FYS5429/9429 MAY 8

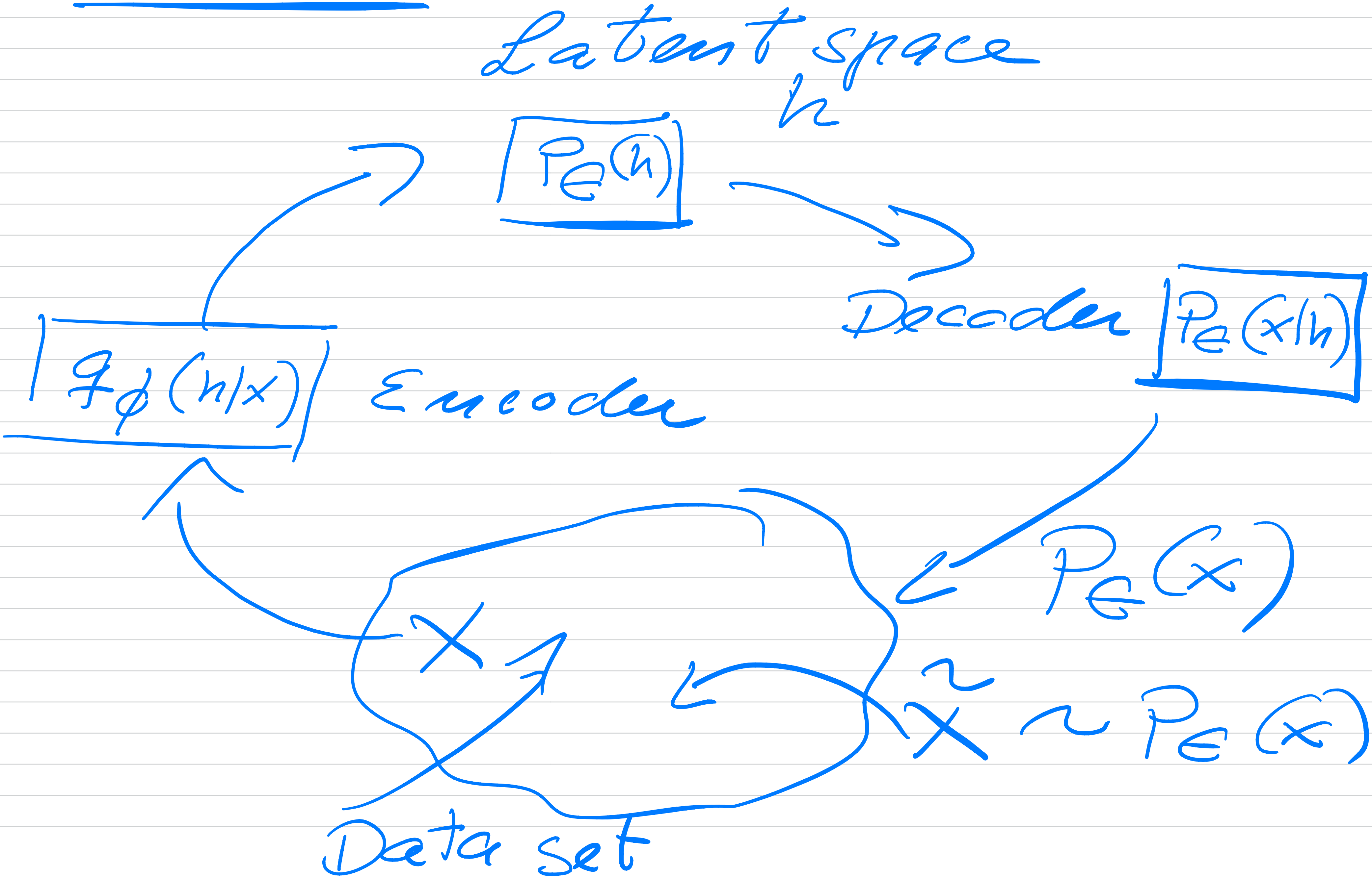
$$p(x) = \int p(x,h)\,dh$$

$$p(x) \rightarrow p_\theta(x)$$

$$\hat{\theta} = \arg\max_\theta \log p_\theta(x)$$

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

VAE



Latent space
h
$p_\theta(h)$
Decoder
$p_\theta(x|h)$
$q_\phi(h|x)$
Encoder
$p_\theta(x)$
x
$\tilde{x} \sim p_\theta(x)$
Data set

