Pseudo Numerical Methods for Diffusion Models on Manifold

Pramook Khungurn

June 7, 2023

This note was written as I read the paper "Pseudo Numerical Methods for Diffusion Models on Manifolds" by Liu et al. [LRLZ22].

1 Introduction

- This paper is often abbreviated as "PNMD," and the algorithm it proposes is called "PLMS." The algorithm is one of the well-knonw solvers tailed to diffusion models. It is distributed with AUTOMATIC1111's Stable Diffusion UI.
- The objective of the paper is to develop a fast algorithm for sampling a diffusion model.

2 Background

2.1 DDIM Sampler

- In this note, we follow the original DDPM formulation Ho et al. [HJA20].
- The forward and backward process is divided into multiple time steps. A time step is denoted by $t \in \{0, 1, \ldots, T\}$.
- The noised sample at time step t is denoted by $\mathbf{x}_t$. So, $\mathbf{x}_0$ is the denoised sample from the data distribution, and $\mathbf{x}_T$ should be distributed like $\mathcal{N}(\mathbf{0}, I)$.
- The forward process is controlled by the parameters $\{\beta_t\}_{t=1}^T$ and is given by:

  $$\mathbf{x}_t \sim \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t I)$$

  for $t = 1, 2, \ldots, T$. We can then deduce that

  $$\mathbf{x}_t \sim \mathcal{N}(\sqrt{\overline{\alpha}_t}\mathbf{x}_0, (1-\overline{\alpha}_t) I)$$

  where $\alpha_t = 1 - \beta_t$, and $\overline{\alpha}_t = \prod_{u=1}^t \alpha_u$.
- According to Ho et al. [HJA20], the backward process can be formulated as:

  $$p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \widetilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \widetilde{\beta}_t I)$$

  where

  $$\widetilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0) = \frac{\beta_t\sqrt{\overline{\alpha}_{t-1}}}{1-\overline{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t, \tag{1}$$

  $$\widetilde{\beta}_t = \beta_t \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}. \tag{2}$$

- Song et al. later showed that the equation above is a special case of the following more general equation:

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \sqrt{\overline{\alpha}_{t-1}}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_{t-1}-\sigma_t^2}\left(\frac{\mathbf{x}_t - \sqrt{\overline{\alpha}_t}\mathbf{x}_0}{\sqrt{1-\overline{\alpha}_t}}\right), \sigma_t^2 I\right)$$

for any $0 \leq \sigma_t < \sqrt{1-\overline{\alpha}_t}$ [SME20]. When $\sigma_t = 0$, the backward process becomes deterministic once $\mathbf{x}_T$ has been sampled according to $\mathcal{N}(\sqrt{\overline{\alpha}_T}\mathbf{x}_0, (1-\overline{\alpha}_T)I)$.

- In a diffusion model, we either train a denoising model $\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ that predicts $\mathbf{x}_0$ from $\mathbf{x}_t$, or we train the noise model $\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ that predicts the Gaussian noise $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, I)$ that was used to construct $\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + (1-\overline{\alpha}_t)\boldsymbol{\xi}$. The models are related to each other via the following equation:

$$\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) = \frac{\mathbf{x}_t - \sqrt{\overline{\alpha}_t}\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{1-\overline{\alpha}_t}}. \tag{3}$$

- The DDIM sampler uses (1) along with the assumption that $\sigma_t = 0$. The update equation is given by:

$$\begin{aligned}
\mathbf{x}_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) + \sqrt{1-\overline{\alpha}_{t-1}}\left(\frac{\mathbf{x}_t - \sqrt{\overline{\alpha}_t}\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{1-\overline{\alpha}_t}}\right) \\
&= \frac{\sqrt{1-\overline{\alpha}_{t-1}}}{\sqrt{1-\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}} - \sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t}}{\sqrt{1-\overline{\alpha}_t}}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\sqrt{1-\overline{\alpha}_{t-1}}}{\sqrt{1-\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{(1-\overline{\alpha}_t)\overline{\alpha}_{t-1} - (1-\overline{\alpha}_{t-1})\overline{\alpha}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}} + \sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\sqrt{1-\overline{\alpha}_{t-1}}}{\sqrt{1-\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{\overline{\alpha}_{t-1} - \overline{\alpha}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}} + \sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t),
\end{aligned} \tag{4}$$

