Rectified Diffusion Guidance for Conditional Generation

Supplementary Material

Mengfei Xia[1,2*] Nan Xue[2] Yujun Shen[2†] Ran Yi[3] Tieliang Gong[4] Yong-Jin Liu[1†]
[1]Tsinghua University [2]Ant Group
[3]Shanghai Jiao Tong University [4]Xi'an Jiao Tong University

Appendix

A. Proofs and derivations

In this section, we will prove the theorems stated in the main manuscript.

A.1. Proof of Theorem 1

We first claim two lemmas which are crucial for the proof.

Lemma 1. *Let $g(\mathbf{x}_t)$ and $h(\mathbf{x}_t, \boldsymbol{\epsilon})$ be integrable functions, then the following equality holds.*

$$\mathbb{E}_{q(\mathbf{x})}[\langle g(\mathbf{x}), \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x})}[h(\mathbf{x}, \boldsymbol{\epsilon})|\mathbf{x}]\rangle] = \mathbb{E}_{q(\mathbf{x}, \boldsymbol{\epsilon})}[\langle g(\mathbf{x}), h(\mathbf{x}, \boldsymbol{\epsilon})\rangle], \tag{S1}$$

in which $\langle \cdot, \cdot \rangle$ is inner product.

Proof of Lemma 1. Note that

$$\begin{aligned}
\mathbb{E}_{q(\mathbf{x})}[\langle g(\mathbf{x}), \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x})}[h(\mathbf{x}, \boldsymbol{\epsilon})|\mathbf{x}]\rangle] &= \int \langle g(\mathbf{x}), \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x})}[h(\mathbf{x}, \boldsymbol{\epsilon})|\mathbf{x}]\rangle q(\mathbf{x}) \mathrm{d}\mathbf{x} &\text{(S2)}\\
&= \int \langle g(\mathbf{x}), \int h(\mathbf{x}, \boldsymbol{\epsilon}) q(\boldsymbol{\epsilon}|\mathbf{x}) \mathrm{d}\boldsymbol{\epsilon} \rangle q(\mathbf{x}) \mathrm{d}\mathbf{x} &\text{(S3)}\\
&= \iint \langle g(\mathbf{x}), h(\mathbf{x}, \boldsymbol{\epsilon})\rangle q(\mathbf{x}) q(\boldsymbol{\epsilon}|\mathbf{x}) \mathrm{d}\boldsymbol{\epsilon} \mathrm{d}\mathbf{x} &\text{(S4)}\\
&= \mathbb{E}_{q(\mathbf{x}, \boldsymbol{\epsilon})}[\langle g(\mathbf{x}), h(\mathbf{x}, \boldsymbol{\epsilon})\rangle], &\text{(S5)}
\end{aligned}$$

in which Eq. (S4) is by linearity of integral. □

Lemma 2. *The following equality of expectation holds:*

$$\mathbb{E}_{\mathbf{x}}[\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)] = \frac{1}{\sigma_t}\mathbb{E}_{\mathbf{x}}[\mathbf{x}] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c, \mathbf{x}_0, \mathbf{x}}[\mathbf{x}_0]. \tag{S6}$$

*Work finished during internship at Ant Group.
†Corresponding author.

Proof of Lemma 2. Note that

$$\nabla_{\mathbf{x}} \log q_t(\mathbf{x}) = \frac{\nabla_{\mathbf{x}} q_t(\mathbf{x})}{q_t(\mathbf{x})} \tag{S7}$$

$$= \frac{\nabla_{\mathbf{x}} \int q_t(\mathbf{x}|c)q(c)\mathrm{d}c}{q_t(\mathbf{x})} \tag{S8}$$

$$= \frac{\int \nabla_{\mathbf{x}} q_t(\mathbf{x}|c)q(c)\mathrm{d}c}{q_t(\mathbf{x})} \tag{S9}$$

$$= \frac{\int q_t(\mathbf{x}|c)q(c)\nabla_{\mathbf{x}} \log q_t(\mathbf{x}|c)\mathrm{d}c}{q_t(\mathbf{x})} \tag{S10}$$

$$= \int \frac{q_t(\mathbf{x}|c)q(c)}{q_t(\mathbf{x})} \nabla_{\mathbf{x}} \log q_t(\mathbf{x}|c)\mathrm{d}c \tag{S11}$$

$$= \mathbb{E}_{q_t(c|\mathbf{x})}[\nabla_{\mathbf{x}} \log q_t(\mathbf{x}|c)|\mathbf{x}]. \tag{S12}$$

Therefore, we have

$$\boldsymbol{\epsilon}_\theta(\mathbf{x}, t) = \mathbb{E}_{q_t(c|\mathbf{x})}[\boldsymbol{\epsilon}_\theta(\mathbf{x}, c, t)|\mathbf{x}] \tag{S13}$$

$$= \mathbb{E}_{q_t(c|\mathbf{x})}\left[\mathbb{E}_{q(\mathbf{x}_0|\mathbf{x},c)}\left[\frac{\mathbf{x} - \alpha_t \mathbf{x}_0}{\sigma_t}\right]|\mathbf{x}\right] \tag{S14}$$

$$= \mathbb{E}_{q_t(c,\mathbf{x}_0|\mathbf{x})}\left[\frac{\mathbf{x} - \alpha_t \mathbf{x}_0}{\sigma_t}|\mathbf{x}\right] \tag{S15}$$

$$= \frac{1}{\sigma_t}\mathbf{x} - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{q_t(c,\mathbf{x}_0|\mathbf{x})}[\mathbf{x}_0|\mathbf{x}], \tag{S16}$$

and

$$\mathbb{E}_{\mathbf{x}}[\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)] = \frac{1}{\sigma_t}\mathbb{E}_{\mathbf{x}}[\mathbf{x}] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{q_t(c,\mathbf{x}_0|\mathbf{x})}[\mathbf{x}_0|\mathbf{x}]] \tag{S17}$$

$$= \frac{1}{\sigma_t}\mathbb{E}_{\mathbf{x}}[\mathbf{x}] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c,\mathbf{x}_0,\mathbf{x}}[\mathbf{x}_0]. \tag{S18}$$

□

Then we start to prove Theorem 1.

