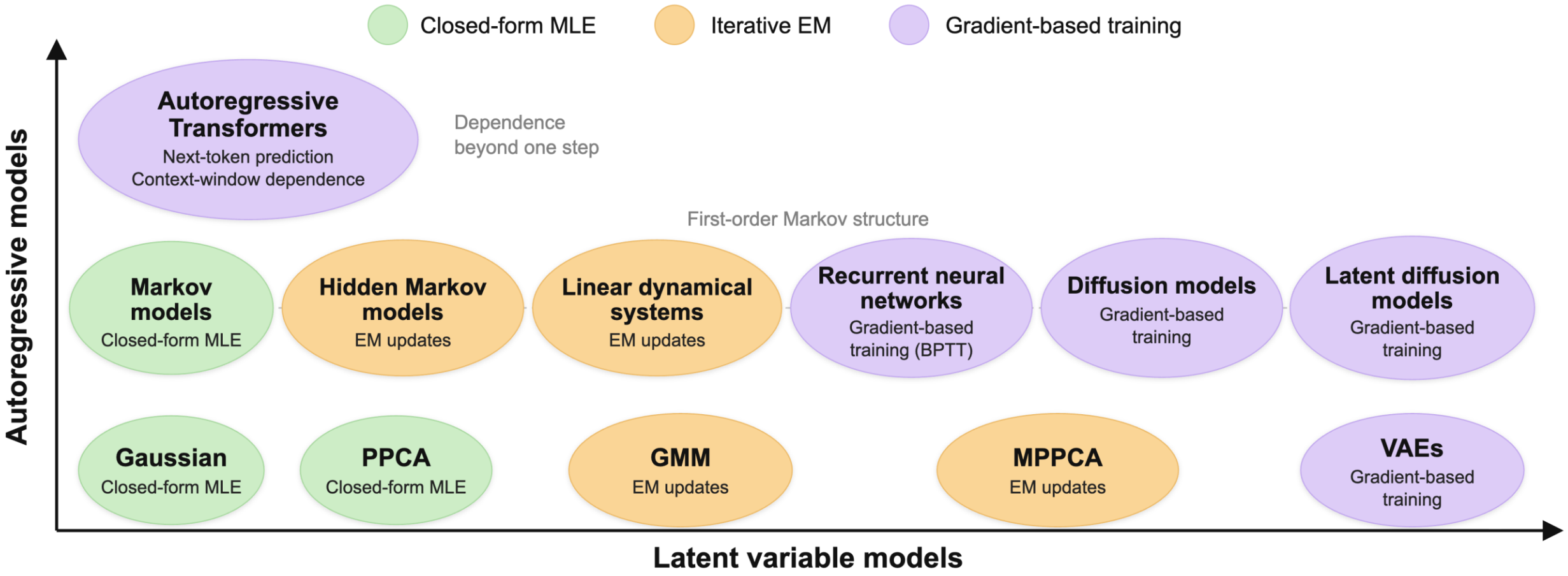
Deep Generative Models: Diffusion Models

Fall Semester 2026

René Vidal
Director of the Center for Innovation in Data Engineering and Science (IDEAS)
Rachleff University Professor, University of Pennsylvania
Amazon Scholar & Chief Scientist at NORCE

