Towards diffusion: (Markovian) hierarchical VAEs

Diffusion models can be seen as hierarchical VAEs with a few restrictions:

- The forward *(encoding)* distribution prescribed as a Markov chain of Gaussians; it is not learned

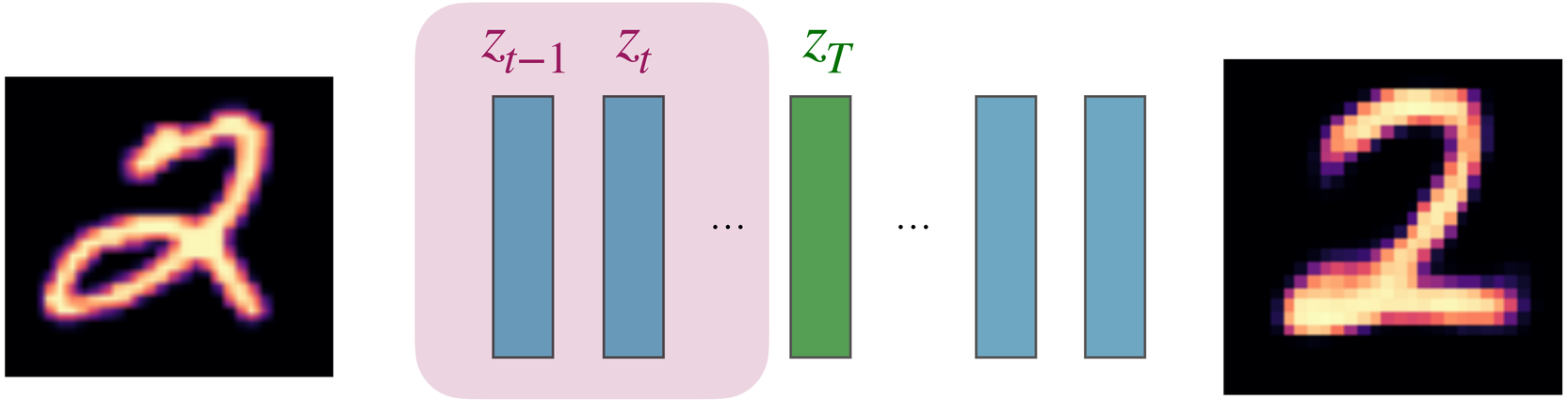
$$q\left(z_t \mid z_{t-1}\right) = \mathcal{N}\left(z_t; \alpha_t z_{t-1}, \beta_t\right)$$

- Distributions of latents at the final timestep T is a standard (unit) Gaussian

$$q(z_T \mid z_{T-1}, \ldots x) = \mathcal{N}(z_T; 0, \mathbb{I})$$

- The dimensionality of latents is the same as the data dimensionality

$$\dim(z_t) = \dim(x)$$



Variational diffusion models

[Kingma et al 2021]

[Gory details: Luo 2022; 2208.11970]

Align the forward and reverse distributions;
variational lower bound (ELBO) as before

$$L = \left\langle \log \frac{q\left(x, z_1, z_2, \cdots, z_T\right)}{p\left(x, z_1, z_2, \cdots, z_T\right)} \right\rangle_{q(x)}$$