Proof of Theorem 1. Similar to derivation in DDIM [28], first rewrite $J_{\delta,\gamma}$ as below:

$$J_{\delta,\gamma} = \mathbb{E}\left[-\log \hat{p}_\theta(\mathbf{x}_0|\mathbf{x}_1, c) + \sum_{t=2}^{T} D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, c)\|\hat{p}_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, c))\right] + C_1, \tag{S19}$$

in which C_1 is a constant not involving γ and θ.

Note that $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) = \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x}_t, c)}[\boldsymbol{\epsilon}|\mathbf{x}_t]$. Hence, for $t > 1$:

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,c)\|\hat{p}_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t,c))] && \text{(S20)}\\
&= \mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,c)\|q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t,\hat{\mathbf{f}}_\theta^t(\mathbf{x}_t,c),c))] && \text{(S21)}\\
&\propto \mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[\|\mathbf{x}_0 - \hat{\mathbf{f}}_\theta^t(\mathbf{x}_t,c)\|_2^2] && \text{(S22)}\\
&\propto \mathbb{E}_{\substack{\mathbf{x}_0\sim q(\mathbf{x}_0|c)\\ \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})\\ \mathbf{x}_t=\alpha_t\mathbf{x}_0+\sigma_t\boldsymbol{\epsilon}}}[\|\boldsymbol{\epsilon} - (\gamma\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t) + (1-\gamma)\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t))\|_2^2] && \text{(S23)}\\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\gamma(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t)) + (1-\gamma)(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t))\|_2^2] && \text{(S24)}\\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\gamma^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t)\|_2^2 + (1-\gamma)^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|_2^2] \\
&\qquad + 2\gamma(1-\gamma)\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t), \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\rangle] && \text{(S25)}\\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\gamma^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t)\|_2^2 + (1-\gamma)^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|_2^2] \\
&\qquad + 2\gamma(1-\gamma)\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon} - \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x}_t,c)}[\boldsymbol{\epsilon}|\mathbf{x}_t], \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\rangle] && \text{(S26)}\\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\gamma^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t)\|_2^2 + (1-\gamma)^2\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|_2^2] \\
&\qquad + 2\gamma(1-\gamma)\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon} - \boldsymbol{\epsilon}, \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\rangle] && \text{(S27)}\\
&= \gamma^2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,c,t)\|_2^2] + (1-\gamma)^2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|_2^2], && \text{(S28)}
\end{aligned}
$$

in which Eq. (S27) is from Lemma 1. As for $t = 1$ we have similar derivation:

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}_1,\mathbf{x}_0|c)}[-\log\hat{p}_\theta(\mathbf{x}_0|\mathbf{x}_1,c))] && \text{(S29)}\\
&\propto \mathbb{E}_{q(\mathbf{x}_1,\mathbf{x}_0|c)}[\|\mathbf{x}_0 - \hat{\mathbf{f}}_\theta^t(\mathbf{x}_1,c)\|_2^2] + C_2 && \text{(S30)}\\
&\propto \mathbb{E}_{\substack{\mathbf{x}_0\sim q(\mathbf{x}_0|c)\\ \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})\\ \mathbf{x}_1=\alpha_1\mathbf{x}_0+\sigma_1\boldsymbol{\epsilon}}}[\|\boldsymbol{\epsilon} - (\gamma\boldsymbol{\epsilon}_\theta(\mathbf{x}_1,c,1) + (1-\gamma)\boldsymbol{\epsilon}_\theta(\mathbf{x}_1,1))\|_2^2] + C_3 && \text{(S31)}\\
&= \gamma^2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_1,c,1)\|_2^2] + (1-\gamma)^2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_1,1)\|_2^2] + C_3, && \text{(S32)}
\end{aligned}
$$

in which C_2 and C_3 are constants not involving γ and θ. Given that CFG involves score matching using both conditional and unconditional distributions, and that $J_{\delta,\gamma}$ is proportional to the score matching objective up to a constant, we confirm the equivalence between $J_{\delta,\gamma}$ and objective of native DPM under CFG.

Note that in native PF-ODE, we have

$$
\begin{aligned}
\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} &= f_t\mathbf{x}_t - \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|c), && \text{(S33)}\\
\mathbb{E}_{q_t(\mathbf{x}_t|c)}[\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|c)] &= \mathbb{E}_{q_t(\mathbf{x}_t|c)}[\mathbb{E}_{q_t(\mathbf{x}_0|\mathbf{x}_t,c)}[\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|\mathbf{x}_0,c)]] && \text{(S34)}\\
&= \mathbb{E}_{q_t(\mathbf{x}_0,\mathbf{x}_t|c)}[\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|\mathbf{x}_0,c)] && \text{(S35)}\\
&= 0, && \text{(S36)}
\end{aligned}
$$

in which Eq. (S36) holds since forward diffusion process $q_t(\mathbf{x}_t|\mathbf{x}_0, c)$ is implemented by adding Gaussian noise. However, according to Eq. (10) and Lemma 2, we have

$$
\begin{aligned}
\mathbb{E}_{\mathbf{x}_t}[s_{t,\gamma}(\mathbf{x}_t,c)] &= \mathbb{E}_{\mathbf{x}_t}[\gamma\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|c) + (1-\gamma)\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t)] && \text{(S37)}\\
&= (1-\gamma)\mathbb{E}_{\mathbf{x}_t}[\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t)] && \text{(S38)}\\
&= \frac{\gamma-1}{\sigma_t^2}(\mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_t] - \alpha_t\mathbb{E}_{c,\mathbf{x}_0,\mathbf{x}_t}[\mathbf{x}_0]) && \text{(S39)}\\
&= \frac{\gamma-1}{\sigma_t^2}(\mathbb{E}_{q_t(\mathbf{x}_t|c)}[\mathbf{x}_t] - \alpha_t\mathbb{E}_{q_0(\mathbf{x}_0,c)}[\mathbf{x}_0]). && \text{(S40)}
\end{aligned}
$$

