Low-Dose Continous SDE

Low-Dose continioues SDE, or LDSDE for short, formulates low-dose to high-dose (qunatum-noise-wise) breast CT image synthesis problem into standard continous SDE problem suitable for diffusion network. The complete math derivation and implementation detail can be found below

Sampling

Suppose we have a perfect projection without any noise, $\boldsymbol{x}_{-1}$, which is also normalized, we can get initial ($t = 0$) high-dose scan and prior ($t = 1$) low-dose scan using the simple addition of quantum and electronic noise. [All projections are normalized by I_0, the high dose]

$$
\begin{gathered}
\text{denote } \boldsymbol{\sigma}_q^2 = \frac{\boldsymbol{x}_{-1}}{I_0} \\
\boldsymbol{x}_0 \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{x}_{-1}, \boldsymbol{\sigma}_q^2 + \sigma_e^2\right) \\
\boldsymbol{x}_0 = \boldsymbol{x}_{-1} + \sqrt{\boldsymbol{\sigma}_q^2 + \sigma_e^2}\boldsymbol{\epsilon} \\
\text{And, with } \alpha_t = \frac{I_t}{I_0} \\
\boldsymbol{x}_t \sim \boldsymbol{\mathcal{N}}(\alpha_t \boldsymbol{x}_{-1}, \alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2) \\
\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_{-1} + \sqrt{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2}\boldsymbol{\epsilon}
\end{gathered}
\tag{38}
$$

Note that the sampling of $\boldsymbol{x}_0$ is equivalent if we define $\alpha_0 = 1$, which is true

Continuous Low Dose SDE Derivation

From LDP [Wang 2014], transition between $\boldsymbol{x}_{t+\Delta t}$ and $\boldsymbol{x}_t$ can be formulated as follows:

$$
\begin{gathered}
\text{Define: } \alpha_t' = \frac{\alpha_{t+\Delta t}}{\alpha_t} \\
\boldsymbol{x}_{t+\Delta t} = \alpha_t' \boldsymbol{x}_t + \boldsymbol{\sigma}_{q,inj}\boldsymbol{\epsilon} + \sigma_{e,inj}\boldsymbol{\epsilon}
\end{gathered}
\tag{39}
$$

To match the noise vairance for both quantum noise and electronic noise:

$$
\begin{aligned}
\boldsymbol{\sigma}_{q,target}^2 = \alpha_{t+\Delta t}\boldsymbol{\sigma}_q^2 &= {\alpha_t'}^2 \boldsymbol{\sigma}_{q,x_t}^2 + \boldsymbol{\sigma}_{q,inj}^2 \\
&= {\alpha_t'}^2 \alpha_t \boldsymbol{\sigma}_q^2 + \boldsymbol{\sigma}_{q,inj}^2
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
\sigma^2_{e,target} = \sigma^2_e = {\alpha'_t}^2 \sigma^2_{e,x_t} + \sigma^2_{e,inj} \\
= {\alpha'_t}^2 \sigma^2_e + \sigma^2_{e,inj}
\end{aligned}
\tag{41}
$$

The injected noise variance:

$$
\boldsymbol{\sigma}^2_{q,inj} = (\alpha_{t+\Delta t} - {\alpha'_t}^2 \alpha_t)\boldsymbol{\sigma}^2_q \tag{42}
$$

$$
\sigma^2_{e,inj} = (1 - {\alpha'_t}^2)\sigma^2_e \tag{43}
$$

Now, we can write the transition from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+\Delta t}$ as follows:

$$
\boldsymbol{x}_{t+\Delta t} = \alpha'_t \boldsymbol{x}_t + \sqrt{(\alpha_{t+\Delta t} - {\alpha'_t}^2 \alpha_t)\boldsymbol{\sigma}^2_q}\boldsymbol{\epsilon} + \sqrt{\left(1 - {\alpha'}^2_t\right)}\sigma_e \boldsymbol{\epsilon} \tag{44}
$$

Subtract both sides of $\boldsymbol{x}_t$

$$
\boldsymbol{x}_{t+\Delta t} - \boldsymbol{x}_t = (\alpha'_t - 1)\boldsymbol{x}_t + \sqrt{\left(\alpha'_t - {\alpha'}^2_t\right)\frac{\boldsymbol{x}_{-1}}{I_0}}\boldsymbol{\epsilon} + \sqrt{\left(1 - {\alpha'}^2_t\right)}\sigma_e \boldsymbol{\epsilon} \tag{45}
$$

