The possibility of skipping steps in the generative process relies entirely on the forward process and its connection to the ELBO.

Given q as in "Readme", we choose to define

$$p_\theta^{(t)}(x_{t-1}|x_t) = \begin{cases} \mathcal{N}(f_\theta^{(1)}(x_1), \sigma_1^2 I) & t=1 \\ q(x_{t-1}|x_t, f_\theta^{(t)}(x_t)) & \text{otherwise} \end{cases}$$

$$f_\theta^{(t)}(x_t) = \frac{x_t - \sqrt{1-\bar{\alpha}_t} \cdot \epsilon_\theta^{(t)}(x_t)}{\sqrt{\bar{\alpha}_t}}$$

Then, $L(x_0) = \mathbb{E}_{x_{0:T} \sim q(x_{0:T})}\Big[\log q(x_T|x_0)$

$+ \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta^{(t)}(x_{t-1}|x_t))$

$- \log p_\theta^{(1)}(x_0|x_1)\Big]$

$= \mathbb{E}[\log q(x_T|x_0)] + \mathbb{E}_{x_0, x_t \sim q(x_0, x_t)}$

$[D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta^{(t)}(x_{t-1}, x_t))]$

$$= C + \sum_{t=1}^{T} \frac{1}{2d\sigma_t^2 \alpha_t} \mathbb{E}\left[\| \varepsilon_\theta^{(t)}(x_t) - \varepsilon_t \|_2^2 \right]$$

$$= L_\gamma(\varepsilon_\theta) + C.$$

$$\gamma_t = \frac{1}{2d\sigma_t^2 \alpha_t}$$

d is the dimension of x_0

And σ_t is as in "Readme".

Next, we also motivate the form of $q(x_{t-1}|x_t, x_0)$ by matching the marginals. Assume

$$q(x_t|x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)$$

for $t \geq k$, for some $k \in \{1, \dots, T\}$

By induction we want to show

$$q(x_{k-1}|x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_{k-1}}\, x_0, (1-\bar{\alpha}_{k-1}) I).$$

The base case $t = T$ is verified by our assumption.

$$q(x_{k-1}|x_0) = \int dx_k\, q(x_k|x_0)\, q(x_{k-1}|x_k, x_0)$$

The product of two Gaussians is still Gaussian with some mean and variance. Let's find them by writing the inside of the exp.

$$-\frac{(x_k-\sqrt{\bar{\alpha}}\,x_0)^2}{2(1-\bar{\alpha}_k)}-\frac{\left(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0-\sqrt{1-\bar{\alpha}_{k-1}-\sigma_k^2}\cdot\frac{(x_k-\sqrt{\bar{\alpha}}\,x_0)}{\sqrt{1-\bar{\alpha}_k}}\right)^2}{2\sigma_k^2}$$

$$=\cancel{\frac{-x_k^2+2\sqrt{\bar{\alpha}_k}\,x_kx_0-\bar{\alpha}_kx_0^2}{2(1-\bar{\alpha}_k)}}$$

$$\frac{-(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0)^2+2(x_{k+1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0)\sqrt{(1-\bar{\alpha}_{k-1}-\sigma_k^2)}\;\frac{x_k-\sqrt{\bar{\alpha}_k}\,x_0}{\sqrt{1-\bar{\alpha}_k}}-\frac{(1-\bar{\alpha}_{k-1}\cancel{-\sigma_k^2})(x_k-\sqrt{\bar{\alpha}_k}\,x_0)^2}{1-\bar{\alpha}_k}}{2\sigma_k^2}$$

$$=-\frac{(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0)^2}{2\sigma_k^2}$$

$$+\,2\left(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0\right)\sqrt{1-\bar{\alpha}_{k-1}-\sigma_k^2}\;\frac{\left(x_k-\sqrt{\bar{\alpha}_k}\,x_0\right)}{\sqrt{1-\bar{\alpha}_k}}$$

$$\frac{-\left(x_k-\sqrt{\bar{\alpha}_k}\,x_0\right)^2\dfrac{(1-\bar{\alpha}_{k-1})}{1-\bar{\alpha}_k}}{2\sigma_k^2}$$

↳ separate the terms that depend on the x_k integration

$$=\frac{\begin{aligned}&-2\left(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0\right)\sqrt{1-\bar{\alpha}_{k-1}-\sigma_k^2}\;\frac{\sqrt{\bar{\alpha}_k}\,x_0}{\sqrt{1-\bar{\alpha}_k}}\\&+\frac{2\left(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}\,x_0\right)\sqrt{1-\bar{\alpha}_{k-1}-\sigma_k^2}\;\underline{x_k}}{\sqrt{1-\bar{\alpha}_k}}\\&-\bar{\alpha}_k x_0^2\,\frac{(1-\bar{\alpha}_{k-1})}{1-\bar{\alpha}_k}\;-\frac{\left(\underline{x_k^2}-2\sqrt{\bar{\alpha}_k}\,\underline{x_k}\,x_0\right)}{\dfrac{(1-\bar{\alpha}_k)}{(1-\bar{\alpha}_{k-1})}}\end{aligned}}{2\sigma_k^2}$$