$$\log p_\theta(x) = \log p_\theta \times 1$$

$$1 = \int q_\phi(h|x)\, dh$$
$$= 1$$

$$\log p_\theta(x) = \int q_\phi(h|x)(\log p_\theta(x))$$
$$\times\, dh$$

use product rule

$$p_\theta(x) \cdot p_\theta(h|x) = p_\theta(x,h)$$

$$= \mathbb{E}_{h \sim q_\phi(h|x)} \left[\log \frac{p_\theta(x,h)}{q_\phi(h|x)} \right]$$

$$+ \mathbb{E}_{h \sim q_\phi(h|x)} \left[\log \frac{q_\phi(h|x)}{p_\theta(h|x)} \right]$$

$$\geq 0$$

$$q_\phi(h|x) \sim \mathcal{N}(h; \mu_\phi(x), \sigma_\phi^2(x))$$

$$p_\theta(h) \sim \mathcal{N}(h; 0, 1)$$

Brief reminder on Markov chains:

$$P_i(t) = \sum_j T_{ij} P_j(t-1)$$

all possible states

T_{ij} is an element of a matrix T, which is a stochastic matrix

$$\sum_j T_{ij} = 1 \qquad \max \lambda(T) = 1$$

$$\sum_i P_i(t) = 1$$

$$t = t_0$$

$$\vec{p}(t_0) = \begin{bmatrix} P_0(t_0) \\ P_1(t_0) \\ \vdots \\ P_{n-1}(t_0) \end{bmatrix}$$

