Flow matching

Following notes from (Holderrieth & Erives, 2025).

Flow matching is a way of modelling a target distribution p_{data} by learning how to transform samples from a simpler distribution p_{noise} into target samples by moving them along a learned *flow*.

Specifically, we learn a time-varying vector field $\vec{u}_t(x)$ for $0 \le t \le 1$ with the property that, given an initial sample $x_0 \sim p_{\text{noise}}$, pushing the sample along the flow of $\vec{u}_t$ by solving the differential equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \vec{u}_t(x_t) \tag{1}$$

for $t = 0 \to 1$ results in a solution such that $x_1 \sim p_{\text{data}}$. In other words, the flow sends samples distributed as p_{noise} to samples distributed as p_{data}.

Probability paths

To come up with a flow which transforms p_{noise} into p_{data}, we need to define the "intermediate" probability distributions, p_t. A **probability path** is a family of probability distributions parametrised by $0 \le t \le 1$.

We will first consider *conditional* probability paths for a single data point, and later consider the full probability path $p_{\text{noise}} \mapsto p_{\text{data}}$.

The continuity equation

If we imagine that the probability path p_t is the time-varying density of a fluid with velocity $\vec{u}_t$ then the *continuity equation*

$$\frac{\partial p_t(x)}{\partial t} = -\nabla \cdot \left(p_t(x)\vec{u}_t(x)\right) \tag{2}$$

is satisfied if the amount of fluid (or total probability) is conserved.

This defines the *flow* $\vec{u}_t$ associated with a probability path.

Note that, if $x_t \sim p_t$ is a particle in the fluid, then $\dot{x}_t = \vec{u}_t(x_t)$ by definition. This relates the "ODE perspective" of eq. 1 to the "physics perspective" in terms of particle flow.

Conditional probability paths

Let $z \in \mathbb{R}^d$ be a data point. Consider the path $t \mapsto p_t(\cdot \mid z)$ through the space of distributions on $\mathbb{R}^d$ given by

$$p_t(\cdot \mid z) = \mathcal{N}\left(\alpha_t z, \beta_t^2 \mathbb{I}\right) \tag{3}$$

where α_t and β_t are fixed *noise schedulers*, for example $\alpha_t = t$ and $\beta_t = 1 - t$. This is the *Gaussian probability path.* It starts at the noise distribution $p_0 = \mathcal{N}(0, \mathbb{I})$ and ends at the Dirac delta $p_1 = \delta_z$.

What is the flow $\vec{u}_t(x|z)$ associated to $p_t(x|z)$?

Note that if $x_t \in p_t(\cdot\,|z)$ then we can write

$$x_t = \alpha_t z + \beta_t \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbb{I})$$

and so $\dot{x}_t = \dot{\alpha}_t z + \dot{\beta}_t \varepsilon$. But note that $\varepsilon = \frac{x_t - \alpha_t z}{\beta_t}$ so we get

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \dot{\alpha}_t z + \frac{\dot{\beta}_t}{\beta_t}(x_t - \alpha_t z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x_t = \vec{u}_t(x_t \mid z)$$

which gives us the *conditional vector field* for the Gaussian probability path.

Since $p_t(x|z)$ and $\vec{u}_t(x|z)$ together solve the continuity equation, then a time-varying random variable

$$x_0 \sim p_0(\cdot\,|z) = \mathcal{N}(0, \mathbb{I})$$

which satisfies the ODE

$$\dot{x}_t = \vec{u}_t(x_t|z)$$

will be distributed as

$$x_t \sim p_t(\cdot\,|z)$$

for all $0 \le t \le 1$, in particular $x_1 \sim p_1(\cdot\,|z) = \delta_z$ so $x_1 = z$.

Marginal probability path

We can construct a *marginal* probability path

$$p_t(x) = \int p_t(x|z) p_{\text{data}}(z)\,\mathrm{d}z \tag{4}$$

by integrating over conditional probability paths from eq. 3.

If we have a conditional probability path $p_t(x|z)$ generated by a conditional vector field $\vec{u}_t(x|z)$, then what vector field $\vec{u}_t(x)$ is associated with the marginal probability path?