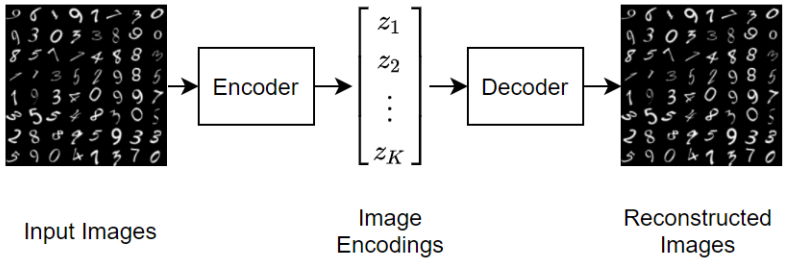


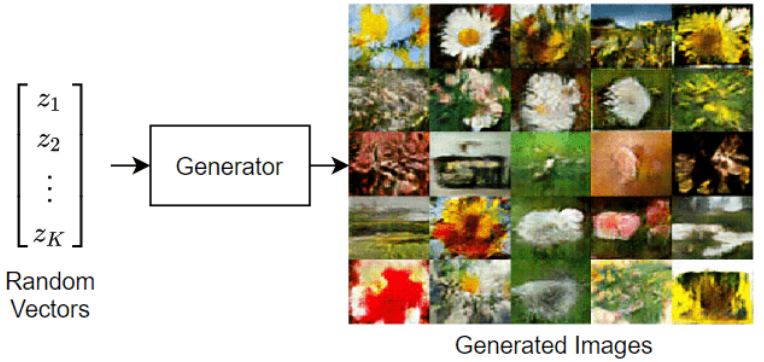
Taxonomy of Generative Models



Diffusion Models

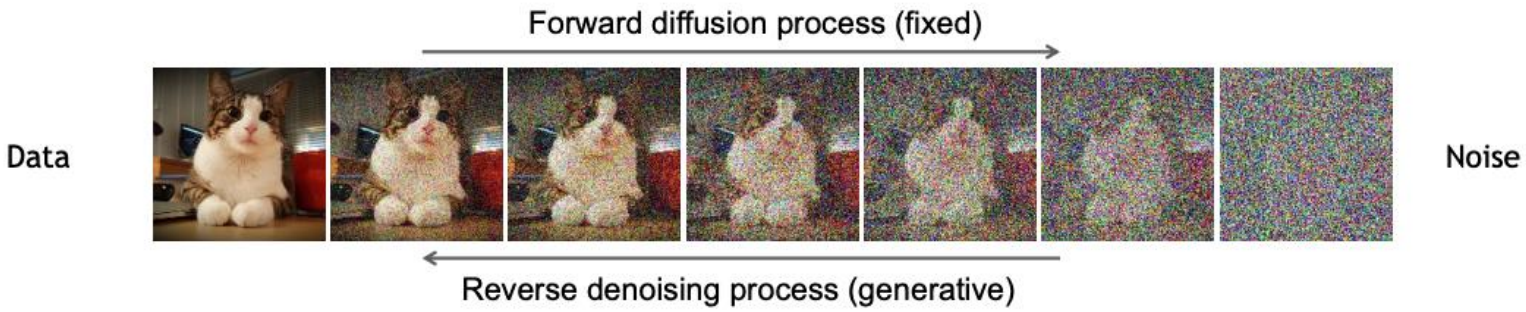
- The journey of generative models has evolved significantly in recent years.
- **Variational Autoencoders (VAEs)** introduce probabilistic modeling for latent representations but struggled with generating high-quality images.
- This led to the rise of **Generative Adversarial Networks (GANs)**, which leverage adversarial learning to produce high-quality, realistic outputs but suffered from issues like mode collapse and unstable training.
- The introduction of **Diffusion Models** achieve state-of-the-art results with superior stability and diversity in generated samples, particularly in multimodal image synthesis.

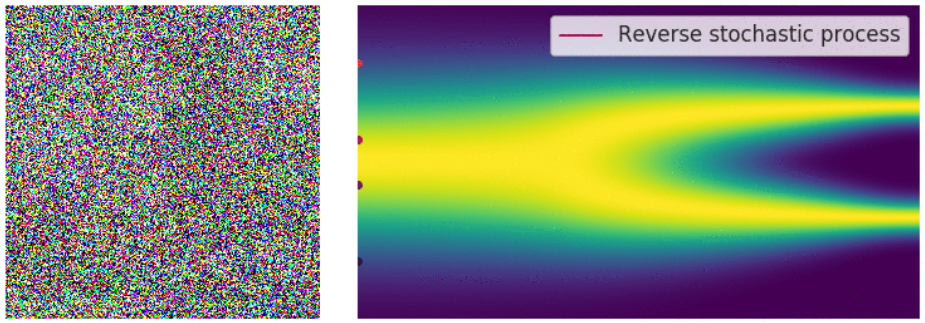




Diffusion Models

- A (Latent) Diffusion Model is a VAE with an autoregressive latent space.
- The diffusion model encoder (a.k.a. the forward diffusion process) maps data to noise by gradually adding Gaussian noise to the input data.
- The diffusion model decoder (a.k.a. the reverse diffusion process) maps noise to data by learning a transformation that aims to reverse the forward diffusion process.