Note that $\mathbb{E}_{q_t(\mathbf{x}_t|c)}[\mathbf{x}_t] = \alpha_t\mathbb{E}_{q_0(\mathbf{x}_0|c)}[\mathbf{x}_0]$, and that $\mathbb{E}_{\mathbf{x}_0,c}[\mathbf{x}_0] = \int\mathbb{E}_{q_0(\mathbf{x}_0|c)}[\mathbf{x}_0]\mathrm{d}c$. Therefore when $\gamma \neq 1$, $\mathbb{E}_{\mathbf{x}_t}[s_{t,\gamma}(\mathbf{x}_t, c)]$ is not guaranteed to be identical with 0. In other words, denoising with CFG cannot be expressed as a reciprocal of diffusion process with Gaussian noise. □

A.2. Proof of Theorem 2

Proof. Given Eq. (9), for $\gamma > 1$, we have

$$s_{t,\gamma}(\mathbf{x}_t, c) = \gamma \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t|c) + (1-\gamma)\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t) \tag{S41}$$

$$= -\gamma \frac{\mathbf{x}_t - c}{t+1} - (1-\gamma)\frac{\mathbf{x}_t}{t+2}, \tag{S42}$$

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -\frac{1}{2} s_{t,\gamma}(\mathbf{x}_t, c) \tag{S43}$$

$$= \mathbf{x}_t \left(\frac{\gamma}{2(t+1)} + \frac{1-\gamma}{2(t+2)} \right) - c\frac{\gamma}{2(t+1)}. \tag{S44}$$

By variation of constants formula, we can analytically solve $q_{0,\gamma}^{\text{deter}}(\mathbf{x}_0|c)$ in Eq. (S44).

$$\mathbf{x}_t = e^{\int_T^t \frac{\gamma}{2(s+1)} + \frac{1-\gamma}{2(s+2)} \mathrm{d}s} \left(C - \int_T^t c \frac{\gamma}{2(s+1)} e^{-\int_s^t \frac{\gamma}{2(r+1)} + \frac{1-\gamma}{2(r+2)} \mathrm{d}r} \mathrm{d}s \right) \tag{S45}$$

$$= (t+1)^{\frac{\gamma}{2}} (t+2)^{\frac{1-\gamma}{2}} \left(C - c\frac{\gamma}{2} \int_T^t (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s \right), \tag{S46}$$

in which C is a constant to determine. Let $t = T$, we can see that

$$C = \frac{\mathbf{x}_T}{(T+1)^{\frac{\gamma}{2}} (T+2)^{\frac{1-\gamma}{2}}}. \tag{S47}$$

Therefore, we achieve the closed-form formula for $q_{0,\gamma}^{\text{deter}}(\mathbf{x}_0|c)$ as below:

$$\mathbf{x}_0 = 2^{\frac{1-\gamma}{2}} \left(\frac{\mathbf{x}_T}{(T+1)^{\frac{\gamma}{2}} (T+2)^{\frac{1-\gamma}{2}}} + c\frac{\gamma}{2} \int_0^T (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s \right). \tag{S48}$$

Since $q_T(\mathbf{x}_T|c) \sim \mathcal{N}(c, T+1)$, we can deduce that

$$q_{0,\gamma}^{\text{deter}}(\mathbf{x}_0|c) \sim \mathcal{N}\left(c\phi(\gamma, T), 2^{1-\gamma} \frac{T+1}{(T+1)^{\gamma}(T+2)^{1-\gamma}} \right), \tag{S49}$$

in which

$$\phi(\gamma, T) = 2^{\frac{1-\gamma}{2}} \left(\frac{1}{(T+1)^{\frac{\gamma}{2}} (T+2)^{\frac{1-\gamma}{2}}} + \frac{\gamma}{2} \int_0^T (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s \right). \tag{S50}$$

It is obvious that

$$\phi(\gamma) = 2^{\frac{1-\gamma}{2}} \frac{\gamma}{2} \int_0^{+\infty} (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s, \tag{S51}$$

$$\lim_{T \to +\infty} \frac{T+1}{(T+1)^{\gamma}(T+2)^{1-\gamma}} = 1. \tag{S52}$$

Then it is suffices to calculate $\phi(\gamma)$ for all $\gamma > 1$. First note that

$$\phi(1) = \frac{1}{2} \int_0^{+\infty} (s+1)^{-\frac{3}{2}} \mathrm{d}s = 1, \tag{S53}$$

$$\phi(3) = 2^{-1} \frac{3}{2} \int_0^{+\infty} (s+1)^{-\frac{5}{2}} (s+2) \mathrm{d}s = 2, \tag{S54}$$

$$\phi(5) = 2^{-2} \frac{5}{2} \int_0^{+\infty} (s+1)^{-\frac{7}{2}} (s+2)^2 \mathrm{d}s = \frac{7}{3}. \tag{S55}$$

For $\gamma > 1$, denote by $I(\gamma)$ with

$$I(\gamma) = \int_0^{+\infty} (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s. \tag{S56}$$

Note that for $\gamma > 1$ we have

$$\begin{aligned}
I(\gamma) &= \int_0^{+\infty} (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s && \text{(S57)}\\
&= \frac{2}{\gamma+1} \int_0^{+\infty} (s+1)^{-\frac{\gamma+2}{2}} \mathrm{d}(s+2)^{\frac{\gamma+1}{2}} && \text{(S58)}\\
&= \frac{2}{\gamma+1} \left((s+1)^{-\frac{\gamma+1}{2}} (s+2)^{\frac{\gamma+1}{2}} \Big|_0^{+\infty} + \frac{\gamma+2}{2} \int_0^{+\infty} (s+1)^{-\frac{\gamma+4}{2}} (s+2)^{\frac{\gamma+1}{2}} \mathrm{d}s \right) && \text{(S59)}\\
&= \frac{2}{\gamma+1} \left(\frac{\gamma+2}{2} I(\gamma+2) - 2^{\frac{\gamma+1}{2}} \right). && \text{(S60)}
\end{aligned}$$