Combining just the x_k terms and define $\alpha := \dfrac{1-\bar{\alpha}_{k-1}}{1-\bar{\alpha}_k}$

$$-\alpha\left(x_K^2 - 2x_K\left(\sqrt{\bar\alpha_K}\,x_0 + \frac{\sqrt{1-\bar\alpha_{K-1}-\sigma_K^2}\,(x_{K-1}-\sqrt{\bar\alpha_{K-1}}\,x_0)}{\alpha\cdot\sqrt{1-\bar\alpha_K}}\right)\right)$$

$$= -\alpha\left(x_K^2 - 2x_K b + b^2 - b^2\right)$$

where I complete the square, and write b as the coefficient of x_K.

$$= -\alpha(x_K - b)^2 + \alpha b^2$$

Exponentials detach in the integral, while the term $-\alpha(x_K - b)^2$ integrates to $\frac{1}{\sqrt{\alpha}}$ because is a Gaussian.

Collecting all the remaining terms:

$$\cancel{-\bar\alpha_K x_0^2}\,\alpha + \alpha_K \cancel{x_0^2}\,\alpha + 2\,\cancel{(x_{K-1}-\sqrt{\bar\alpha_{K-1}}\,x_0)}$$

$$\sqrt{1-\bar\alpha_{K-1}-\sigma_K^2}\cdot\cancel{\sqrt{\bar\alpha_K}\,x_0}\cdot\frac{1}{\sqrt{1-\bar\alpha_K}}$$

$$+\frac{(1-\cancel{\bar{\alpha}_{k-1}}-\sigma_k^2)(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}x_0)^2}{2(1-\bar{\alpha}_k)}$$

$$\cancel{-2(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}x_0)\sqrt{1-\bar{\alpha}_{k-1}-\sigma_k^2}\,\frac{\sqrt{\bar{\alpha}_k}x_0}{\sqrt{1-\bar{\alpha}_k}}}$$

$$\cancel{-(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}x_0)^2}$$

$$=-\frac{1}{2(1-\bar{\alpha}_k)}(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}x_0)^2$$

$$=-\frac{1}{(1-\bar{\alpha}_{k-1})}(x_{k-1}-\sqrt{\bar{\alpha}_{k-1}}x_0)^2$$

So, this is a Gaussian $\mathcal{N}(\sqrt{\bar{\alpha}_{k-1}}x_0, (1-\bar{\alpha}_{k-1})I)$, which completes the induction. When I didn't write the $\frac{1}{\sigma_k^2}$ factors above, I was ignoring them because it is just a common factor.

With all of there, we can write for the case of $q_\tau(x_{1:T}|x_0) = q_{\sigma,\tau}(x_T|x_0)$

$$\prod_{i=1}^{S} q_\tau(x_{\tau_{i-1}}|x_{\tau_i}, x_0) \prod_{t\in\bar{\tau}} q_\tau(x_t|x_0),$$

where τ is an increasing rule-sequence of $\{1,\dots,T\}$ with lengths S and $\tau_S = T$. Let $\bar{\tau} := \{1,\dots T\}\setminus\tau$ be the complement.

Under the assumptions:

$$q_\tau(x_t|x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\,x_0, (1-\bar{\alpha}_t)I),$$

(#) $t\in\{T\}\cup\bar{\tau}$

$$q_\tau(x_{\tau_{i-1}}|x_{\tau_i}, x_0) = \mathcal{N}\Big(\sqrt{\alpha_{\tau_{i-1}}}\,x_0 + \sqrt{1-\bar{\alpha}_{\tau_{i-1}}-\sigma_{\tau_i}^2}\,\frac{x_{\tau_i}-\sqrt{\alpha_\tau}\,x_0}{\sqrt{1-\alpha_{\tau_i}}}, \sigma_{\tau_i}^2 I\Big),$$

(#) $i\in S$.

So, $q_\tau(x_{\tau_i}|x_0)$ match the marginals.

The generative process becomes

$$p_\theta(x_{0:T}) = p_\theta(x_T) \prod_{i=1}^{S} p_\theta^{(\tau_i)}(x_{\tau_{i-1}} | x_{\tau_i})$$

$$\prod_{t \in \bar{\tau}} p_\theta^{(t)}(x_0 | x_t), \text{ where}$$

$$p_\theta^{(\tau_i)}(x_{\tau_{i-1}} | x_{\tau_i}) = q_\tau(x_{\tau_{i-1}} | x_{\tau_i}, f_\theta^{(\tau_i)}(x_{\tau_i}))$$

for $i \in S, i > 1$

$$p_\theta^{(t)}(x_0 | x_t) = \mathcal{N}(f_\theta^{(t)}(x_t), \sigma_t^2 I) \text{ otherwise}$$

Similarly, we can show $\mathcal{L}(x_0)$ reduces to the old ELBO. So, minimizing the old ELBO is equivalent to learning in the background a fast sampling non-Markovian process.