Outline

- **Markov Hierarchical Variational Auto Encoders (MHVAEs)**
  - Autoregressive Encoder and Autoregressive Decoder of an MHVAE
  - Derivation of the ELBO of an MHVAE
- Diffusion Models: MHVAEs with a Linear Gaussian Autoregressive Latent Space
  - Forward Diffusion Process
  - Reverse Diffusion Process
  - ELBO for Diffusion Models as a particular case of the ELBO for MHVAEs
- Implementation Details: UNet architecture, Training and Sampling Strategies
- Application of Diffusion Models
  - Text-Conditioned Diffusion Model: Stable Diffusion
  - Multimodal Control for Consistent Synthesis: ControlNet
  - Image Editing: DDIM, P2P

Recall the Variational Autoencoder (VAE)



Datapoint $\boldsymbol{x}$

Encoder
$q_\phi(\boldsymbol{z} \mid \boldsymbol{x})$

$\boldsymbol{\mu}$

$\log \sigma^2$

Sampler
$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{\mu}_\phi(\boldsymbol{x}), \sigma^2_\phi(\boldsymbol{x})\boldsymbol{I})$

Decoder
$p_\theta(\boldsymbol{x} \mid \boldsymbol{z})$

$\boldsymbol{\mu}_\theta(\boldsymbol{z})$

ELBO Objective

$$\mathbb{E}_{\boldsymbol{z} \sim q_\psi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z})] - \mathrm{KL}[q_\psi(\boldsymbol{z} \mid \boldsymbol{x}) || p(\boldsymbol{z})]$$

Recall the Evidence Lower Bound (ELBO)

- The ELBO is the sum of a reconstruction term and a prior matching term

$$\log p_\theta(\boldsymbol{x}) \geq \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]$$

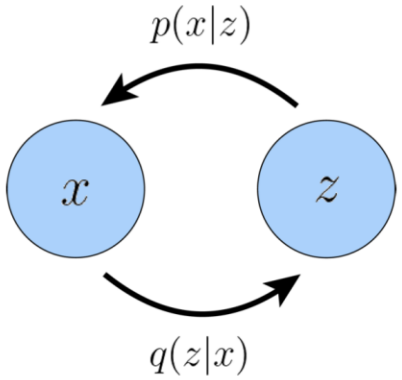
$$= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]$$

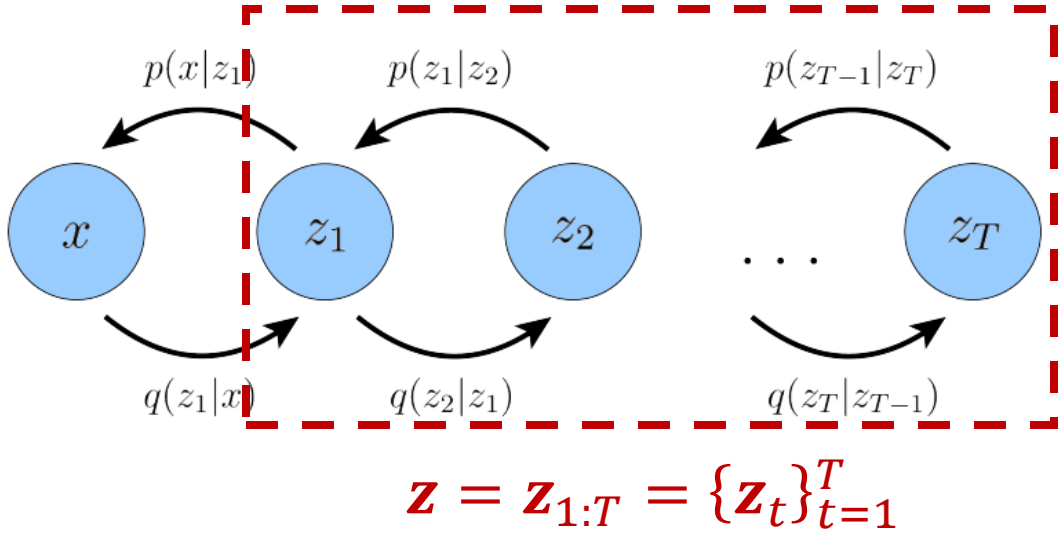
$$= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z})] + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]$$