Therefore, for $\gamma > 1$ we have

$$\phi(\gamma) = \frac{2\gamma}{\gamma+1} (\phi(\gamma+2) - 1), \tag{S61}$$

$$\phi(\gamma+2) = 1 + \frac{\gamma+1}{2\gamma} \phi(\gamma). \tag{S62}$$

From Eqs. (S53) to (S55) we have

$$I(1) = 2, \quad I(3) = \frac{8}{3}, \quad I(5) = \frac{56}{15}. \tag{S63}$$

For $\gamma \in [1,3]$, by Cauchy-Schwarz inequality with $p \in [0,1]$, we have

$$\begin{aligned}
&\Big(I(\gamma)\Big)^p \Big(I(5)\Big)^{1-p} && \text{(S64)}\\
&= \left(\int_0^{+\infty} (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s \right)^p \left(\int_0^{+\infty} (s+1)^{-\frac{7}{2}} (s+2)^2 \mathrm{d}s \right)^{1-p} && \text{(S65)}\\
&\geqslant \int_0^{+\infty} \left((s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \right)^p \left((s+1)^{-\frac{7}{2}} (s+2)^2 \right)^{1-p} \mathrm{d}s && \text{(S66)}\\
&= \int_0^{+\infty} (s+1)^{-\frac{\gamma p - 5p + 7}{2}} (s+2)^{-\frac{5p - \gamma p - 4}{2}} \mathrm{d}s. && \text{(S67)}
\end{aligned}$$

Let $p = \frac{2}{5-\gamma} \in [0,1]$ for $\gamma \in [1,3]$, from Eq. (S67) we have

$$I(\gamma) \geqslant \Big(I(3)\Big)^{\frac{5-\gamma}{2}} \Big(I(5)\Big)^{\frac{\gamma-3}{2}} = \left(\frac{8}{3}\right)^{\frac{5-\gamma}{2}} \left(\frac{56}{15}\right)^{\frac{\gamma-3}{2}}. \tag{S68}$$

Therefore for $\gamma \in [1,3]$, we have

$$\phi(\gamma) \geqslant 2^{\frac{1-\gamma}{2}} \frac{\gamma}{2} \left(\frac{8}{3}\right)^{\frac{5-\gamma}{2}} \left(\frac{56}{15}\right)^{\frac{\gamma-3}{2}} = \gamma \frac{7}{15} \left(\frac{10}{7}\right)^{\frac{5-\gamma}{2}} =: h_1(\gamma) \tag{S69}$$

Since $\frac{1}{\gamma} - \frac{1}{2} \log \frac{10}{7} > 0$ for $\gamma \in [1,3]$, $h_1(\gamma)$ increases monotonically on $\in [1,3]$ and $h_1(1) = \frac{20}{21}$, $h_1(3) = 2$. Similarly, for

$\gamma \in [3,5]$, by Cauchy-Schwarz inequality with $p \in [0,1]$, we have

$$\Big(I(1)\Big)^{1-p}\Big(I(\gamma)\Big)^{p} \tag{S70}$$

$$= \left(\int_0^{+\infty}(s+1)^{-\frac{3}{2}}\mathrm{d}s\right)^{1-p}\left(\int_0^{+\infty}(s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\mathrm{d}s\right)^{p} \tag{S71}$$

$$\geqslant \int_0^{+\infty}\left((s+1)^{-\frac{3}{2}}\right)^{1-p}\left((s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\right)^{p}\mathrm{d}s \tag{S72}$$

$$= \int_0^{+\infty}(s+1)^{-\frac{3-p+\gamma p}{2}}(s+2)^{-\frac{p-\gamma p}{2}}\mathrm{d}s. \tag{S73}$$

Let $p = \frac{2}{\gamma-1} \in [0,1]$ for $\gamma \in [3,5]$, from Eq. (S73) we have

$$I(\gamma) \geqslant \Big(I(1)\Big)^{\frac{3-\gamma}{2}}\Big(I(3)\Big)^{\frac{\gamma-1}{2}} = 2^{\frac{3-\gamma}{2}}\left(\frac{8}{3}\right)^{\frac{\gamma-1}{2}}. \tag{S74}$$

Therefore for $\gamma \in [3,5]$

$$\phi(\gamma) \geqslant 2^{\frac{1-\gamma}{2}}\frac{\gamma}{2}2^{\frac{3-\gamma}{2}}\left(\frac{8}{3}\right)^{\frac{\gamma-1}{2}} = \gamma\left(\frac{2}{3}\right)^{\frac{\gamma-1}{2}} =: h_2(\gamma) \tag{S75}$$

It is easy to see that $h_2(\gamma) \geqslant 2$ for $\gamma \in [3,5]$. Then by mathematical induction and Eq. (S62), we have $\phi(\gamma) \geqslant 2$ for all $\gamma \geqslant 3$. Specially, we have

$$\lim_{\gamma\to+\infty}\phi(\gamma) = 2. \tag{S76}$$

And specifically, for $\gamma \in \mathbb{N}$, $\gamma > 1$, we analytically calculate $\phi(\gamma, T)$ for $\gamma = 2n+1$ and $\gamma = 2n$, respectively. First let $\gamma = 2n+1$, $n \in \mathbb{N}$. We can see that

$$(s+2)^{-\frac{1-\gamma}{2}} = (s+2)^n = \sum_{k=0}^{n} C_n^k (s+1)^k. \tag{S77}$$

$$\int_0^T (s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\mathrm{d}s = \int_0^T (s+1)^{-\frac{\gamma+2}{2}}\left(\sum_{k=0}^{n} C_n^k (s+1)^k\right)\mathrm{d}s \tag{S78}$$

$$= \sum_{k=0}^{n}\left(C_n^k\int_0^T (s+1)^{\frac{2k-\gamma-2}{2}}\mathrm{d}s\right) \tag{S79}$$

$$= \sum_{k=0}^{n}\left(C_n^k\frac{2}{2k-\gamma}\left((T+1)^{\frac{2k-\gamma}{2}}-1\right)\right). \tag{S80}$$

Since $2k-\gamma < 0$ for $k = 0, 1, \cdots, n$, we have $(T+1)^{\frac{2k-\gamma}{2}} \to 0$ as T goes to infinity and hence

$$\phi(\gamma, T) \tag{S81}$$

$$= 2^{\frac{1-\gamma}{2}}\left(\frac{1}{(T+1)^{\frac{\gamma}{2}}(T+2)^{\frac{1-\gamma}{2}}} + \frac{\gamma}{2}\int_0^T (s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\mathrm{d}s\right) \tag{S82}$$

$$= 2^{\frac{1-\gamma}{2}}\left(\frac{1}{(T+1)^{\frac{\gamma}{2}}(T+2)^{\frac{1-\gamma}{2}}} + \frac{\gamma}{2}\left(\sum_{k=0}^{n} C_n^k\frac{2}{2k-\gamma}\left((T+1)^{\frac{2k-\gamma}{2}}-1\right)\right)\right) \tag{S83}$$

$$= 2^{\frac{1-\gamma}{2}}\left(\frac{1}{(T+1)^{\frac{\gamma}{2}}(T+2)^{\frac{1-\gamma}{2}}} + \left(\sum_{k=0}^{n} C_n^k\frac{2n+1}{2n-2k+1}\left(1-(T+1)^{\frac{2k-\gamma}{2}}\right)\right)\right). \tag{S84}$$

