Note on Flow Matching and Diffusion Models

Anderson Lee

September 5, 2025

Abstract

A note for studying diffusion and flow matching models.

Contents

Chapter 1

Understanding Stable Diffusion v3.5

1.1 Generalized Setting

1.1.1 Constructing Conditional Probability Path

Let $p_1 = \mathcal{N}(0,1)$. The goal is to construct a probability path $p_t(z_t)$ such that $p_1 = \mathcal{N}(0,1)$ and $p_0 = p_{data}$. Instead of constructing p_t, we first construct the conditional probability path $p_t(z_t|\epsilon)$, where $z_t = a_t x_0 + b_t\epsilon$ and $\epsilon \sim \mathcal{N}(0,\mathcal{I})$. With such z_t, the marginal probability path $p_t(z_t) = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\mathcal{I})}[p_t(z_t|\epsilon)]$ satisfies the condition we want, $p_0 = p_{data}$ and $p_1 = \mathcal{N}(0,1)$ when $a_0 = 1, b_0 = 0, a_1 = 0, b_1 = 1$ is satisfied.

1.1.2 Derive Conditional Vector Field $u_t(z|\epsilon)$

Recall that **flow** is a solution to the ODE, z_t is the trajectory, and $u_t(z|\epsilon)$ is the conditional vector field. The ODE conditioned on ϵ is defined as

$$\frac{d}{dt}z_t = u_t(z|\epsilon) \tag{1.1}$$

where $z_t \sim p_t(\cdot|\epsilon)$. We use $\psi_t(\cdot|\epsilon)$ to describe the flow conditioned on ϵ

$$\psi_t(\cdot|\epsilon) : x_0 \mapsto a_t x_0 + b_t\epsilon \tag{1.2}$$

$$u_t(z|\epsilon) := \frac{d}{dt}\psi_t(\psi_t^{-1}(z|\epsilon)|\epsilon) \tag{1.3}$$

Equation (1.3) takes in z conditioned on ϵ, reverse to x_0 using ψ_t^{-1}, and then take the derivative with respect to t after applying ψ_t. We can plug in the values we have to derive $u_t(z|\epsilon)$ as follows:

$$\begin{aligned}
u_t(z|\epsilon) &= \frac{d}{dt}a_t(\psi_t^{-1}(z|\epsilon)) + \frac{d}{dt}b_t\epsilon && (1.4)\\
&= \dot{a}_t(\psi_t^{-1}(z|\epsilon)) + \dot{b}_t\epsilon && (1.5)\\
&= \dot{a}_t(\frac{z - b_t\epsilon}{a_t}) + \dot{b}_t\epsilon && (1.6)\\
&= \frac{\dot{a}_t}{a_t}\cdot z - \epsilon(\frac{\dot{a}_t b_t}{a_t} - \dot{b}_t) && (1.7)\\
&= \frac{\dot{a}_t}{a_t}\cdot z - \epsilon(\frac{\dot{a}_t b_t - a_t\dot{b}_t}{a_t}) && (1.8)
\end{aligned}$$

where $\dot{a}_t = \frac{d}{d_t}a_t, \dot{b}_t = \frac{d}{dt}b_t$. Now, let $\lambda_t := \log\frac{a_t^2}{b_t^2}$ be the *signal-to-noise ratio*, and therefore $\dot{\lambda}_t = 2(\frac{\dot{a}_t}{a_t} - \frac{\dot{b}_t}{b_t})$. We can reparameterize

$$u_t(z|\epsilon) = \frac{\dot{a}_t}{a_t}\cdot z - \frac{b_t}{2}\dot{\lambda}_t\epsilon \tag{1.9}$$

1.1.3 Conditional Flow Matching (CFM) Objective

$$\mathcal{L}_{CFM} = \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta(z,t) - u_t(z|\epsilon)\|_2^2\right] \tag{1.10}$$
$$= \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta(z,t) - \frac{\dot{a}_t}{a_t}z + \frac{b_t}{2}\dot{\lambda}_t\epsilon\|_2^2\right] \tag{1.11}$$
$$= \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left(-\frac{b_t}{2}\dot{\lambda}_t\right)^2\left[\|(-\frac{2}{\dot{\lambda}_t b_t})(v_\theta(z,t) - \frac{\dot{a}_t}{a_t}z) - \epsilon\|_2^2\right] \tag{1.12}$$
$$= \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left(-\frac{b_t}{2}\dot{\lambda}_t\right)^2\left[\|\epsilon_\theta(z,t) - \epsilon\|_2^2\right] \tag{1.13}$$

where $\epsilon_\theta(z,t) = (-\frac{2}{\dot{\lambda}_t b_t})(v_\theta(z,t) - \frac{\dot{a}_t}{a_t}z)$ is used to reparameterize, and therefore we obtain a noise prediction network $\epsilon_\theta(z,t)$ to optimize for.

1.2 Rectified Flow

In Section 1.1, we build a generalized form of $\mathcal{L}_{CFM}$ with a general a_t, b_t as long as $a_0 = 1, b_0 = 0, a_1 = 0, b_1 = 1$. While there are many ways to specify the schedule of a_t, b_t, rectified flow uses a simple schedule where $a_t = (1-t), b_t = t$. i.e.

$$z_t = (1-t)x_0 + t\epsilon \tag{1.14}$$

The parameterized loss function then become

$$\mathcal{L}_{CFM}^{RF} = \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta(z,t) - u_t(z|\epsilon)\|_2^2\right] \tag{1.15}$$
$$= \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta(z,t) - \left(\frac{-1}{1-t}\cdot z - \epsilon\cdot\frac{-t-(1-t)}{1-t}\right)\|_2^2\right] \tag{1.16}$$
$$= \mathbb{E}_{t,z\sim p_t(\cdot|\epsilon),\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta(z,t) - \frac{\epsilon - z}{1-t}\|_2^2\right] \tag{1.17}$$

Instead of sampling from sampling $z \sim p_t(\cdot|\epsilon)$, in training time, we need to sample $x_0 \sim p_{data}$, and $z_t = (1-t)x_0 + t\epsilon$.

$$\mathcal{L}_{CFM}^{RF} = \mathbb{E}_{t,x_0\sim p_{data},\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta((1-t)x_0 + t\epsilon, t) - \frac{\epsilon - ((1-t)x_0 + t\epsilon)}{1-t}\|_2^2\right] \tag{1.18}$$
$$= \mathbb{E}_{t,x_0\sim p_{data},\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta((1-t)x_0 + t\epsilon, t) - \frac{(1-t)\cdot(\epsilon - x_0)}{1-t}\|_2^2\right] \tag{1.19}$$
$$= \mathbb{E}_{t,x_0\sim p_{data},\epsilon\sim\mathcal{N}(0,1)}\left[\|v_\theta((1-t)x_0 + t\epsilon, t) - (\epsilon - x_0)\|_2^2\right] \tag{1.20}$$

Chapter 2

Known Bugs

Lecture 2: Second Lecture

2.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

2.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

Remark. Oops! footnote somehow shows up earlier than expect![a]

[a]This is a footnote!
[a]This is another footnote!

Bugs caught![b]

[b]The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

2.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. $\circledast$

$\blacksquare$

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Chapter 3

Test

Lecture 2: Second Lecture

3.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

3.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

> **Remark.** Oops! footnote somehow shows up earlier than expect![a]
>
> [a]This is a footnote!
> [a]This is another footnote!

Bugs caught![b]

[b]The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

3.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. ⊛

■

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Appendix

Appendix A

Additional Proofs

A.1 Proof of ??

We can now prove **??**.

Proof of ??. See here. ■