or, equivalently,

$$\begin{aligned}
\mathbf{x}_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\left(\frac{\mathbf{x}_t - \sqrt{1-\overline{\alpha}_t}\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{\overline{\alpha}_t}}\right) + \sqrt{1-\overline{\alpha}_{t-1}}\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\sqrt{\overline{\alpha}_{t-1}}}{\sqrt{\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} - \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\sqrt{\overline{\alpha}_{t-1}}}{\sqrt{\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{(1-\overline{\alpha}_{t-1})\overline{\alpha}_t - \overline{\alpha}_{t-1}(1-\overline{\alpha}_t)}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t})}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\sqrt{\overline{\alpha}_{t-1}}}{\sqrt{\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{\overline{\alpha}_t - \overline{\alpha}_{t-1}}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t})}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)
\end{aligned} \tag{5}$$

2.2 ODEs for DDIM Sampling

- We start by turning the DDIM update equation (5) into an ODE. Subtracting $\mathbf{x}_t$ from both sides of the equation, we have

$$\begin{aligned}
\mathbf{x}_{t-1} - \mathbf{x}_t &= \frac{\sqrt{\overline{\alpha}_{t-1}} - \sqrt{\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}\mathbf{x}_t + \left(\frac{\overline{\alpha}_t - \overline{\alpha}_{t-1}}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t})}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= \frac{\overline{\alpha}_{t-1} - \overline{\alpha}_t}{\sqrt{\overline{\alpha}_t}(\sqrt{\overline{\alpha}_{t-1}} + \sqrt{\overline{\alpha}_t})}\mathbf{x}_t + \left(\frac{\overline{\alpha}_t - \overline{\alpha}_{t-1}}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t})}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
&= (\overline{\alpha}_{t-1} - \overline{\alpha}_t)\left(\frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}(\sqrt{\overline{\alpha}_{t-1}} + \sqrt{\overline{\alpha}_t})} - \frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-1}}\sqrt{1-\overline{\alpha}_t})}\right).
\end{aligned}$$

Now, we replace $t-1$ with $t-\delta$:

$$\mathbf{x}_{t-\delta} - \mathbf{x}_t = (\overline{\alpha}_{t-\delta} - \overline{\alpha}_t)\left(\frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}(\sqrt{\overline{\alpha}_{t-\delta}} + \sqrt{\overline{\alpha}_t})} - \frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-\delta}}\sqrt{\overline{\alpha}_t} + \sqrt{\overline{\alpha}_{t-\delta}}\sqrt{1-\overline{\alpha}_t})}\right).$$

Dividing both sides by δ and taking the limit as $\delta \to 0$, we have

$$\begin{aligned}
\lim_{\delta\to 0}\frac{\mathbf{x}_{t-\delta}-\mathbf{x}_t}{\delta} &= \lim_{\delta\to 0}\frac{\overline{\alpha}_{t-\delta}-\overline{\alpha}_t}{\delta}\left(\frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}(\sqrt{\overline{\alpha}_{t-\delta}}+\sqrt{\overline{\alpha}_t})}-\frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-\delta}}\sqrt{\overline{\alpha}_t}+\sqrt{\overline{\alpha}_{t-\delta}}\sqrt{1-\overline{\alpha}_t})}\right)\\
-\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} &= -\overline{\alpha}_t'\left(\frac{\mathbf{x}_t}{2\overline{\alpha}_t}-\frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{2\overline{\alpha}_t\sqrt{1-\overline{\alpha}_t}}\right)\\
\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} &= \frac{\overline{\alpha}_t'}{2\overline{\alpha}_t}\left(\mathbf{x}_t-\frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{1-\overline{\alpha}_t}}\right). \qquad (6)
\end{aligned}$$

- Using (3), we have an equivalent ODE:

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \frac{\overline{\alpha}_t'}{2(1-\overline{\alpha}_t)}\left(\frac{\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{\overline{\alpha}_t}}-\mathbf{x}_t\right). \qquad (7)$$

- Note that the above equation can also be derived by staring with (4).