When $T \to +\infty$, we have

$$\phi(2n+1) = 2^{-n}\left(\sum_{k=0}^{n} C_n^k \frac{2n+1}{2n-2k+1}\right). \tag{S85}$$

Then let $\gamma = 2n$, $n \in \mathbb{N}$, and $n \geqslant 1$. We have

$$\int (s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\mathrm{d}s \tag{S86}$$

$$= \int 2(s+1)^{-n-1}(\sqrt{s+2})^{2n}\mathrm{d}\sqrt{s+2} \tag{S87}$$

$$= \int 2(u^2-1)^{-n-1}u^{2n}\mathrm{d}u \tag{S88}$$

$$= -\frac{1}{n}u^{2n-1}(u^2-1)^{-n} + \frac{2n-1}{n}\int (u^2-1)^{-n}u^{2n-2}\mathrm{d}u, \tag{S89}$$

in which Eq. (S88) is due to integration by substitution with $u = \sqrt{s+2} > 1$, and Eq. (S89) is due to integration by parts. Denote by I_n with

$$I_n = \int 2(u^2-1)^{-n-1}u^{2n}\mathrm{d}u, \quad n \geqslant 0, \tag{S90}$$

then we have

$$I_n = -\frac{1}{n}u^{2n-1}(u^2-1)^{-n} + \frac{2n-1}{2n}I_{n-1}, \quad n \geqslant 1. \tag{S91}$$

For $n \geqslant 1$, let $I_n = \frac{(2n-1)!!}{(2n)!!}A_n$, then we have

$$\frac{(2n-1)!!}{(2n)!!}A_n = -\frac{1}{n}u^{2n-1}(u^2-1)^{-n} + \frac{2n-1}{2n}\frac{(2n-3)!!}{(2n-2)!!}A_{n-1}, \quad n \geqslant 2, \tag{S92}$$

$$A_n = A_{n-1} - \frac{1}{n}\frac{(2n)!!}{(2n-1)!!}u^{2n-1}(u^2-1)^{-n}, \quad n \geqslant 2. \tag{S93}$$

Therefore for $n \geqslant 2$, we have

$$A_n = A_1 - \sum_{k=2}^{n}\frac{1}{k}\frac{(2k)!!}{(2k-1)!!}u^{2k-1}(u^2-1)^{-k}, \tag{S94}$$

and

$$I_n = \begin{cases} -\dfrac{u}{u^2-1} + \dfrac{1}{2}\log\dfrac{u-1}{u+1}, & n = 1, \\ \dfrac{(2n-1)!!}{(2n)!!}\left(-\dfrac{2u}{u^2-1} + \log\dfrac{u-1}{u+1} - \displaystyle\sum_{k=2}^{n}\frac{1}{k}\frac{(2k)!!}{(2k-1)!!}\frac{u^{2k-1}}{(u^2-1)^k}\right), & n \geqslant 2. \end{cases} \tag{S95}$$

Therefore, for $\gamma = 2$ we have

$$\phi(\gamma, T) = 2^{\frac{1-\gamma}{2}}\left(\frac{1}{(T+1)^{\frac{\gamma}{2}}(T+2)^{\frac{1-\gamma}{2}}} + \frac{\gamma}{2}\int_0^T (s+1)^{-\frac{\gamma+2}{2}}(s+2)^{-\frac{1-\gamma}{2}}\mathrm{d}s\right) \tag{S96}$$

$$= 2^{\frac{1-\gamma}{2}}\frac{1}{(T+1)^{\frac{\gamma}{2}}(T+2)^{\frac{1-\gamma}{2}}} - 2^{\frac{1-\gamma}{2}}\frac{\gamma}{2}\left(\frac{\sqrt{T+2}}{T+1} - \sqrt{2}\right) + 2^{\frac{1-\gamma}{2}}\frac{\gamma}{2}\frac{1}{2}\left(\log\frac{\sqrt{T+2}-1}{\sqrt{T+2}+1} - \log\frac{\sqrt{2}-1}{\sqrt{2}+1}\right), \tag{S97}$$

and for $\gamma \geqslant 4$ we have

$$\phi(\gamma, T) \tag{S98}$$

$$= 2^{\frac{1-\gamma}{2}} \left(\frac{1}{(T+1)^{\frac{\gamma}{2}} (T+2)^{\frac{1-\gamma}{2}}} + \frac{\gamma}{2} \int_0^T (s+1)^{-\frac{\gamma+2}{2}} (s+2)^{-\frac{1-\gamma}{2}} \mathrm{d}s \right) \tag{S99}$$

$$\begin{aligned} &= 2^{\frac{1-\gamma}{2}} \frac{1}{(T+1)^{\frac{\gamma}{2}} (T+2)^{\frac{1-\gamma}{2}}} \\ &\quad - 2^{\frac{1-\gamma}{2}} \frac{\gamma}{2} \frac{(2n-1)!!}{(2n)!!} \left(\frac{2\sqrt{T+2}}{T+1} - 2\sqrt{2} \right) \\ &\qquad + 2^{\frac{1-\gamma}{2}} \frac{\gamma}{2} \frac{(2n-1)!!}{(2n)!!} \left(\log \frac{\sqrt{T+2}-1}{\sqrt{T+2}+1} - \log \frac{\sqrt{2}-1}{\sqrt{2}+1} \right) \\ &\qquad\quad - 2^{\frac{1-\gamma}{2}} \frac{\gamma}{2} \frac{(2n-1)!!}{(2n)!!} \left(\sum_{k=2}^{n} \frac{1}{k} \frac{(2k)!!}{(2k-1)!!} \Big(\frac{(T+2)^{\frac{2k-1}{2}}}{(T+1)^k} - 2^{\frac{2k-1}{2}} \Big) \right). \end{aligned} \tag{S100}$$