$$P_i(t_0+1) = \sum_j \overline{T}_{ij} P_j(t_0)$$

as matrices & vectors

$$\vec{p}(t_0+1) = \overline{T}\vec{p}(t_0)$$

time t_0 goes to infinity

$$t_0 \Rightarrow t$$

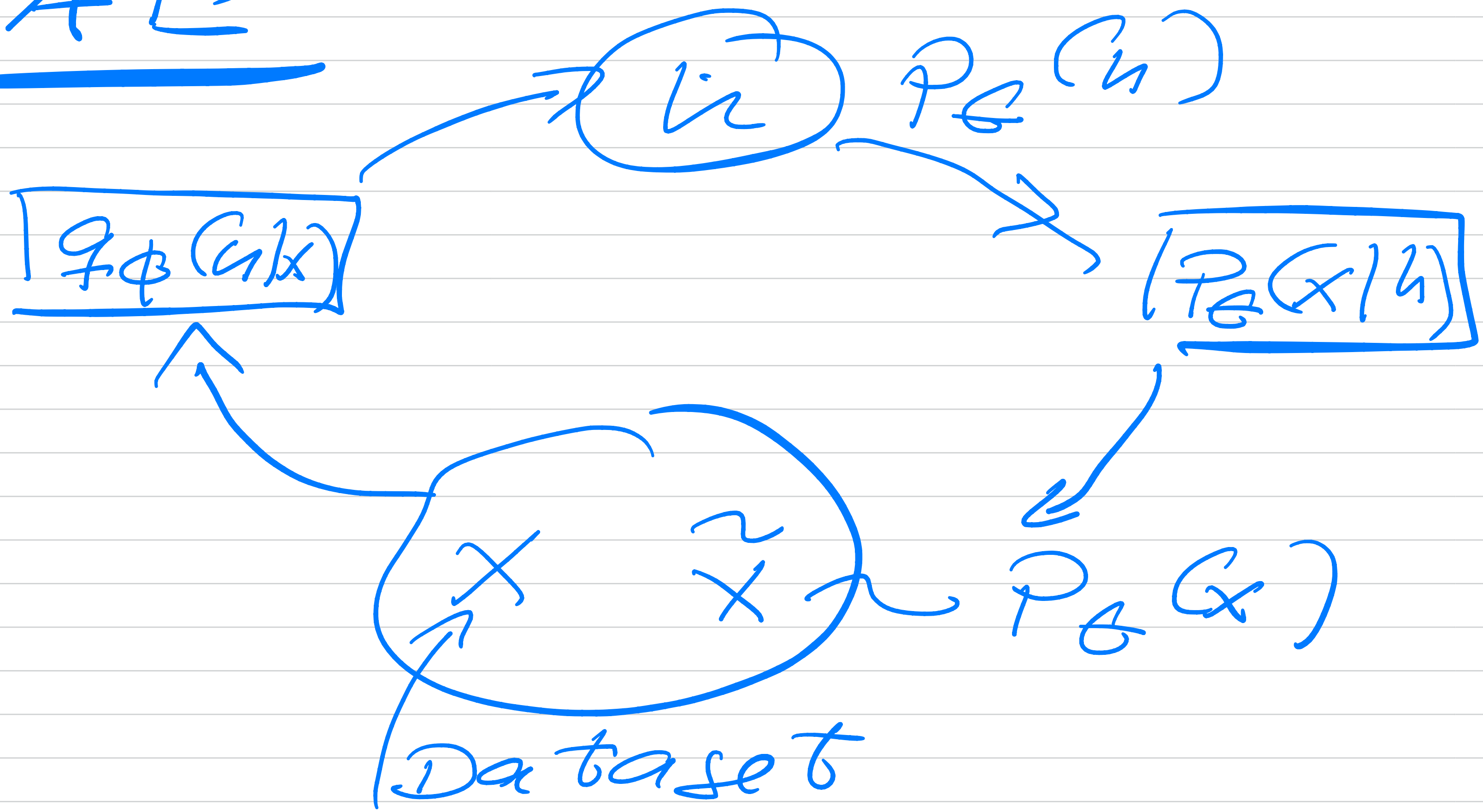
$$\vec{p}(t) = \overline{T}\vec{p}(t) \Rightarrow$$

$$\vec{p} = \overline{T}\vec{p}$$

$$\vec{p}(t_m) = T^m \vec{p}(t_0)$$

$$\begin{array}{ccccccc} p(t_0) & & p(t_1) & & p(t_2) & & p(t_m) \\ t_0 & \xrightarrow[T(t_0 \to t_1)]{} & t_1 & \xrightarrow[T(t_1 \to t_2)]{} & t_2 & \cdots \xrightarrow[T(t_{m-1} \to t_m)]{} & t_m \end{array}$$