For the $\boldsymbol{x}_t$ term

$$
\begin{aligned}
(\alpha'_t - 1)\boldsymbol{x}_t &= \frac{\alpha_{t+\Delta t} - \alpha_t}{\alpha_t}\boldsymbol{x}_t \\
&= \frac{\alpha_{t+\Delta t} - \alpha_t}{\Delta t}\frac{\boldsymbol{x}_t}{\alpha_t}\Delta t
\end{aligned}
\tag{46}
$$

Take limit $\lim_{\Delta t \to 0}$

$$
\lim_{\Delta t \to 0}(\alpha'_t - 1)\boldsymbol{x}_t = \frac{d\alpha_t}{dt}\frac{\boldsymbol{x}_t}{\alpha_t}dt \tag{47}
$$

For the Quantum noise term

$$
\begin{aligned}
\sqrt{(\alpha_{t+\Delta t} - {\alpha'_t}^2 \alpha_t)\boldsymbol{\sigma}^2_q}\boldsymbol{\epsilon} &= \sqrt{(\alpha'_t \alpha_t - {\alpha'_t}^2 \alpha_t)\boldsymbol{\sigma}^2_q}\boldsymbol{\epsilon} \\
&= \sqrt{\alpha_t \alpha'_t (1 - \alpha'_t)\boldsymbol{\sigma}^2_q}\boldsymbol{\epsilon} \\
&= \sqrt{\alpha_t}\sqrt{\alpha'_t (1 - \alpha'_t)\boldsymbol{\sigma}^2_q}\boldsymbol{\epsilon}
\end{aligned}
\tag{48}
$$

$$
\begin{aligned}
\sqrt{\alpha_t}\sqrt{\alpha'_t\left(1 - {\alpha'}_t\right)} &= \sqrt{\alpha_t}\sqrt{\alpha'_t \frac{\alpha_t - \alpha_{t+\Delta t}}{\alpha_t}} \\
&= \sqrt{\alpha_t}\sqrt{\frac{\alpha'_t}{\alpha_t}\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}\Delta t} \\
&= \sqrt{\alpha_t}\sqrt{\frac{\alpha_{t+\Delta t}}{\alpha^2_t}}\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t} \\
&= \sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}}\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}
\end{aligned}
\tag{49}
$$

$$
\begin{aligned}
\sqrt{(\alpha_{t+\Delta t} - {\alpha'_t}^2 \alpha_t)\boldsymbol{\sigma}_q^2}\boldsymbol{\epsilon} &= \sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}}\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}\sqrt{\boldsymbol{\sigma}_q^2}\boldsymbol{\epsilon} \\
&= \sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}}\boldsymbol{\sigma}_q\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}\boldsymbol{\epsilon}
\end{aligned}
\tag{50}
$$

For the Electronic noise term

$$
\sqrt{\left(1 - {\alpha'}_t^2\right)}\sigma_e\boldsymbol{\epsilon} = \sqrt{(1+{\alpha'}_t)(1-{\alpha'}_t)}\sigma_e\boldsymbol{\epsilon}
\tag{51}
$$

$$
\begin{aligned}
\sqrt{(1+{\alpha'}_t)(1-{\alpha'}_t)} &= \sqrt{(1+{\alpha'}_t)\frac{\alpha_t - \alpha_{t+\Delta t}}{\alpha_t}} \\
&= \sqrt{\frac{(1+{\alpha'}_t)}{\alpha_t}\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}\Delta t} \\
&= \sqrt{\frac{\alpha_t + \alpha_{t+\Delta t}}{\alpha_t^2}}\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}
\end{aligned}
\tag{52}
$$

$$
\begin{aligned}
\sqrt{\left(1 - {\alpha'}_t^2\right)}\sigma_e\boldsymbol{\epsilon} &= \sqrt{\frac{\alpha_t + \alpha_{t+\Delta t}}{\alpha_t^2}}\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}\sigma_e\boldsymbol{\epsilon} \\
&= \sqrt{\frac{\alpha_t + \alpha_{t+\Delta t}}{\alpha_t^2}}\sigma_e\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}\boldsymbol{\epsilon}
\end{aligned}
\tag{53}
$$

Adding equation (13) and (16) together:

$$
\left(\sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}}\boldsymbol{\sigma}_q + \sqrt{\frac{\alpha_t + \alpha_{t+\Delta t}}{\alpha_t^2}}\sigma_e\right)\sqrt{\frac{\alpha_t - \alpha_{t+\Delta t}}{\Delta t}}\sqrt{\Delta t}\boldsymbol{\epsilon}
\tag{54}
$$

Then take $\lim_{\Delta t \to 0}$:

Using Taylor expansion $\lim_{a\to 0}\frac{f(x+a)}{f(x)} = 1$ and $\lim_{a\to 0}\frac{f(x+a)+f(x)}{f^2(x)} = \frac{2}{f(x)}$

$$
\left(\boldsymbol{\sigma}_q + \sqrt{\frac{2}{\alpha_t}}\sigma_e\right)\sqrt{-\frac{d\alpha_t}{dt}}\sqrt{dt}\boldsymbol{\epsilon} = \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-\frac{1}{\alpha_t}\frac{d\alpha_t}{dt}}\sqrt{dt}\boldsymbol{\epsilon}
\tag{55}
$$

Now, with $\lim_{\Delta t\to 0}\boldsymbol{x}_{t+\Delta t} - \boldsymbol{x}_t = d\boldsymbol{x}$, the whole SDE becomes:

$$
\begin{aligned}
dx &= \frac{d\alpha_t}{dt}\frac{\boldsymbol{x}_t}{\alpha_t}dt + \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-\frac{1}{\alpha_t}\frac{d\alpha_t}{dt}}\sqrt{dt}\boldsymbol{\epsilon} \\
&= \frac{A(t)}{\alpha(t)}\boldsymbol{x}_t dt + \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-\frac{A(t)}{\alpha(t)}}d\boldsymbol{w} \\
&= D(t)\boldsymbol{x}_t dt + \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-D(t)}d\boldsymbol{w}
\end{aligned}
\tag{56}
$$

Where we define

$$\text{Tube Output Scaling Factor: } \alpha(t) \equiv \alpha_t = \frac{I_t}{I_0} \tag{57}$$

$$A(t) = \frac{d\alpha(t)}{dt} \tag{58}$$

$$\text{Dose Scheduling: } D(t) = \frac{A(t)}{\alpha(t)} \tag{59}$$

$$d\boldsymbol{w} = \sqrt{dt}\boldsymbol{\epsilon} \tag{60}$$

The final standard SDE from

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + \boldsymbol{g}(t)d\boldsymbol{w} \tag{61}$$

Where

$$\boldsymbol{f}(\boldsymbol{x}, t) = D(t)\boldsymbol{x}_t \tag{62}$$

$$\boldsymbol{g}(t) = \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-D(t)} \tag{63}$$

Reverse Low-Dose SDE

Follow the similar derivation in Fourier Diffusion SDE [Tivnan, 2023]

Reverse SDE

$$d\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - \boldsymbol{g}^2(t)\nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_{-1})\right]dt + \boldsymbol{g}(t)d\boldsymbol{w} \tag{64}$$

The main point is then to approximate the score function $\nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_{-1})$ with neural network $\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$. However, the important thing is to get the ground truth score function.

Recall that

$$\begin{gathered} \boldsymbol{x}_t = \alpha_t\boldsymbol{x}_{-1} + \sqrt{\alpha_t\boldsymbol{\sigma}_q^2 + \sigma_e^2}\boldsymbol{\epsilon} \\ \boldsymbol{x}_t \sim \boldsymbol{\mathcal{N}}(\alpha_t\boldsymbol{x}_{-1}, \alpha_t\boldsymbol{\sigma}_q^2 + \sigma_e^2) \end{gathered} \tag{65}$$

with $\boldsymbol{\sigma} = \sqrt{\alpha_t\boldsymbol{\sigma}_q^2 + \sigma_e^2}$, score function is

$$\begin{gathered} \boldsymbol{u}(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_{-1}) = \frac{1}{p(\boldsymbol{x}_t|\boldsymbol{x}_{-1})}\frac{\partial p(\boldsymbol{x}_t|\boldsymbol{x}_{-1})}{\partial \boldsymbol{x}_t} \\ p(\boldsymbol{x}_t|\boldsymbol{x}_{-1}) = \frac{1}{\boldsymbol{\sigma}\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{\boldsymbol{x}_t - \alpha_t\boldsymbol{x}_{-1}}{\boldsymbol{\sigma}})^2} \\ \frac{\partial p(\boldsymbol{x}_t|\boldsymbol{x}_{-1})}{\partial \boldsymbol{x}_t} = \frac{-(\boldsymbol{x}_t - \alpha_t\boldsymbol{x}_{-1})}{\boldsymbol{\sigma}^2}p(\boldsymbol{x}_t|\boldsymbol{x}_{-1}) \\ \nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_{-1}) = \boldsymbol{u}(\boldsymbol{x}_t) = \frac{-(\boldsymbol{x}_t - \alpha_t\boldsymbol{x}_{-1})}{\boldsymbol{\sigma}^2} \end{gathered} \tag{66}$$