$$\begin{aligned}
&\mathbf{x}_{t-1}-\mathbf{x}_t\\
&= \frac{\sqrt{1-\overline{\alpha}_{t-1}}-\sqrt{1-\overline{\alpha}_t}}{\sqrt{1-\overline{\alpha}_t}}\mathbf{x}_t+\left(\frac{\overline{\alpha}_{t-1}-\overline{\alpha}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)\\
&= \frac{\overline{\alpha}_t-\overline{\alpha}_{t-1}}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_t})}\mathbf{x}_t+\left(\frac{\overline{\alpha}_{t-1}-\overline{\alpha}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)\\
&= (\overline{\alpha}_{t-1}-\overline{\alpha}_t)\left(-\frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_t})}+\frac{\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}\right)\\
&= (\overline{\alpha}_{t-1}-\overline{\alpha}_t)\left(\frac{\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_{t-1}}\sqrt{\overline{\alpha}_t})}-\frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-1}}+\sqrt{1-\overline{\alpha}_t})}\right).
\end{aligned}$$

This becomes

$$\mathbf{x}_{t-\delta}-\mathbf{x}_t = (\overline{\alpha}_{t-\delta}-\overline{\alpha}_t)\left(\frac{\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_t}\sqrt{\overline{\alpha}_{t-\delta}}+\sqrt{1-\overline{\alpha}_{t-\delta}}\sqrt{\overline{\alpha}_t})}-\frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}(\sqrt{1-\overline{\alpha}_{t-\delta}}+\sqrt{1-\overline{\alpha}_t})}\right),$$

which ultimately yields

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \frac{\overline{\alpha}_t'}{2(1-\overline{\alpha}_t)}\left(\frac{\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{\overline{\alpha}_t}}-\mathbf{x}_t\right).$$

2.3 Simpler ODEs for DDIM Sampling

- Teng gave in his paper a simpler formulation of the ODE for DDIM sampling [WS23].
- Starting from (4), we have that

$$\begin{aligned}
\mathbf{x}_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\left(\frac{\mathbf{x}_t-\sqrt{1-\overline{\alpha}_t}\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)}{\sqrt{\overline{\alpha}_t}}\right)+\sqrt{1-\overline{\alpha}_{t-1}}\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)\\
\frac{\mathbf{x}_{t-1}}{\sqrt{\overline{\alpha}_{t-1}}} &= \frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}}+\left(\frac{\sqrt{1-\overline{\alpha}_{t-1}}}{\sqrt{\overline{\alpha}_{t-1}}}-\frac{\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t)\\
\frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}}-\frac{\mathbf{x}_{t-1}}{\sqrt{\overline{\alpha}_{t-1}}} &= \left(\frac{\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}-\frac{\sqrt{1-\overline{\alpha}_{t-1}}}{\sqrt{\overline{\alpha}_{t-1}}}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t).
\end{aligned}$$

Replacing $t-1$ with $t-\delta$, we have

$$\frac{\mathbf{x}_t}{\sqrt{\overline{\alpha}_t}}-\frac{\mathbf{x}_{t-\delta}}{\sqrt{\overline{\alpha}_{t-\delta}}} = \left(\frac{\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}-\frac{\sqrt{1-\overline{\alpha}_{t-\delta}}}{\sqrt{\overline{\alpha}_{t-\delta}}}\right)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t,t).$$

Defining $\overline{\mathbf{x}}_t := \mathbf{x}_t/\sqrt{\overline{\alpha}_t}$ and $\gamma_t := \sqrt{1-\overline{\alpha}_t}/\sqrt{\overline{\alpha}_t}$, we have that

$$\overline{\mathbf{x}}_t - \overline{\mathbf{x}}_{t-\delta} = (\gamma_t - \gamma_{t-\delta})\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t). \tag{8}$$

In other words,

$$\frac{\overline{\mathbf{x}}_t - \overline{\mathbf{x}}_{t-\delta}}{\gamma_t - \gamma_{t-\delta}} = \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t).$$

Taking the limit as $\delta \to 0$, we have that

$$\frac{\mathrm{d}\overline{\mathbf{x}}_t}{\mathrm{d}\gamma_t} = \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \tag{9}$$

- Alternatively, we can start from (4).