We know that $p_t(x)$ and $\vec{u}_t$ together must satisfy the continuity equation

$$\dot{p}_t(x) = -\nabla \cdot (p_t \vec{u}_t)(x).$$

Substituting the definition eq. 4,

$$\begin{aligned} \dot{p}_t(x) &= \int \dot{p}_t(x|z) p_{\text{data}}(z)\,\mathrm{d}z \\ &= -\int \nabla \cdot \left(p_t(x|z)\vec{u}_t(x|z)\right) p_{\text{data}}(z)\,\mathrm{d}z \\ &= -\nabla \cdot \left(\int p_t(x|z)\vec{u}_t(x|z) p_{\text{data}}(z)\,\mathrm{d}z\right) \end{aligned}$$

so we have to vector fields whose divergences are equal everywhere. If we equate them,

$$p_t(x)\vec{u}_t(x) = \int p_t(x|z)\vec{u}_t(x|z) p_{\text{data}}(z)\,\mathrm{d}z,$$

then we get the marginal vector field:

$$\vec{u}_t(x) = \int \vec{u}_t(x|z)\left[\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}\right]\mathrm{d}z \tag{5}$$

Since $p_t(x)$ and $\vec{u}_t(x)$ together solve the continuity equation, then a time-varying random variable

$$x_0 \sim p_0 = \mathcal{N}(0, \mathbb{I})$$

which satisfies the ODE

$$\dot{x}_t = \vec{u}_t(x_t)$$

will be distributed as

$$x_t \sim p_t$$

for all $0 \le t \le 1$, in particular $x_1 \sim p_1(\cdot) = p_{\text{data}}$.

Learning $\vec{u}_t$

Suppose we wish to approximate the marginal vector field $\vec{u}_t$ with a neural network $\vec{u}_t^\theta$ with parameters θ. The *flow matching* loss function is defined as

$$\mathcal{L}(\theta) = \mathbb{E}_{t,x}\left[\left\|\vec{u}_t^\theta(x) - \vec{u}_t(x)\right\|^2\right] \tag{6}$$

where $t \sim \text{Uniform}(0,1)$ and $x \sim p_t$. In other words, we try to match the outputs the model $\vec{u}^\theta(x)$ with the true vector field $\vec{u}_t(x)$ at every possible input (t, x).

Evaluating this loss function amounts to drawing t and then $x \sim p_t$, or equivalently drawing t and $z \sim p_{\text{data}}$ then $x \sim p_t(\cdot \mid z)$. Assuming a Gaussian probability path where $p_t(\cdot \mid z) = \mathcal{N}\left(\alpha_t z \mid \beta_t^2 t\right)$, this second method amounts to picking a sample z and adding some noise to it, $x = \alpha_t z + \beta_t \varepsilon$ for $\varepsilon \in \mathcal{N}(0, \mathbb{I})$.

Note that $\vec{u}_t(x)$ cannot be evaluated directly, since eq. 5 is intractable. However, if we expand eq. 6 we get a cross term

$$\begin{aligned} \mathbb{E}_{t,x}\left[\vec{u}_t^\theta(x)\vec{u}_t(x)\right] &= \int_0^1 \mathrm{d}t \int \mathrm{d}x\, p_t(x)\vec{u}_t^\theta(x) \int \mathrm{d}z\, \vec{u}_t(x|z)\left[\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}\right] \\ &= \int_0^1 \mathrm{d}t \int \mathrm{d}z\, p_{\text{data}}(z) \int \mathrm{d}x\, p_t(x|z)\vec{u}_t^\theta(x)\vec{u}_t(x|z) \\ &= \mathbb{E}_{t,z,x\sim p_t(\cdot|z)}\left[\vec{u}_t^\theta(x)\vec{u}_t(x|z)\right] \end{aligned}$$

which only uses the conditional vector field.