Then, the loss function is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta};\boldsymbol{\sigma}) &= \left| \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) + \frac{(\boldsymbol{x}_t - \alpha_t \boldsymbol{x}_{-1})}{\boldsymbol{\sigma}^2} \right|^2 \\
\mathcal{L}(\boldsymbol{\theta};\boldsymbol{\sigma}) &= \left| \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) + \frac{(\boldsymbol{x}_t - \alpha_t \boldsymbol{x}_{-1})}{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2} \right|^2
\end{aligned} \tag{67}
$$

However, the major problem is that even after the network if fully learned, we still have $\boldsymbol{x}_{-1}$ term in $\boldsymbol{g}(t)$ in order to go the reverse SDE. The solution to this is provided below

Improved Loss And Removal of Need for $\boldsymbol{x}_{-1}$

Notice that if we expand the score function

$$
\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t | \boldsymbol{x}_{-1}) = \frac{-(\boldsymbol{x}_t - \alpha_t \boldsymbol{x}_{-1})}{\boldsymbol{\sigma}^2} = \frac{-(\sqrt{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2}\boldsymbol{\epsilon})}{\left(\sqrt{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2}\right)^2} = \frac{-\boldsymbol{\epsilon}}{\sqrt{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2}} = \frac{-\boldsymbol{\epsilon}}{\boldsymbol{\sigma}} \tag{68}
$$

Which means, we can train the network to predict $-\boldsymbol{\epsilon}$ instead, we can also use the prior image $\boldsymbol{x}_T$ as conditioning

$$
\mathcal{L}(\boldsymbol{\theta};\boldsymbol{\sigma}) = |\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \boldsymbol{x}_T, t) + \boldsymbol{\epsilon}|^2 \tag{69}
$$

Once fully trained, we can use the network as follows:

using $\boldsymbol{x}_T$ as approximation of $\boldsymbol{x}_{-1}$

$$
\begin{aligned}
\boldsymbol{x}_{-1} &\approx \boldsymbol{x}_T / \alpha_T \\
\boldsymbol{\sigma}_q^2 &= \frac{\boldsymbol{x}_{-1}}{I_0} \\
\boldsymbol{\sigma} &= \sqrt{\alpha_t \boldsymbol{\sigma}_q^2 + \sigma_e^2} \\
\boldsymbol{f}(\boldsymbol{x}, t) &= D(t)\boldsymbol{x}_t \\
\boldsymbol{g}(t) &= \left(\sqrt{\alpha_t}\boldsymbol{\sigma}_q + \sqrt{2}\sigma_e\right)\sqrt{-D(t)} \\
\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t | \boldsymbol{x}_{-1}) &= \frac{-\boldsymbol{\epsilon}}{\boldsymbol{\sigma}} \approx \frac{\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \boldsymbol{x}_T, t)}{\boldsymbol{\sigma}} \\
d\boldsymbol{x} &= \left[\boldsymbol{f}(\boldsymbol{x}, t) - \boldsymbol{g}^2(t)\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t | \boldsymbol{x}_{-1})\right] dt + \boldsymbol{g}(t) d\boldsymbol{w}
\end{aligned} \tag{70}
$$

Dose Scheduling (Linear)

To obtain the dose scheduling term $D(t)$, we start with

$$
\begin{gathered}
I_0 : \text{Inital High Dose} \\
I_{min} : \text{Target Low Dose} \\
t \in [0, 1] \\
I_0 \to I_{min} \text{ is linear of } t \\
I(t) = (I_{min} - I_0)t + I_0
\end{gathered} \tag{71}
$$

Recall that by definition

$$\alpha(t) = \frac{I(t)}{I_0}$$
$$\alpha(t) = \frac{(I_{min} - I_0)t + I_0}{I_0} = 1 + \left(\frac{I_{min}}{I_0} - 1\right)t \tag{72}$$

Therefore, by definition of $\mathrm{A}(t)$

$$\mathrm{A}(t) = \frac{d\alpha(t)}{dt} = \left(\frac{I_{min}}{I_0} - 1\right) \tag{73}$$

Subsequently,

$$D(t) = \frac{\mathrm{A}(t)}{\alpha(t)} = \frac{\frac{I_{min}}{I_0} - 1}{1 + \left(\frac{I_{min}}{I_0} - 1\right)t} = \frac{I_{min} - I_0}{I_0 + (I_{min} - I_0)t} \tag{74}$$