When $T \to +\infty$, we have

$$\phi(2n) = \begin{cases} 2^{\frac{1}{2}-n} n \left(\sqrt{2} - \frac{1}{2} \log \frac{\sqrt{2}-1}{\sqrt{2}+1} \right), & n = 1, \\ \frac{(2n-1)!!\sqrt{2}n}{(2n)!!2^n} \left(\left(\sum_{k=2}^{n} \frac{1}{k} \frac{(2k)!!}{(2k-1)!!} 2^{k-\frac{1}{2}} \right) + 2\sqrt{2} - \log \frac{\sqrt{2}-1}{\sqrt{2}+1} \right), & n \geqslant 2. \end{cases} \tag{S101}$$

□

A.3. Proof of Theorem 3

Proof. We first write the closed-form expressions of DDIM sampler as below:

$$\mathbf{x}_{t-1} = \frac{\alpha_{t-1}}{\alpha_t} \mathbf{x}_t + (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t} \sigma_t) \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t), \tag{S102}$$

$$\tilde{\mathbf{x}}_{t-1} = \frac{\alpha_{t-1}}{\alpha_t} \tilde{\mathbf{x}}_t + (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t} \sigma_t) (\gamma_1 \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t) + \gamma_0 \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)). \tag{S103}$$

Then we have

$$\Delta_{t-1} = \mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_{t-1}] - \mathbb{E}_{\tilde{\mathbf{x}}_t}[\tilde{\mathbf{x}}_{t-1}] \tag{S104}$$

$$\begin{aligned} &= \frac{\alpha_{t-1}}{\alpha_t} (\mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_t] - \mathbb{E}_{\tilde{\mathbf{x}}_t}[\tilde{\mathbf{x}}_t]) \\ &\quad + (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t} \sigma_t) (\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c)] - \mathbb{E}_{\tilde{\mathbf{x}}_t}[\gamma_1 \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t) + \gamma_0 \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)]). \end{aligned} \tag{S105}$$

Note that

$$\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) = \mathbb{E}_{q(\mathbf{x}_0|\mathbf{x}_t, c)} \left[\frac{\mathbf{x}_t - \alpha_t \mathbf{x}_0}{\sigma_t} | \mathbf{x}_t \right], \quad \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t) = \mathbb{E}_{q(\mathbf{x}_0|\tilde{\mathbf{x}}_t, c)} \left[\frac{\tilde{\mathbf{x}}_t - \alpha_t \mathbf{x}_0}{\sigma_t} | \tilde{\mathbf{x}}_t \right]. \tag{S106}$$

Therefore, by $q_t(\mathbf{x}_t|c) = \int q_0(\mathbf{x}_0|c) q_{0t}(\mathbf{x}_t|\mathbf{x}_0, c) \mathrm{d}\mathbf{x}_0$ and Lemma 1 we have

$$\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t)] = \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t} \left[\frac{\mathbf{x}_t - \alpha_t \mathbf{x}_0}{\sigma_t} \right] = \frac{1}{\sigma_t} \mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_t] - \frac{\alpha_t}{\sigma_t} \mathbb{E}_{\mathbf{x}_0}[\mathbf{x}_0]. \tag{S107}$$

Similarly, by $\hat{p}_\theta(\tilde{\mathbf{x}}_t|c) = \int q_0(\mathbf{x}_0|c) q_{0T}(\mathbf{x}_T|\mathbf{x}_0, c) \hat{p}_\theta(\tilde{\mathbf{x}}_t|\mathbf{x}_T, c) \mathrm{d}\mathbf{x}_0 \mathrm{d}\mathbf{x}_T$ we have

$$\mathbb{E}_{\tilde{\mathbf{x}}_t}[\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t)] = \mathbb{E}_{\mathbf{x}_0, \tilde{\mathbf{x}}_t} \left[\frac{\tilde{\mathbf{x}}_t - \alpha_t \mathbf{x}_0}{\sigma_t} \right] = \frac{1}{\sigma_t} \mathbb{E}_{\tilde{\mathbf{x}}_t}[\tilde{\mathbf{x}}_t] - \frac{\alpha_t}{\sigma_t} \mathbb{E}_{\mathbf{x}_0}[\mathbf{x}_0]. \tag{S108}$$

Then we can simplify Δ_t as below:

$$\Delta_{t-1} = \frac{\alpha_{t-1}}{\alpha_t}\Delta_t + (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t}\sigma_t)(\frac{1}{\sigma_t}\Delta_t - \mathbb{E}_{\tilde{\mathbf{x}}_t}[(\gamma_1 - 1)\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)]) \tag{S109}$$

$$= \frac{\sigma_{t-1}}{\sigma_t}\Delta_t - (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t}\sigma_t)\mathbb{E}_{\tilde{\mathbf{x}}_t}[(\gamma_1 - 1)\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)]) \tag{S110}$$

$$= \frac{\sigma_{t-1}}{\sigma_t}\Delta_t - (\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t}\sigma_t)\mathbb{E}_{\tilde{\mathbf{x}}_t}[\boldsymbol{\epsilon}_{\gamma_1,\gamma_0}(\tilde{\mathbf{x}}_t)]). \tag{S111}$$

$\Delta_t = 0$ implies that $\mathbb{E}_{q_t(\mathbf{x}_t|c)}[\mathbf{x}_t] = \mathbb{E}_{\hat{p}_\theta(\tilde{\mathbf{x}}_t|c)}[\tilde{\mathbf{x}}_t]$. Therefore, by Eqs. (S107) and (S108) we have $\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t)] = \mathbb{E}_{\tilde{\mathbf{x}}_t}[\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, c, t)]$. According to Lemma 2 and by calculating the expectation over $\mathbf{x}_t$ and $\tilde{\mathbf{x}}_t$ respectively, we have

$$\mathbb{E}_{\tilde{\mathbf{x}}_t}[\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)] = \frac{1}{\sigma_t}\mathbb{E}_{\tilde{\mathbf{x}}_t}[\tilde{\mathbf{x}}_t] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c,\mathbf{x}_0,\tilde{\mathbf{x}}_t}[\mathbf{x}_0] = \frac{1}{\sigma_t}\mathbb{E}_{\tilde{\mathbf{x}}_t}[\tilde{\mathbf{x}}_t] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c,\mathbf{x}_0}[\mathbf{x}_0], \tag{S112}$$

$$\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)] = \frac{1}{\sigma_t}\mathbb{E}_{\tilde{\mathbf{x}}_t}[\mathbf{x}_t] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c,\mathbf{x}_0,\mathbf{x}_t}[\mathbf{x}_0] = \frac{1}{\sigma_t}\mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_t] - \frac{\alpha_t}{\sigma_t}\mathbb{E}_{c,\mathbf{x}_0}[\mathbf{x}_0]. \tag{S113}$$