We can put this cross term back into the square to obtain a simpler loss function plus a constant term.

$$\mathcal{L}(\theta) = \underbrace{\mathbb{E}_{t,z,x}\left[\left\|\vec{u}_t^\theta(x) - \vec{u}_t(x|z)\right\|^2\right]}_{\mathcal{L}_C(\theta)} + \underbrace{\left\|\vec{u}_t(x)\right\|^2 - \left\|\vec{u}_t(x|z)\right\|^2}_{\text{independent of } \theta}$$

Therefore, by minimising the *conditional flow matching* loss $\mathcal{L}_C(\theta)$, we minimise the flow matching loss $\mathcal{L}(\theta)$.

The final loss function is therefore

$$\mathcal{L}_C(\theta) = \mathbb{E}_{t,z,x}\left[\left\|\vec{u}_t^\theta(x) - \vec{u}_t(x|z)\right\|^2\right]$$

where $t \sim \text{Uniform}[0,1]$, $z \sim p_{\text{data}}$ and $x \sim p_t(\cdot \mid z)$. In the Gaussian case with eq. 3 this simplifies even further to:

Gaussian conditional flow matching loss:

$$\mathcal{L}_C(\theta) = \mathbb{E}_{t,z,\varepsilon}\left[\left\|\vec{u}_t^\theta(\alpha_t z + \beta_t\varepsilon) - \left(\dot{\alpha}_t z + \dot{\beta}_t\varepsilon\right)\right\|^2\right]$$

where $t \sim \text{Uniform}[0,1]$, $z \sim p_{\text{data}}$ and $\varepsilon \sim \mathcal{N}(0, \mathbb{I})$.

Summary of the Gaussian case

Let $\varepsilon \sim \mathcal{N}(0,1)$.

Conditional probability flow:

$$x \sim p_t(\cdot \mid z) = \mathcal{N}\left(\alpha_t z, \beta_t^2\right) \iff x = \alpha_t z + \beta_t \varepsilon$$

Conditional vector field:

$$\vec{u}_t(x|z) = \dot{x}_t = \dot{\alpha}_t z + \dot{\beta}_t \varepsilon = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x_t$$

Conditional loss:

$$\begin{aligned} \mathcal{L}(\theta) &= \mathbb{E}_{t\sim\text{Unif}[0,1], z\sim p_{\text{data}}, x\sim p_t(\cdot \mid z)}\left[\left\|\vec{u}_t^\theta(x) - \vec{u}_t(x|z)\right\|^2\right] \\ &= \mathbb{E}_{t\sim\text{Unif}[0,1], z\sim p_{\text{data}}, \varepsilon\sim\mathcal{N}(0,1)}\left[\left\|\vec{u}_t^\theta(\alpha_t z + \beta_t\varepsilon) - \left(\dot{\alpha}_t z + \dot{\beta}_t\varepsilon\right)\right\|^2\right] \end{aligned}$$

So to train θ to minimise $\mathcal{L}$:

given θ and (?)
- $t \sim \text{Unif}[0,1]$
- $z \sim p_{\text{data}}$
- $\varepsilon \sim \mathcal{N}(0,1)$
- $\nabla\theta \leftarrow \text{gradient}_\theta\left(\left\|\vec{u}_t^\theta(\alpha_t z + \beta_t\varepsilon) - \left(\dot{\alpha}_t z + \dot{\beta}_t\varepsilon\right)\right\|^2\right)$

Details I have ignored

You can add a stochastic term to get a diffusion model where we solve an SDE instead of an ODE.

Score functions

The conditional score function is $\nabla \ln p_t(x)$ and the marginal score function is

$$\begin{aligned} \nabla \ln p_t(x) &= \frac{1}{p_t(x)} \int \nabla p_t(x|z) p_{\text{data}}(z)\,\mathrm{d}z \\ &= \int \nabla \ln p_t(x|z)\left[\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}\right]\mathrm{d}z \end{aligned}$$

For a Gaussian conditional probability path, the conditional score function is:

$$\nabla \ln p_t(x|z) = \frac{\alpha_t z - x}{\beta_t^2}$$

References

Holderrieth, P., & Erives, E. (2025, July 12). *An Introduction to Flow Matching and Diffusion Models.* https://doi.org/10.48550/arXiv.2506.02070