$$\begin{aligned}
\mathbf{x}_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) + \sqrt{1-\overline{\alpha}_{t-1}}\left(\frac{\mathbf{x}_t - \sqrt{\overline{\alpha}_t}\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\sqrt{1-\overline{\alpha}_t}}\right) \\
\frac{\mathbf{x}_{t-1}}{\sqrt{1-\overline{\alpha}_{t-1}}} &= \frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}} + \left(\frac{\sqrt{\overline{\alpha}_{t-1}}}{\sqrt{1-\overline{\alpha}_{t-1}}} - \frac{\sqrt{\overline{\alpha}_t}}{\sqrt{1-\overline{\alpha}_t}}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \\
\frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}} - \frac{\mathbf{x}_{t-1}}{\sqrt{1-\overline{\alpha}_{t-1}}} &= \left(\frac{\sqrt{\overline{\alpha}_t}}{\sqrt{1-\overline{\alpha}_t}} - \frac{\sqrt{\overline{\alpha}_{t-1}}}{\sqrt{1-\overline{\alpha}_{t-1}}}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t).
\end{aligned}$$

Again, replacing $t-1$ with $t-\delta$, we have

$$\frac{\mathbf{x}_t}{\sqrt{1-\overline{\alpha}_t}} - \frac{\mathbf{x}_{t-\delta}}{\sqrt{1-\overline{\alpha}_{t-\delta}}} = \left(\frac{\sqrt{\overline{\alpha}_t}}{\sqrt{1-\overline{\alpha}_t}} - \frac{\sqrt{\overline{\alpha}_{t-\delta}}}{\sqrt{1-\overline{\alpha}_{t-\delta}}}\right)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t).$$

Define $\widetilde{\mathbf{x}}_t := \mathbf{x}_t/\sqrt{1-\overline{\alpha}_t}$ and $\omega_t = \sqrt{\overline{\alpha}_t}/\sqrt{1-\overline{\alpha}_t} = \gamma_t^{-1}$. We have that the above equation can be rewritten as

$$\widetilde{\mathbf{x}}_t - \widetilde{\mathbf{x}}_{t-\delta} = (\omega_t - \omega_{t-\delta})\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t). \tag{10}$$

Finally, taking the limit as $\delta \to 0$, we have

$$\frac{\mathrm{d}\widetilde{\mathbf{x}}_t}{\mathrm{d}\omega_t} = \mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t). \tag{11}$$

2.4 Numerical Methods

- A numerical method in this context is a way to numerically solve the initial value problem:

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}s} &= \mathbf{f}(\mathbf{y}, s) \\
\mathbf{y}(0) &= \mathbf{y}_0
\end{aligned}$$

where $\mathbf{f}$ is an arbitrary function and $\mathbf{y}_0$ is a given value. By "numerically solving" the initial value problem, we mean to compute $\mathbf{y}(S)$ for any $S > 0$.

- In a typical setting, we would divide the interval $[0, S]$ into K equal subintervals. We let $\Delta s = S/K$. Numerical methods will compute $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_K$ that are supposed to approximate $\mathbf{y}(\Delta s)$, $\mathbf{y}(2\Delta s)$, $\mathbf{y}(3\Delta s)$, $\ldots$, $\mathbf{y}(K\Delta s)$, respectively. A numerical method generally gives how to compute $\mathbf{y}_{k+1}$ from $\mathbf{y}_k$, $\mathbf{y}_k$, or any other terms that come before it.

- Here are some famous numerical methods. For brevity, let us s_k denote $k\Delta s$.

– Euler method.

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s\, \mathbf{f}(\mathbf{y}_k, s_k).$$

– Heun's method.

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{f}(\mathbf{y}_k, s_k), \\
\mathbf{f}_2 &= \mathbf{f}(\mathbf{y}_k + \Delta s \mathbf{f}_1, s_{k+1}), \\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \Delta s \left(\frac{\mathbf{f}_1 + \mathbf{f}_2}{2} \right).
\end{aligned}$$

– Runge–Kutta method. The following is the 4th order version of the method.

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{f}(\mathbf{y}_k, s_k), \\
\mathbf{f}_2 &= \mathbf{f}\left(\mathbf{y}_k + \frac{\Delta s}{2} \mathbf{f}_1, s_{k+1/2} \right), \\
\mathbf{f}_3 &= \mathbf{f}\left(\mathbf{y}_k + \frac{\Delta s}{2} \mathbf{f}_2, s_{k+1/2} \right), \\
\mathbf{f}_4 &= \mathbf{f}(\mathbf{y}_k + \Delta s \mathbf{f}_3, s_{k+1}), \\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \Delta s \left(\frac{\mathbf{f}_1 + 2\mathbf{f}_2 + 2\mathbf{f}_3 + \mathbf{f}_4}{6} \right).
\end{aligned}$$

– Linear multi-step method. Each of this is a collection of numerical methods. The ith method in the collection is referred to as the "ith-order" method. For these methods, we define

$$\mathbf{f}_k = \mathbf{f}(\mathbf{s}_k, s_k).$$

Now, $\mathbf{y}_{k+1}$ would be defined in terms of $\mathbf{y}_k$ and $\mathbf{f}_k$, $\mathbf{f}_{k-1}$, and so on. One well-known linear multi-step method is the Adams–Bashforth method.