$$= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z})]}_{\text{reconstruction term}} - \underbrace{\mathrm{D}_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z} \mid \boldsymbol{x}) \,||\, p(\boldsymbol{z})\right)}_{\text{prior matching term}}$$

Latent Diffusion Models as "Autoregressive VAEs"

- A Latent Diffusion Model is as a Markovian Hierarchical Variational Autoencoder (MHVAE) with T hierarchical latents $\boldsymbol{z} = \boldsymbol{z}_{1:T} = \{\boldsymbol{z}_t\}_{t=1}^{T}$ modeled by a Markov chain where each latent $\boldsymbol{z}_t$ is generated only from the previous latent $\boldsymbol{z}_{t+1}$.



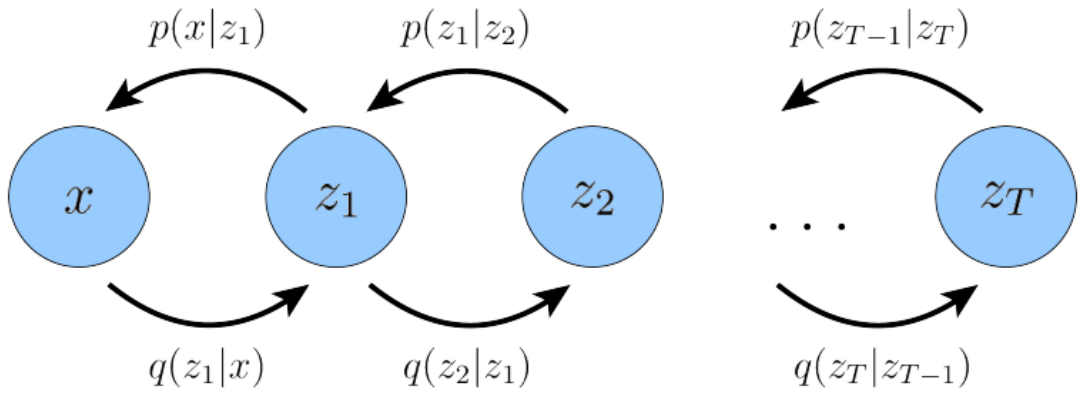


$$\boldsymbol{z} = \boldsymbol{z}_{1:T} = \{\boldsymbol{z}_t\}_{t=1}^{T}$$

- What is the VAE encoder $q_\phi(\boldsymbol{z} \mid \boldsymbol{x})$ of a Diffusion Model ?
- What is the VAE decoder $p_\theta(\boldsymbol{x} \mid \boldsymbol{z})$ of a Diffusion Model ?
- What is the ELBO of a Diffusion Model ?

