1 Maths

The *noising process* $q(\mathbf{z}_t \mid \mathbf{z}_{t-1}, \mathbf{x})$ is a Markov process which iteratively adds Gaussian noise to data $\mathbf{x}$ until the signal is destroyed, $\mathbf{z}_T \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$.

We learn a Gaussian *generative denoising process* p_θ that aims to approximate the *true denoising process* $q(\mathbf{z}_{t-1} \mid \mathbf{z}_t, \mathbf{x})$ without knowledge of the data $\mathbf{x}$. Cross-entropy loss (often called NLL in the literature) is intractible, so a variational bound is used:

$$\mathbb{E}\left(-\log p_\theta(\mathbf{x})\right) \leq \mathbb{E}\left(-p_\theta(\mathbf{x} \mid \mathbf{z}_0)\right) + \sum_{t=1}^{T} \mathbb{E}\left[\mathrm{KL}\left(q(\mathbf{z}_{t-1} \mid \mathbf{z}_t, \mathbf{x}) \mid\mid p_\theta(\mathbf{z}_{t-1} \mid \mathbf{z}_t)\right)\right]$$

Parameterisation The noising process has a marginal distribution $\mathbf{z}_t \overset{\text{dist.}}{=} \alpha_t \mathbf{x} + \beta_t \boxed{\boldsymbol{\varepsilon}}$ where $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$.

$$q(\mathbf{z}_{t-1} \mid \mathbf{z}_t, \boxed{\mathbf{x}})$$

$$p_\theta(\mathbf{z}_{t-1} \mid \mathbf{z}_t) \coloneqq q(\mathbf{z}_{t-1} \mid \mathbf{z}_t, \boxed{\hat{\mathbf{x}}_\theta(\mathbf{z}_t, t)}) \qquad \frac{1}{\alpha_t}\mathbf{z}_t - \frac{\beta_t}{\alpha_t}\boxed{\hat{\boldsymbol{\varepsilon}}_\theta(\mathbf{z}_t, t)}$$

$$\min \mathbb{E}_{t,\mathbf{x},\mathbf{z}_t}\left(\left\|\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}_\theta(\mathbf{z}_t, t)\right\|^2\right)$$