* The first-order method is the Euler method.

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s \mathbf{f}_k.$$

* The 2nd-order method is given by:

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s \left(\frac{3\mathbf{f}_k - \mathbf{f}_{k-1}}{2} \right).$$

* The 3rd-order method is given by

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s \left(\frac{23\mathbf{f}_k - 16\mathbf{f}_{k-1} + 5\mathbf{f}_{k-2}}{12} \right).$$

* The 4th-order method is given by

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s \left(\frac{55\mathbf{f}_k - 59\mathbf{f}_{k-1} + 37\mathbf{f}_{k-2} - 9\mathbf{f}_{k-3}}{24} \right).$$

3 Method

3.1 Problems with Conventional Numerical Methods

- Once we have an ODE like (6) or (7), we can apply numerical methods discussed previously.

- However, there are difficulties in applying numerical methods to (6) and (7).
  - Both ODEs require the term $\overline{\alpha}_t'$. Nevertheless, t is discrete in case of the original DDPM of Ho et al. [HJA20], and so how are we supposed to find the derivative of a discrete function to begin with? We can do something about it by trying to make $\overline{\alpha}_t$ a continuous function, so this is just a nuisance.
  - The RHSs of the ODEs blow up near $t = 0$ and $t = T$.
- On the other hand, the ODEs of Section 2.3, which are (9) and (11), do not have any of the above problems. There are no derivate terms. Moreover, the RHSs do not blow up.
- The problem with these equations is that the derivatives on the LHSs are with respect to γ_t and ω_t, not t. So, the numerical methods cannot be applied directly. However, we will see immediately in the next section how to circumvent this problem.
- Theoretical problems aside, Salimans and Ho observed that using higher-order numerical methods on (6) or (7) can introduce noise when the number of steps are small [SH22].
- The paper suggests that the above noise problem is due to the fact that diffusion models are typically trained to denoise $\mathbf{x}_t$ of the form $\sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\xi}$ where $\boldsymbol{\xi}$ is a univariate Gaussian noise. So, there is a "manifold" where the noised samples belong. If the trajectory of a numerical method strays too far away from this manifold, then the results would be off. The thesis is that, if we use the update rule based on (6) or (7), it becomes easy for the trajectory to veer off course.

3.2 Pseudo Numerical Methods

- The paper observes that the numerical methods in Section 2.4 can be divided into two steps.
  - **Gradient step.** This involves evaluating the function $\mathbf{f}$ at several points and then combining the values somehow to compute and estimate of $\mathrm{d}\mathbf{y}/\mathrm{d}s$.
  - **Transfer step.** This involves taking the gradient from the previous step and update the variable $\mathbf{y}$.
- Let us take the 2nd-order Adams–Bashforth method

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s\left(\frac{3\mathbf{f}_k - \mathbf{f}_{k-1}}{2}\right)$$

  as an example. Its gradient stepn is give by

$$\overline{\mathbf{f}} = \frac{3\mathbf{f}_k - \mathbf{f}_{k-1}}{2},$$

  and its transfer step is given by

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \Delta s\,\overline{\mathbf{f}}. \tag{12}$$

  Note that the equation of the transfer step above is the same for all numerical methods in Section 2.4.
- There's nothing wrong with the gradient step, but the transfer step (12) only works when the timesteps are of the form $s_k = k\Delta s$, which is a linear function of k. However, for ODEs like (9) and (9), the timesteps are $\gamma_t = \sqrt{1-\overline{\alpha}_t}/\sqrt{\overline{\alpha}_t}$ and $\omega_t = \sqrt{\overline{\alpha}_t}/\sqrt{1-\overline{\alpha}_t}$, which are not linear functions of t.