MHVAE Encoder, Decoder, and ELBO

- A MHVAE is a VAE whose encoder and decoder are autoregressive models:



$$p_\theta(\boldsymbol{x}, \boldsymbol{z}_{1:T}) = p_\theta(\boldsymbol{z}_T) p_\theta(\boldsymbol{x} \mid \boldsymbol{z}_1) \prod_{t=2}^{T} p_\theta(\boldsymbol{z}_{t-1} \mid \boldsymbol{z}_t)$$

$$q_\phi(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}) = q_\phi(\boldsymbol{z}_1 \mid \boldsymbol{x}) \prod_{t=2}^{T} q_\phi(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})$$

- Given this joint distribution and posterior, we can rewrite the ELBO for MHVAE as:

$$\mathbb{E}_{q_\phi(\boldsymbol{z}_{1:T}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{x}, \boldsymbol{z}_{1:T})}{q_\phi(\boldsymbol{z}_{1:T} \mid \boldsymbol{x})}\right] = \mathbb{E}_{q_\phi(\boldsymbol{z}_{1:T}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{z}_T) p_\theta(\boldsymbol{x} \mid \boldsymbol{z}_1) \prod_{t=2}^{T} p_\theta(\boldsymbol{z}_{t-1} \mid \boldsymbol{z}_t)}{q_\phi(\boldsymbol{z}_1 \mid \boldsymbol{x}) \prod_{t=2}^{T} q_\phi(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})}\right]$$

Decomposition of the ELBO for an MHVAE

- Let us make the change of variables $\boldsymbol{x} \rightarrow \boldsymbol{x}_0$ and $\boldsymbol{z}_{1:T} \rightarrow \boldsymbol{x}_{1:T}$.
- The ELBO is hard to evaluate because it requires sampling from $q_\phi(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)$.

- **Theorem**: The ELBO for a MHVAE can be written as

$$\mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\prod_{t=2}^{T} p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)\prod_{t=2}^{T} q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})}\right] =$$

$$\underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)]}_{\text{reconstruction term}} - \underbrace{D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_T)\right)}_{\text{prior matching term}}$$

$$-\sum_{t=2}^{T} \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)\right)\right]}_{\text{score matching term}}$$

Decomposition of the ELBO for an MHVAE

- **Proof (1/2):** Reversing $q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$

$$q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) = q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{x}_0) = \frac{q_\phi(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}{q_\phi(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}.$$

- Substituting $q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ and using telescopic product to cancel factors

$$\log p(\boldsymbol{x}) \geq \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)} \left[\log \frac{p_\theta(\boldsymbol{x}_T) p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) \prod_{t=2}^{T} p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0) \prod_{t=2}^{T} q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})} \right]$$

$$= \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)} \left[\log \frac{p_\theta(\boldsymbol{x}_T) p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)}{q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)} \prod_{t=2}^{T} \frac{p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{\frac{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) q_\phi(\boldsymbol{x}_t \mid \boldsymbol{x}_0)}{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_0)}} \right]$$

$$= \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)} \left[\log \frac{p_\theta(\boldsymbol{x}_T) p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1) q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)}{q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0) q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)} \prod_{t=2}^{T} \frac{p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0)} \right]$$

Decomposition of the ELBO for an MHVAE

- **Proof (2/2):** expanding into three terms and simplifying expectations

$$\log p(\boldsymbol{x}) \geq \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)} \prod_{t=2}^{T} \frac{p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0)}\right]$$

$$= \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\right] + \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right] + \sum_{t=2}^{T} \mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0)}\right]$$

$$= \mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\right] + \mathbb{E}_{q_\phi(\boldsymbol{x}_T|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right] + \sum_{t=2}^{T} \mathbb{E}_{q_\phi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0)}\right]$$

$$= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)\right]}_{\text{reconstruction term}} - \underbrace{D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_T)\right)}_{\text{prior matching term}} - \sum_{t=2}^{T} \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)\right)\right]}_{\text{score matching term}}$$

Why can we Simplify Expectations?

