A proof of the well-known equality between the noise function $\epsilon_\theta(\mathbf{x}_t, t)$ and the score function $\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t)$ in denoising diffusion probabilistic models

KiHyun Yun *

October 30, 2024

Abstract

The equality $\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)$ plays a fundamental role in Denoising Diffusion Probabilistic Models (DDPM) and is widely acknowledged in the field. But although this proof is relatively straightforward, it is not easy to find a rigorous and accessible presentation of the derivation. This paper aims to present a resource for those seeking foundational understanding of this essential concept in diffusion models.

1 Introduction

In this paper, we use the definitions and notations as presented in the DDPM paper [1]. For $t = 1, \cdots, T$,

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbb{I}).$$

From the definition of $\epsilon_\theta(\cdot, t)$ in DDPM, we find the parameter $\tilde{\theta}$ minimizing the expection

$$\mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\epsilon_t \sim \mathcal{N}(\mathbf{0},\mathbb{I})}\left[\left\|\epsilon_t - \epsilon_{\tilde{\theta}}\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right]\right].$$

We define $\epsilon_\theta(\cdot, t)$ as

$$\begin{aligned}&\epsilon_\theta(\cdot, t) = f_*(\cdot)\\ &= \operatorname{argmin}_{f(\cdot)}\mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\epsilon_t \sim \mathcal{N}(\mathbf{0},\mathbb{I})}\left[\left\|\epsilon_t - f\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t\right)\right\|^2\right]\right].\end{aligned}$$

Please note that the above f_* is not used commonly regardless of the choice of t. There is a different function f_* for each t.

Theorem 1.1 *Let $\epsilon_\theta(\cdot, t)$ be as given above. Then we have*

$$\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t).$$

The equality is found in (11) in [2].

*Department of Mathematics, Hankuk University of Foreign Studies, Youngin-si, Gyeonggi-do 449-791, Republic of Korea (kihyun.yun@gmail.com).

2 Proof of the theorem 1.1

We begin by defining $\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)$ as

$$\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) = \epsilon_t,$$

where

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbb{I}).$$

Then we have

$$\frac{\mathbf{x}_0 - \sqrt{\bar{\alpha}_t}\mathbf{x}_0}{\sqrt{1-\bar{\alpha}_t}} = \epsilon_t(\mathbf{x}_t|\mathbf{x}_0) \sim \mathcal{N}(\mathbf{0}, \mathbb{I}), \tag{2.1}$$

$$q(\mathbf{x}_t|\mathbf{x}_0) = p(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)|\mathbf{x}_0) = \mathcal{N}(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)|\mathbf{0}, \mathbb{I}). \tag{2.2}$$

In details,

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t|\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbb{I}) = \frac{1}{Z}\exp\left(-\frac{1}{2(1-\bar{\alpha}_t)}\left\|\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0\right\|^2\right).$$

Then,

$$\begin{aligned}\nabla_{\mathbf{x}_t}\log q(\mathbf{x}_t|\mathbf{x}_0) &= -\frac{1}{1-\bar{\alpha}_t}\left(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0\right)\\ &= -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t(\mathbf{x}_t|\mathbf{x}_0). \end{aligned}\tag{2.3}$$

From the definition of $\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)$,

$$\begin{aligned}&\mathbb{E}_{\mathbf{x}_0\sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\epsilon_t=\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)\sim\mathcal{N}(\mathbf{0},\mathbb{I})}\left[\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right]\right]\\ &= \mathbb{E}_{\mathbf{x}_0\sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\epsilon_t\sim\mathcal{N}(\mathbf{0},\mathbb{I})}\left[\|\epsilon_t - \epsilon_\theta(\mathbf{x}_t, t)\|^2 \;\middle|\; \mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t\right]\right],\end{aligned}$$

and the function $\epsilon_\theta(\cdot, t)$ of $\mathbf{x}_t$ is a function $f_*(\cdot)$ minimizing the expectation as follows:

$$\begin{aligned}&\epsilon_\theta(\cdot, t) = f_*(\cdot)\\ &= \text{argmin}_{f(\cdot)}\\ &\mathbb{E}_{\mathbf{x}_0\sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)\sim\mathcal{N}(\mathbf{0},\mathbb{I})}\left[\|\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f(\mathbf{x}_t)\|^2 \;\middle|\; \mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t(\mathbf{x}_t|\mathbf{x}_0)\right]\right]\\ &= \text{argmin}_{f(\cdot)}\mathbb{E}_{\mathbf{x}_0\sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\mathbf{x}_t\sim q(\mathbf{x}_t|\mathbf{x}_0)}\left[\|\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f(\mathbf{x}_t)\|^2\right]\right].\end{aligned}\tag{2.4}$$