- To make the transfer step works for (9) and (11), we rewrite

$$\mathbf{y}_{k+1} = \mathbf{y}_k + (s_{k+1} - s_k)\overline{\mathbf{f}}.$$

Note that this equation is very similar to (8) and (10), which we reproduce (in rewritten forms) below:

$$\begin{aligned}\overline{\mathbf{x}}_{t-\delta} &= \overline{\mathbf{x}}_t + (\gamma_{t-\delta} - \gamma_t)\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t),\\ \widetilde{\mathbf{x}}_{t-\delta} &= \widetilde{\mathbf{x}}_t + (\omega_{t-\delta} - \omega_t)\mathbf{x}_{\boldsymbol{\theta}}(\mathbf{x}_t, t).\end{aligned}$$

So, the transfer steps for (9) and (11) of the form

$$\begin{aligned}\overline{\mathbf{x}}_{t-\delta} &= \overline{\mathbf{x}}_t + (\gamma_{t-\delta} - \gamma_t)\widehat{\boldsymbol{\xi}},\\ \widetilde{\mathbf{x}}_{t-\delta} &= \widetilde{\mathbf{x}}_t + (\omega_{t-\delta} - \omega_t)\widehat{\boldsymbol{x}}\end{aligned}$$

where $\widehat{\boldsymbol{\xi}}$ is some linear combination of values obtained by evaluating $\boldsymbol{\xi}_{\boldsymbol{\theta}}$ at several points, and $\widehat{\boldsymbol{x}}$ is also some linear combination of values obtained by evaluating $\mathbf{x}_{\boldsymbol{\theta}}$ at several points.

- Before we go ahead, let us take stock. What does the transfer step like

$$\overline{\mathbf{x}}_{t-\delta} = \overline{\mathbf{x}}_t + (\gamma_{t-\delta} - \gamma_t)\widehat{\boldsymbol{\xi}}$$

actually mean? Does it mean that we have to compute in terms of $\overline{\mathbf{x}}_t = \mathbf{x}_t/\sqrt{\overline{\alpha}_t}$ of $\mathbf{x}_t$? If so, this would not be very nice because this value would blow up near $t = 0$. The good news is that we can still do everything in terms of $\mathbf{x}_t$ because the above equation is just another form of the DDIM update rule (5). In other words, after rearranging, it becomes:

$$\mathbf{x}_{t-\delta} = \sqrt{\overline{\alpha}_{t-\delta}}\left(\frac{\mathbf{x}_t - \widehat{\boldsymbol{\xi}}\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}\right) + \widehat{\boldsymbol{\xi}}\sqrt{1-\overline{\alpha}_{t-\delta}}.$$

For convenience, let us define

$$\boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, s) := \sqrt{\overline{\alpha}_s}\left(\frac{\mathbf{x}_t - \widehat{\boldsymbol{\xi}}\sqrt{1-\overline{\alpha}_t}}{\sqrt{\overline{\alpha}_t}}\right) + \widehat{\boldsymbol{\xi}}\sqrt{1-\overline{\alpha}_s}.$$

So, we may write the transfer step as:

$$\mathbf{x}_{t-\delta} = \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-\delta).$$

- Similarly, the transfer step

$$\widetilde{\mathbf{x}}_{t-\delta} = \widetilde{\mathbf{x}}_t + (\omega_{t-\delta} - \omega_t)\widehat{\boldsymbol{x}}$$

is equivalent to

$$\mathbf{x}_{t-\delta} = \boldsymbol{\phi}_{\mathbf{x}}(\mathbf{x}_t, \widehat{\mathbf{x}}, t, s)$$

where

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{x}_t, \widehat{\mathbf{x}}, t, s) := \sqrt{\overline{\alpha}_s}\widehat{\mathbf{x}} + \sqrt{1-\overline{\alpha}_s}\left(\frac{\mathbf{x}_t - \sqrt{\overline{\alpha}_t}\widehat{\mathbf{x}}}{\sqrt{1-\overline{\alpha}_t}}\right).$$

- Because the transfer steps above are not the same as the transfer step of conventional numerical methods, the paper calls methods that use them "pseudo numerical."
  - However, I think this nomenclature is not really necessary. The formulation in Section 2.3 tells us that these methods are just conventional numerical methods applied to different ODEs.