VAE

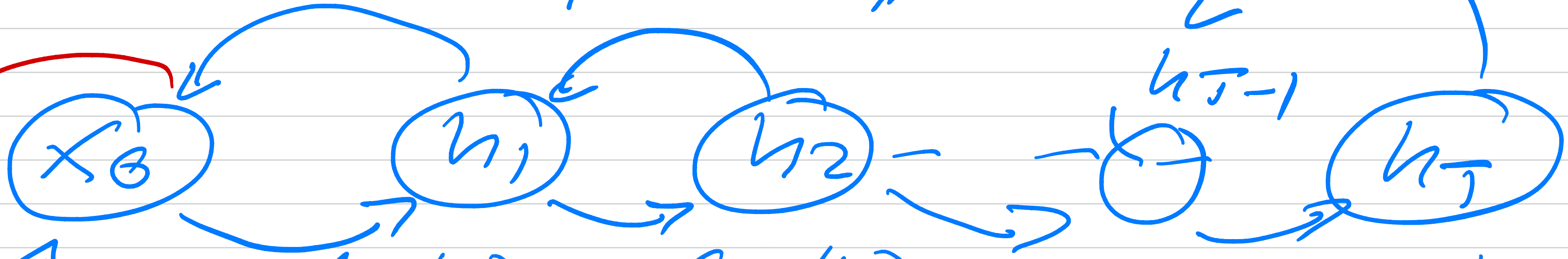


$$p_\theta(x) = \int dz\, p_\theta(x|z)\, p_\theta(z)$$

$$= \int dz\, p_\theta(x, z)$$

Diffusion model

$p(x_0|h_1)\ p(h_1|h_2)\ \sim\ p(h_{T-1}|h_T)$



$q(h_1|x_0)\ q(h_2|h_1)\ \ q(h_T|h_{T-1})$

$p(x_0)$

$\underline{p_\theta(x_0)}$

generates $\tilde{x}_0$

$$p(x, h_{1:T}) = p(h_T)\, p_\theta(x \mid h_1) \times \prod_{t=2}^{T} p_\theta(h_{t-1} \mid h_t)$$

$$q_\phi(h_{1:T}) = q_\phi(h_1 \mid x) \times \prod_{t=2}^{T} q(h_t \mid h_{t-1})$$

Final marginal probability

$$\log p(x) = \log\left(\int p(x, h_{1:T}) \times dh_{1:T}\right)$$

$h_1, h_2, \ldots h_T \rightarrow x_1\, x_2 \ldots x_T$

x_0 is the input data

x_0 x_1 x_2 — — — x_T

we want a model $p(x_0)$

$$\tilde{x}_0 \sim p(x_0) \quad (\text{skip } \phi, \theta)$$

$$q_\phi(\underset{z_{t+1}}{x_{t+1}} \mid \underset{z_t}{x_t})$$

$$\log p(x_0) = \log \int p(x_{0:T}) \times \underbrace{dx_{1:T}}_{\text{latent variables}}$$

$$= \log \int \frac{p(x_{0:T})\, q(x_{1:T}|x_0)}{q(x_{1:T}|x_0)}\, dx_{1:T}$$

$$= 1$$

$$= \log \mathbb{E}_{q(x_{1:T}|x_0)}\left[\frac{p(x_{0:T})}{q(x_{1:T}|x_0)} \right]$$

$$= \log \mathbb{E}_{q(x_{1:T}|x_0)}\left[\frac{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T} q(x_t|x_{t-1})}\right]$$

$$=$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log\left(\frac{p(x_T)\,p_\theta(x_0|x_1)\prod_{t=2}^{T} p_\theta(x_{t-1}|x_t)}{q(x_T|x_{T-1})\prod_{t=1}^{T} q(x_t|x_{t-1})}\right)\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log\left(\frac{p(x_T)\,p_\theta(x_0|x_1)\prod_{t=1}^{T} p_\theta(x_t|x_{t+1})}{q(x_T|x_{T-1})\prod_{t=1}^{T} q(x_t|x_{t-1})}\right)\right]$$

$$= \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_0|x_1)\right]$$

$$+ \mathbb{E}_{q(x_{1:T}|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_{T-1})}\right]$$

$$+ \mathbb{E}_{q(x_{1:T}|x_0)}\left[\sum_{t=1}^{T} \log\left(\frac{p_\theta(x_t|x_{t+1})}{q(x_t|x_{t-1})}\right)\right] \geq 0$$

$$= \underbrace{\mathbb{E}_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right]}_{\text{reconstruction}}$$

$$+ \mathbb{E}_{q(x_{T-1}, x_T|x_0)}\left[\log \frac{p(x_T)}{q(x_T|x_{T-1})}\right]$$

$$+ \sum_{t=1}^{T-1} \mathbb{E}_{q(x_{t-1}, x_t, x_{t+1}|x_0)}\left[\log \frac{p_\theta(x_t|x_{t+1})}{q(x_t|x_{t-1})}\right]$$