Since $\Delta_t = 0$, we have $\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)] = \mathbb{E}_{\tilde{\mathbf{x}}_t}[\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t, t)]$, and thus

$$\Delta_{t-1} = -(\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t}\sigma_t)\mathbb{E}_{\mathbf{x}_t}[(\gamma_1 - 1)\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)]) \tag{S114}$$

$$= -(\sigma_{t-1} - \frac{\alpha_{t-1}}{\alpha_t}\sigma_t)\mathbb{E}_{\mathbf{x}_t}[\boldsymbol{\epsilon}_{\gamma_1,\gamma_0}(\mathbf{x}_t)]. \tag{S115}$$

□

A.4. Proof of Theorem 4

Proof. Given Eq. (28), for any γ_1 and γ_0, we have

$$s_{t,\gamma_1,\gamma_0}(\mathbf{x}_t, c) = \gamma_1\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t|c) + \gamma_0\nabla_{\mathbf{x}_t}\log q_t(\mathbf{x}_t) \tag{S116}$$

$$= -\gamma_1\frac{\mathbf{x}_t - c}{t+1} - \gamma_0\frac{\mathbf{x}_t}{t+2}, \tag{S117}$$

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -\frac{1}{2}s_{t,\gamma_1,\gamma_0}(\mathbf{x}_t, c) \tag{S118}$$

$$= \mathbf{x}_t\left(\frac{\gamma_1}{2(t+1)} + \frac{\gamma_0}{2(t+2)}\right) - c\frac{\gamma_1}{2(t+1)}. \tag{S119}$$

By variation of constants formula, we can analytically solve $q^{\text{deter}}_{0,\gamma_1,\gamma_0}(\mathbf{x}_0|c)$ in Eq. (S119).

$$\mathbf{x}_t = e^{\int_T^t \frac{\gamma_1}{2(s+1)} + \frac{\gamma_0}{2(s+2)}\mathrm{d}s}\left(C - \int_T^t c\frac{\gamma_1}{2(s+1)}e^{-\int_s^t \frac{\gamma_1}{2(r+1)} + \frac{\gamma_0}{2(r+2)}\mathrm{d}r}\mathrm{d}s\right) \tag{S120}$$

$$= (t+1)^{\frac{\gamma_1}{2}}(t+2)^{\frac{\gamma_0}{2}}\left(C - c\frac{\gamma_1}{2}\int_T^t (s+1)^{-\frac{\gamma_1+2}{2}}(s+2)^{-\frac{\gamma_0}{2}}\mathrm{d}s\right), \tag{S121}$$

in which C is a constant to determine. Let $t = T$, we can see that

$$C = \frac{\mathbf{x}_T}{(T+1)^{\frac{\gamma_1}{2}}(T+2)^{\frac{\gamma_0}{2}}}. \tag{S122}$$

Therefore, we achieve the closed-form formula for $q^{\text{deter}}_{0,\gamma}(\mathbf{x}_0|c)$ as below:

$$\mathbf{x}_0 = 2^{\frac{\gamma_0}{2}}\left(\frac{\mathbf{x}_T}{(T+1)^{\frac{\gamma_1}{2}}(T+2)^{\frac{\gamma_0}{2}}} + c\frac{\gamma_1}{2}\int_0^T (s+1)^{-\frac{\gamma_1+2}{2}}(s+2)^{-\frac{\gamma_0}{2}}\mathrm{d}s\right). \tag{S123}$$

Since $q_T(\mathbf{x}_T|c) \sim \mathcal{N}(c, T+1)$, we can deduce that

$$\mathrm{var}_{q^{\text{deter}}_{0,\gamma_1,\gamma_0}(\mathbf{x}_0|c)}[\mathbf{x}_0] = 2^{\gamma_0}(T+1)^{1-\gamma_1}(T+2)^{-\gamma_0}. \tag{S124}$$

□

A.5. Proof of Theorem 5

Proof. According to Eqs. (32) and (33), we can write the variational lower bound of $\hat{p}_\theta(\mathbf{x}_{0:T}|c)$ as below:

$$
\begin{aligned}
J_{\delta,\gamma_1,\gamma_0} &= \mathbb{E}_{q_\delta(\mathbf{x}_{0:T}|c)}[\log q_\delta(\mathbf{x}_{1:T}|\mathbf{x}_0, c) - \log \hat{p}_\theta(\mathbf{x}_{0:T}|c)] \quad \text{(S125)} \\
&= \mathbb{E}\left[-\log \hat{p}_\theta(\mathbf{x}_0|\mathbf{x}_1, c)\right] \\
&\quad + \mathbb{E}\left[\sum_{t=2}^{T} D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, c)\|\hat{p}_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, c))\right] \\
&\quad + C_1, \quad \text{(S126)}
\end{aligned}
$$

in which C_1 is a constant not involving γ_1, γ_0, and θ.