- Armed with the new transfer steps, we can turn any numerical methods in Section 2.4 pseudo numerical.
  - The pseudo numerical Euler method is just the DDIM sampling method.

$$\begin{aligned}\widehat{\boldsymbol{\xi}} &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t),\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

  - Pseudo Huen's method.

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_1 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\\ \widehat{\mathbf{x}}_1 &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}_1, t, t-1)\\ \widehat{\boldsymbol{\xi}}_2 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_1, t-1)\\ \widehat{\boldsymbol{\xi}} &= \frac{\widehat{\boldsymbol{\xi}}_1 + \widehat{\boldsymbol{\xi}}_2}{2}\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

  - Pseudo Runge-Kutta method (PRK).

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_1 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\\ \widehat{\mathbf{x}}_1 &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}_1, t, t-0.5)\\ \widehat{\boldsymbol{\xi}}_2 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_1, t-0.5)\\ \widehat{\mathbf{x}}_2 &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}_2, t, t-0.5)\\ \widehat{\boldsymbol{\xi}}_3 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_2, t-0.5)\\ \widehat{\mathbf{x}}_3 &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}_3, t, t-1)\\ \widehat{\boldsymbol{\xi}}_4 &= \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_3, t-1)\\ \widehat{\boldsymbol{\xi}} &= \frac{\widehat{\boldsymbol{\xi}}_1 + 2\widehat{\boldsymbol{\xi}}_2 + 2\widehat{\boldsymbol{\xi}}_3 + \widehat{\boldsymbol{\xi}}_4}{6}\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

  - Pseudo linear multi-step method (PLMS). Let $\widehat{\boldsymbol{\xi}}_t = \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$.
    - Second order (PLMS2).

$$\begin{aligned}\widehat{\boldsymbol{\xi}} &= \frac{3\widehat{\boldsymbol{\xi}}_t - \widehat{\boldsymbol{\xi}}_{t+1}}{2}\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

    - Third order (PLMS3).

$$\begin{aligned}\widehat{\boldsymbol{\xi}} &= \frac{23\widehat{\boldsymbol{\xi}}_t - 16\widehat{\boldsymbol{\xi}}_{t+1} + 5\widehat{\boldsymbol{\xi}}_{t+2}}{12}\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

    - Fourth order (PLMS4).

$$\begin{aligned}\widehat{\boldsymbol{\xi}} &= \frac{55\widehat{\boldsymbol{\xi}}_t - 59\widehat{\boldsymbol{\xi}}_{t+1} + 37\widehat{\boldsymbol{\xi}}_{t+2} - 9\widehat{\boldsymbol{\xi}}_{t+3}}{24}\\ \mathbf{x}_{t-1} &= \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)\end{aligned}$$

3.3 Putting It All Together

- The methods we covered in the last section are ways to generate $\mathbf{x}_{t-1}$ from $\mathbf{x}_t$, $\mathbf{x}_{t+1}$, and so on. However, a complete sampling algorithm must generate $\mathbf{x}_0$ from scratch.
- The paper's main contribution is a sampling algorithm where most of the sampling steps is PLMS4. However, PLMS4 can be used to compute $\mathbf{x}_{t-1}$ only when $\widehat{\boldsymbol{\xi}}_t$, $\widehat{\boldsymbol{\xi}}_{t+1}$, $\widehat{\boldsymbol{\xi}}_{t+2}$, and $\widehat{\boldsymbol{\xi}}_{t+3}$ have been computed. As a result, we cannot use it to compute $\mathbf{x}_{T-1}$, $\mathbf{x}_{T-2}$, and $\mathbf{x}_{T-3}$. A sampling algorithm thus must specify how to compute these values.
- There are two implementations of the sampling algorithm using PLMS.
  - The one proposed by Liu et al. in the paper. It uses pseudo Runge–Kutta to compute $\mathbf{x}_{T-1}$, $\mathbf{x}_{T-2}$, and $\mathbf{x}_{T-3}$.
  - The one documented in Teng's paper [WS23]. It uses PLMS methods of lower orders to compute $\mathbf{x}_{T-1}$, $\mathbf{x}_{T-2}$, and $\mathbf{x}_{T-3}$.
- Let's start with Teng's implementation because it's the easier both to understand and to implement.

  Sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, I)$.
  for $t \leftarrow T, T-1, \ldots, 1$ **do**
    $\widehat{\boldsymbol{\xi}}_t \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$
    if $t = T$ **then**
      $\widehat{\boldsymbol{\xi}} \leftarrow \widehat{\boldsymbol{\xi}}_t$
    else if $t = T-1$ **then**
      $\widehat{\boldsymbol{\xi}} \leftarrow \frac{1}{2}(3\widehat{\boldsymbol{\xi}}_t - \widehat{\boldsymbol{\xi}}_{t+1})$
    else if $t = T-2$ **then**
      $\widehat{\boldsymbol{\xi}} \leftarrow \frac{1}{12}(23\widehat{\boldsymbol{\xi}}_t - 16\widehat{\boldsymbol{\xi}}_{t+1} + 5\widehat{\boldsymbol{\xi}}_{t+2})$
    else
      $\widehat{\boldsymbol{\xi}} \leftarrow \frac{1}{24}(55\widehat{\boldsymbol{\xi}}_t - 59\widehat{\boldsymbol{\xi}}_{t+1} + 37\widehat{\boldsymbol{\xi}}_{t+2} - 9\widehat{\boldsymbol{\xi}}_{t+3})$
    end if
    $\mathbf{x}_{t-1} \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)$
  end for
  return $\mathbf{x}_0$.

- Liu et al.'s implementation requires PRK as a subroutine.

  PRK$(\mathbf{x}, t, s)$
  begin
    $\widehat{\boldsymbol{\xi}}_1 \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}, t)$
    $\widehat{\mathbf{x}}_1 \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}, \widehat{\boldsymbol{\xi}}_1, t, (t+s)/2)$
    $\widehat{\boldsymbol{\xi}}_2 \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_1, (t+s)/2)$
    $\widehat{\mathbf{x}}_2 \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}, \widehat{\boldsymbol{\xi}}_2, t, (t+s)/2)$
    $\widehat{\boldsymbol{\xi}}_3 \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_2, (t+s)/2)$
    $\widehat{\mathbf{x}}_3 \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}, \widehat{\boldsymbol{\xi}}_3, t, s)$
    $\widehat{\boldsymbol{\xi}}_4 \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\widehat{\mathbf{x}}_3, s)$
    $\widehat{\boldsymbol{\xi}} \leftarrow \frac{1}{6}(\widehat{\boldsymbol{\xi}}_1 + 2\widehat{\boldsymbol{\xi}}_2 + 2\widehat{\boldsymbol{\xi}}_3 + \widehat{\boldsymbol{\xi}}_4)$
    $\widehat{\mathbf{x}} \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}, \widehat{\boldsymbol{\xi}}, t, s)$

return $\widehat{\mathbf{x}}$, $\widehat{\boldsymbol{\xi}}$

end

- Liu et al.'s implementation is as follows.

Sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, I)$.
for $t \leftarrow T, T-1, \ldots, 1$ **do**
  if $t \in \{T, T-1, T-2\}$ **then**
    $\mathbf{x}_{t-1}, \widehat{\boldsymbol{\xi}}_t \leftarrow \textsc{Prk}(\mathbf{x}_t, t, t-1)$
  else
    $\widehat{\boldsymbol{\xi}}_t \leftarrow \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$
    $\widehat{\boldsymbol{\xi}} \leftarrow \frac{1}{24}(55\widehat{\boldsymbol{\xi}}_t - 59\widehat{\boldsymbol{\xi}}_{t+1} + 37\widehat{\boldsymbol{\xi}}_{t+2} - 9\widehat{\boldsymbol{\xi}}_{t+3})$
    $\mathbf{x}_{t-1} \leftarrow \boldsymbol{\phi}_{\boldsymbol{\xi}}(\mathbf{x}_t, \widehat{\boldsymbol{\xi}}, t, t-1)$
  end if
end for
return $\mathbf{x}_0$.

- Liu et al.'s implementation might not be as versatile as one might hope. The thing is that it requires evaluating $\boldsymbol{\xi}_{\boldsymbol{\theta}}$ at $t = T - 1/2$, $T - 3/2$, and $T - 5/2$. The results might not be accurate on models that are trained on discrete times.

References

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020.

[LRLZ22] Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds, 2022.

[SH22] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *CoRR*, abs/2202.00512, 2022.

[SME20] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2020.

[WS23] Suttisak Wizadwongsa and Supasorn Suwajanakorn. Accelerating guided diffusion sampling with splitting numerical methods, 2023.