1 Forward process

Adding Noise in T steps using:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t)$$

How can we sample $\boldsymbol{x}_t$ in one step? We use the following definitions and identities:

$$\alpha_t = 1 - \beta_t \ , \ \bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$$

$$\mathcal{N}(x; \mu_1, \sigma_1^2) * \mathcal{N}(x; \mu_2, \sigma_2^2) = \mathcal{N}\big(x; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\big)$$

In the following $\boldsymbol{\varepsilon}$ is a random number taken from a standard normal distribution:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t)$$

$$\Rightarrow \boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\varepsilon}_t$$

Performing one recursive step yields:

$$\begin{aligned}
\boldsymbol{x}_t &= \sqrt{\alpha_t}\sqrt{\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t\beta_{t-1}}\boldsymbol{\varepsilon}_{t-1} + \sqrt{\beta_t}\boldsymbol{\varepsilon}_t \\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t\beta_{t-1} + \beta_t}\boldsymbol{\varepsilon} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{1 - \alpha_t\alpha_{t-1}}\boldsymbol{\varepsilon}
\end{aligned}$$

After t steps we get:

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}$$

This means that $\boldsymbol{x}_t$ is also normally distributed:

$$\boldsymbol{x}_t \sim q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t, \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, 1-\bar{\alpha}_t)$$

We can precompute $\bar{\alpha}_t \Rightarrow$ we can compute $\boldsymbol{x}_t$ directly from $\boldsymbol{x}_0$

2 Reverse process

For the reverse process we need to find $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$. Bayes' theorem gives us:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)q(\boldsymbol{x}_t|\boldsymbol{x}_0) = q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0)q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)$$

The markov property of the forward process means:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0) = q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$$

Since $q(\boldsymbol{x}_t|\boldsymbol{x}_0)$ is independent of $\boldsymbol{x}_{t-1}$ we can write:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \propto q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)$$

We know the second term from the forward process. We can rewrite $q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ in terms of $\boldsymbol{x}_{t-1}$ as follows:

$$\begin{aligned}
q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) &\propto \exp\left(-\frac{1}{2\beta_t}\left(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1}\right)^2\right) \\
&= \exp\left(-\frac{1}{2\beta_t}\left(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1}\right)^2\right) \\
&\propto \exp\left(-\frac{\alpha_t}{2\beta_t}\left(\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\sqrt{\alpha_t}}\right)^2\right)
\end{aligned}$$

With this we can now write down an expression for the needed conditional distribution:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \propto \mathcal{N}\left(\boldsymbol{x}_{t-1}; \frac{\boldsymbol{x}_t}{\sqrt{\alpha_t}}, \frac{\beta_t}{\alpha_t}\right) \mathcal{N}\left(\boldsymbol{x}_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0, 1-\bar{\alpha}_{t-1}\right)$$

When multiplying two gaußians the result is also proportional to a gaußian. The following rules apply for the variance and mean:

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad , \quad \frac{\mu}{\sigma^2} = \frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}$$

Using this we finally get an analytic tractable form:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \propto \mathcal{N}\left(\boldsymbol{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0), \tilde{\beta}_t\right)$$

With the following variance and mean:

$$\tilde{\beta}_t = \frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$$

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t$$

Now we express the mean in terms of $\boldsymbol{x}_0$ only, by using the following identity:

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon} \implies \boldsymbol{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}(\boldsymbol{x}_t, t)\right)$$

This results in the following simplified formula for the mean:

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\varepsilon}(\boldsymbol{x}_t, t)\right)$$

Here the $\boldsymbol{\varepsilon}(\boldsymbol{x}_t, t)$ is not just a random number from a standard normal distribution, it now depends on $\boldsymbol{x}_t$ and t. This makes it the perfect quantity to be predicted by our model. We will call our model prediction $\boldsymbol{\varepsilon}_\theta(\boldsymbol{x}_t, t)$.

3 Algorithm

Training

1: repeat
2: $\boldsymbol{x}_0 \sim q(\boldsymbol{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\varepsilon},\, t\right)\right\|^2$
6: until converged

Sampling

1: $\boldsymbol{x}_T \sim \mathcal{N}(0, 1)$
2: **for** $t = T, \ldots, 1$ **do**
3: $\boldsymbol{z} \sim \mathcal{N}(0, 1)$ if $t > 1$, else $\boldsymbol{z} = 0$
4: $\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\varepsilon}_\theta(\boldsymbol{x}_t, t)\right) + \sigma_t \boldsymbol{z}$
5: **end for**
6: return x_0