- **For the first term:**

$$\mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}[\log(p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1))] = \int \log(p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1))\, q_\phi(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{1:T}$$

$$= \int \log(p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1))\, q_\phi(\boldsymbol{x}_1, \boldsymbol{x}_{2:T} \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{2:T} d\boldsymbol{x}_1$$

$$\int q_\phi(\boldsymbol{x}_1, \boldsymbol{x}_{2:T} \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{2:T} = q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)$$

$$= \int \log(p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1))\, q_\phi(\boldsymbol{x}_1 \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_1 = \mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}[\log(p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1))]$$

- **For the second term:**

$$\mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right] = \int \log\left(\frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right) q_\phi(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{1:T}$$

$$= \int \log\left(\frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right) q_\phi(\boldsymbol{x}_{1:T-1}, \boldsymbol{x}_T \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{1:T-1} d\boldsymbol{x}_T$$

$$\int q_\phi(\boldsymbol{x}_{1:T-1}, \boldsymbol{x}_T \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_{1:T-1} = q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)$$

$$= \int \log\left(\frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right) q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)\, d\boldsymbol{x}_T = \mathbb{E}_{q_\phi(\boldsymbol{x}_T|\boldsymbol{x}_0)}\left[\log \frac{p_\theta(\boldsymbol{x}_T)}{q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0)}\right]$$

Why can we Simplify Expectations?

- **For the third term:**

$$\mathbb{E}_{q_\phi(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log\left(\frac{p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)}\right)\right]=\int\log\left(\frac{p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)}\right)q_\phi(\boldsymbol{x}_{1:T}\mid\boldsymbol{x}_0)d\boldsymbol{x}_{1:T}$$

$$=\int\log\left(\frac{p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)}\right)q_\phi(\boldsymbol{x}_{1:t-2},\boldsymbol{x}_{t-1},\boldsymbol{x}_t,\boldsymbol{x}_{t+1:T}\mid\boldsymbol{x}_0)d\boldsymbol{x}_{1:t-2}d\boldsymbol{x}_{t+1:T}d\boldsymbol{x}_{t-1}d\boldsymbol{x}_t$$

$$=\int\log\left(\frac{p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)}\right)q_\phi(\boldsymbol{x}_{t-1},\boldsymbol{x}_t\mid\boldsymbol{x}_0)d\boldsymbol{x}_{t-1}d\boldsymbol{x}_t$$

$$=\int\log\left(\frac{p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)}{q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)}\right)q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)q_\phi(\boldsymbol{x}_t\mid\boldsymbol{x}_0)d\boldsymbol{x}_{t-1}d\boldsymbol{x}_t$$

$$=-\int D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)\,\|\,p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)\right)q_\phi(\boldsymbol{x}_t\mid\boldsymbol{x}_0)d\boldsymbol{x}_t$$

$$=-\mathbb{E}_{q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t,\boldsymbol{x}_0)\,\|\,p_\theta(\boldsymbol{x}_{t-1}\mid\boldsymbol{x}_t)\right)\right]$$

$$\int q_\phi(\boldsymbol{x}_{1:t-2},\boldsymbol{x}_{t-1:t},\boldsymbol{x}_{t+1:T}\mid\boldsymbol{x}_0)d\boldsymbol{x}_{1:t-2}d\boldsymbol{x}_{t+1:T}=q_\phi(\boldsymbol{x}_{t-1:t}\mid\boldsymbol{x}_0)$$

Interpretation of the ELBO of an MHVAE

$$\log p(\boldsymbol{x}) \geq \mathcal{L}_{\text{ELBO}}(\theta, \phi)$$

$$= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)]}_{\text{reconstruction term}} - \underbrace{D_{\text{KL}}\left(q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_T)\right)}_{\text{prior matching term}} - \sum_{t=2}^{T} \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\text{KL}}\left(q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)\right)\right]}_{\text{score matching term}}$$

- $\mathbb{E}_{q_\phi(\boldsymbol{x}_1|\boldsymbol{x}_0)}[\log p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{x}_1)]$ can be interpreted as a reconstruction term; like its analogue in the ELBO of a vanilla VAE. This term can be approximated and optimized using a Monte Carlo estimate.
- $D_{\text{KL}}\left(q_\phi(\boldsymbol{x}_T \mid \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_T)\right)$ represents how close the distribution of the final latent distribution is to the standard Gaussian prior.
- $\mathbb{E}_{q_\phi(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\text{KL}}\left(q_\phi(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0) \,||\, p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)\right)\right]$ is a score matching term.