Let $h(\cdot)$ be arbitrary function, and s be a real number. Then,

$$(f_* + sh)(\mathbf{x}_t) = f_*(\mathbf{x}_t) + sh(\mathbf{x}_t).$$

At $s = 0$, the function $(f_* + sh)(\mathbf{x}_t)$ minimizes the following expection

$$F_h(s) = \mathbb{E}_{\mathbf{x}_0\sim q(\mathbf{x}_0)}\left[\mathbb{E}_{\mathbf{x}_t\sim q(\mathbf{x}_t|\mathbf{x}_0)}\left[\|\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - (f_* + sh)(\mathbf{x}_t)\|^2\right]\right].$$

Let $\Omega(\mathbf{x}_t)$ be the set of all $\mathbf{x}_t$, and it is the common domain of f_* and h. Then,

$$\begin{aligned}
0 &= \frac{dF_h}{ds}(0) \\
&= \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0)} \Big[\mathbb{E}_{\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)} \Big[2\left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) h(\mathbf{x}_t)\Big]\Big] \\
&= 2\int_{\Omega_{(\mathbf{x}_0)}} \left(\int_{\Omega_{(\mathbf{x}_t)}} \left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) h(\mathbf{x}_t) q(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_t\right) q(\mathbf{x}_0) d\mathbf{x}_0 \\
&= 2\int_{\Omega_{(\mathbf{x}_t)}} \left(\int_{\Omega_{(\mathbf{x}_0)}} \left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_0) d\mathbf{x}_0\right) h(\mathbf{x}_t) d\mathbf{x}_t \\
&= 2\int_{\Omega_{(\mathbf{x}_t)}} g(\mathbf{x}_t) h(\mathbf{x}_t) d\mathbf{x}_t, \qquad (2.5)
\end{aligned}$$

where

$$g(\mathbf{x}_t) = \int_{\Omega_{(\mathbf{x}_0)}} \left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_0) d\mathbf{x}_0.$$

Note that g is regardless of h. Since h is a arbitrary function, the equality (2.5) implies

$$\begin{aligned}
0 = g(\mathbf{x}_t) &= \int_{\Omega_{(\mathbf{x}_0)}} \left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_0) d\mathbf{x}_0 \\
&= \int_{\Omega_{(\mathbf{x}_0)}} \left(\epsilon_t(\mathbf{x}_t|\mathbf{x}_0) - f_*\left(\mathbf{x}_t\right)\right) q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 \\
&= \int_{\Omega_{(\mathbf{x}_0)}} \epsilon_t(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 - f_*\left(\mathbf{x}_t\right) \int_{\Omega_{(\mathbf{x}_0)}} q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 \\
&= \int_{\Omega_{(\mathbf{x}_0)}} \epsilon_t(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 - f_*\left(\mathbf{x}_t\right) q(\mathbf{x}_t).
\end{aligned}$$

By (2.3)

$$\begin{aligned}
f_*\left(\mathbf{x}_t\right) q(\mathbf{x}_t) &= \int_{\Omega_{(\mathbf{x}_0)}} \epsilon_t(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 \\
&= \int_{\Omega_{(\mathbf{x}_0)}} -\sqrt{1-\bar{\alpha}_t}\Big(\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t|\mathbf{x}_0)\Big) q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_0) d\mathbf{x}_0 \\
&= -\sqrt{1-\bar{\alpha}_t} \int_{\Omega_{(\mathbf{x}_0)}} \left(\frac{\nabla_{\mathbf{x}_t} q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)}\right) q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x}_0) d\mathbf{x}_0 \\
&= -\sqrt{1-\bar{\alpha}_t} \nabla_{\mathbf{x}_t} \int_{\Omega_{(\mathbf{x}_0)}} \left(\frac{q(\mathbf{x}_t, \mathbf{x}_0)}{q(\mathbf{x}_0)}\right) q(\mathbf{x}_0) d\mathbf{x}_0 \\
&= -\sqrt{1-\bar{\alpha}_t} \nabla_{\mathbf{x}_t} \int_{\Omega_{(\mathbf{x}_0)}} q(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 \\
&= -\sqrt{1-\bar{\alpha}_t} \nabla_{\mathbf{x}_t} q(\mathbf{x}_t).
\end{aligned}$$

Therefore,

$$f_*(\mathbf{x}_t) = -\sqrt{1-\bar{\alpha}_t} \frac{\nabla_{\mathbf{x}_t} q(\mathbf{x}_t)}{q(\mathbf{x}_t)} = -\sqrt{1-\bar{\alpha}_t} \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t).$$

By (2.4), this means that

$$\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t) = -\frac{f_*(\mathbf{x}_t)}{\sqrt{1-\bar{\alpha}_t}} = -\frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1-\bar{\alpha}_t}}.$$

$\square$

Acknowledgements

Originally, this was written as a class material, but it evolved into this paper. In any case, I hope it serves as a helpful resource for the related study.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel, Denoising diffusion probabilistic models, arXiv:2006.11239, 2020

[2] Prafulla Dhariwal, Alex Nichol, Diffusion Models Beat GANs on Image Synthesis, arXiv:2105.05233, 2021