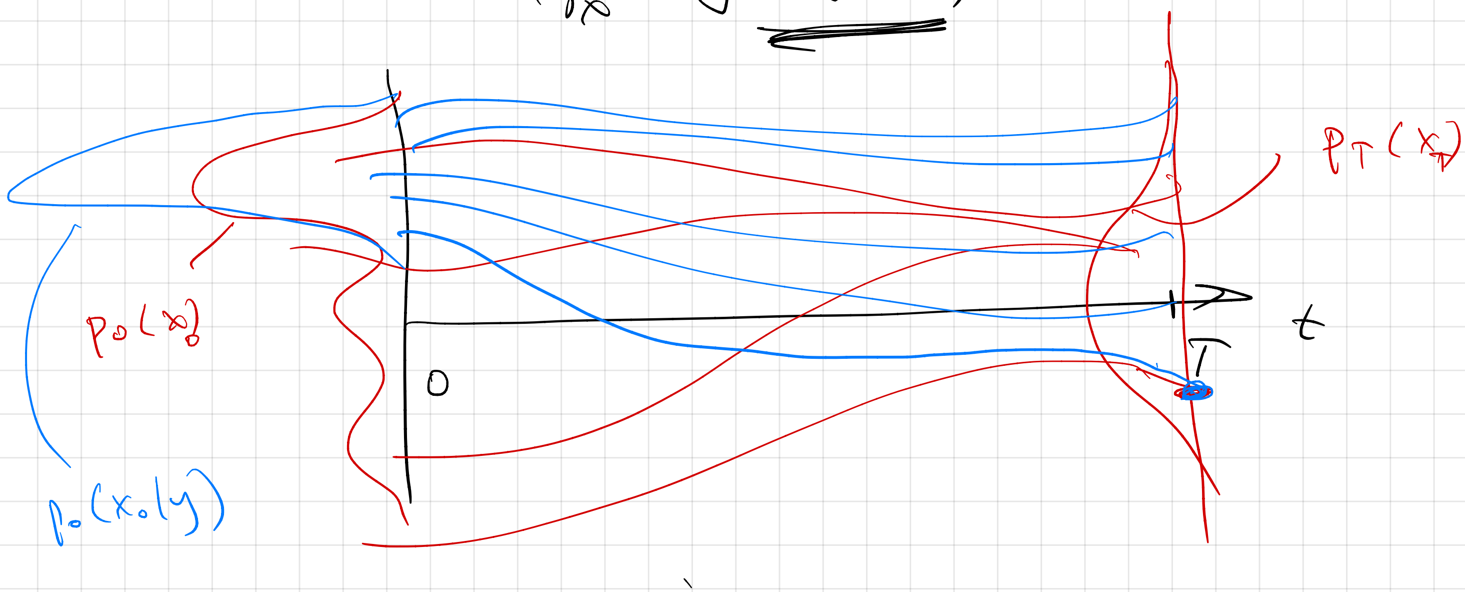
$$dx = \left(f(x,t) - \frac{g^2(t)}{2} \left(\frac{\partial}{\partial x} \log p_t(x) \right) \right) dt$$

$$\Updownarrow$$

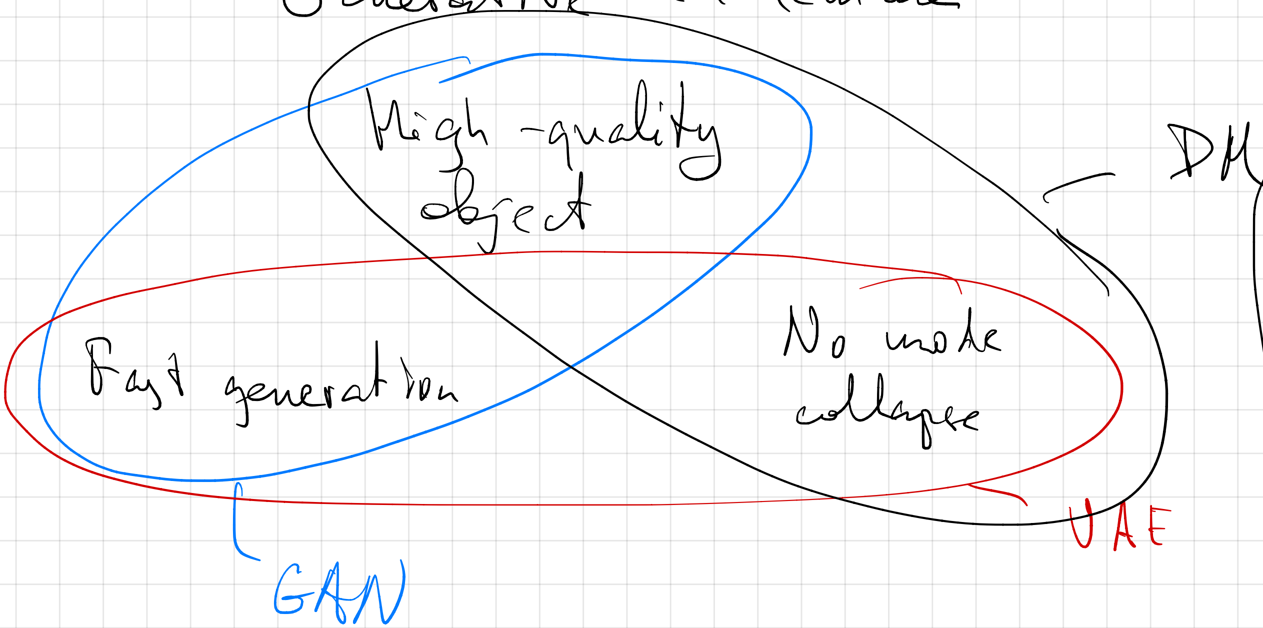
$$\frac{\partial}{\partial x} \log p_t(x|y) =$$

$$= \frac{\partial}{\partial x} \log p_t(x) + \frac{\partial}{\partial x} \log p(y|x_t)$$ – classifier guidance

To perform classifier guidance we should be able to $\frac{\partial}{\partial x} \log p(y|x_t)$ $\forall t$



Generative trilemma



Requirements for DM

- $q(x_t|x_{t-1})$
- $q(x_t|x_0)$