Note that $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) = \mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x}_t, c)}[\boldsymbol{\epsilon}|\mathbf{x}_t]$. Hence, for $t > 1$:

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, c)\|\hat{p}_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, c))] \quad \text{(S127)} \\
&= \mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[D_{KL}(q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, c)\|q_\delta(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{f}}_\theta^t(\mathbf{x}_t, c), c))] \quad \text{(S128)} \\
&\propto \mathbb{E}_{q(\mathbf{x}_t,\mathbf{x}_0|c)}[\|\mathbf{x}_0 - \hat{\mathbf{f}}_\theta^t(\mathbf{x}_t, c)\|_2^2] \quad \text{(S129)} \\
&\propto \mathbb{E}_{\substack{\mathbf{x}_0\sim q(\mathbf{x}_0|c) \\ \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I}) \\ \mathbf{x}_t=\alpha_t\mathbf{x}_0+\sigma_t\boldsymbol{\epsilon}}}[\|\boldsymbol{\epsilon} - (\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t))\|_2^2] \quad \text{(S130)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}\|_2^2 + \|\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2] \\
&\qquad - 2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon}, \gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\rangle] \quad \text{(S131)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}\|_2^2 + \|\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2] \\
&\qquad - 2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\mathbb{E}_{q(\boldsymbol{\epsilon}|\mathbf{x}_t,c)}[\boldsymbol{\epsilon}|\mathbf{x}_t], \gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\rangle] \quad \text{(S132)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}\|_2^2 + \|\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2] \\
&\qquad - 2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t), \gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\rangle] \quad \text{(S133)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t)\|_2^2 + \|\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2] \\
&\qquad - 2\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\langle\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t), \gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t)\rangle] \\
&\qquad\quad + \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}\|_2^2 - \|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t)\|_2^2] \quad \text{(S134)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) - (\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t))\|_2^2] + C_2 \quad \text{(S135)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|(\gamma_1 - 1)\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2] + C_2, \quad \text{(S136)}
\end{aligned}
$$

in which Eq. (S132) is from Lemma 1, and $C_2 = \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}\|_2^2 - \|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, c, t)\|_2^2]$ is constant not involving γ_1 and γ_0. As for $t = 1$ we have similar derivation:

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}_1,\mathbf{x}_0|c)}[-\log \hat{p}_\theta(\mathbf{x}_0|\mathbf{x}_1, c))] \quad \text{(S137)} \\
&\propto \mathbb{E}_{q(\mathbf{x}_1,\mathbf{x}_0|c)}[\|\mathbf{x}_0 - \hat{\mathbf{f}}_\theta^t(\mathbf{x}_1, c)\|_2^2] + C_3 \quad \text{(S138)} \\
&\propto \mathbb{E}_{\substack{\mathbf{x}_0\sim q(\mathbf{x}_0|c) \\ \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I}) \\ \mathbf{x}_1=\alpha_1\mathbf{x}_0+\sigma_1\boldsymbol{\epsilon}}}[\|\boldsymbol{\epsilon} - (\gamma_1\boldsymbol{\epsilon}_\theta(\mathbf{x}_1, c, 1) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_1, 1))\|_2^2] + C_4 \quad \text{(S139)} \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}[\|(\gamma_1 - 1)\boldsymbol{\epsilon}_\theta(\mathbf{x}_1, c, 1) + \gamma_0\boldsymbol{\epsilon}_\theta(\mathbf{x}_1, 1))\|_2^2] + C_5, \quad \text{(S140)}
\end{aligned}
$$

in which C_3, C_4, and C_5 are constants not involving γ_1 and γ_0. □

B. Pseudo-codes of Lookup Table

We below propose the pseudo-codes to achieve the lookup table and corresponding guided sampling in Algorithms 1 and 2.

Algorithm 1 Pseudo-code to achieve lookup table of `ReCFG` in a PyTorch-like style.

```
def calculate_lookup_table(net, gnet, data_loader, timesteps):
    """Defines the function to maintain the lookup table.

    Args:
        net: Noise prediction model for conditional score function.
        gnet: Noise prediction model for unconditional score function.
        data_loader: Dataloader to calculate score functions.
        timesteps: All timesteps under the given sampling trajectory.

    Returns:
        coeffs: Lookup list under all timesteps and conditions.
    """
    sum1_dict, sum2_dict = dict(), dict()
    # Iterate the dataloader.
    for x, c in data_loader:
        # Iterate for all timesteps.
        sum1s, sum2s = list(), list()
        for nfe_idx, t in enumerate(timesteps):
            # Forward process.
            noise = torch.randn_like(x)
            x_t = alpha_t * x + sigma_t * noise

            # Calculate score functions first.
            eps_cond, eps_uncond = net(x_t, c, t), gnet(x_t, t)

            # Calculate the expectation.
            sum1s.append(eps_cond.mean(dim=0, keepdim=True))
            sum2s.append(eps_uncond.mean(dim=0, keepdim=True))

        # Save the results.
        update_dict(sum1_dict, sum2_dict, c, sum1s, sum2s)

    # Calculate coefficients according to Eq. (34) for all timesteps.
    coeffs = {c: sum1_dict[c] / sum2_dict[c] for c in sum1_dict}

    # Record the mean coefficient for other conditions.
    coeffs.update(avg=sum(coeffs.values()) / len(coeffs)

    return coeffs
```

Algorithm 2 Pseudo-code for guided sampling by lookup table of `ReCFG` in a PyTorch-like style.

```
def guided_sampler(sampler, net, gnet, gamma_1, noise, c, timesteps, coeffs):
    """Defines the guided sampling with lookup table.

    Args:
        sampler: Native sampler without guidance, e.g., DDIM sampler.
        net: Noise prediction model for conditional score function.
        gnet: Noise prediction model for unconditional score function.
        gamma_1: Guidance strength similar to CFG of type `float`.
        noise: Initial random noise to denoise.
        c: Input label.
        timesteps: All timesteps under the given sampling trajectory.
        coeffs: Pre-calculated lookup table.

    Returns:
        x: A batch of samples by guided sampling.
    """
    # Calculate gamma_0.
    if c in coeffs:
        gamma_0s = (1. - gamma_1) * coeffs[c]
    else:
        gamma_0s = (1. - gamma_1) * coeffs['avg']
    # Ensure gamma_0 <= 0 and gamma_1 + gamma_0 >= 1.
    gamma_0s = clamp(gamma_0s, gamma_1)

    # Guided sampling using gamma_1 and gamma_0.
    x = noise
    for t, gamma_0 in zip(timesteps, gamma_0s):
        # Calculate score functions and apply guided sampling.
        eps_cond, eps_uncond = net(x, c, t), gnet(x, t)
        eps = eps_cond * gamma_1 + eps_uncond * gamma_0
        x = sampler(x, eps, t)

    return x
```