Normalizing flows and flow matching

Sunday, February 1, 2026 9:08 PM

data point : $x \in \mathbb{R}^d$.

time-dependent density : $p: [0,1] \times \mathbb{R}^d \to \mathbb{R}_{>0}$ $:= (0, +\infty)$

$\forall t \in [0,1]$, $p_t = p(t, \cdot) : \mathbb{R}^d \to \mathbb{R}_{>0}$ s.t. $\int p_t(x)\,dx = 1$

time-dependent vector field: $v: [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$

$\forall t \in [0,1]$, $v_t = v(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$

Def. 1. [diffeomorphic map]

A map $f: \mathbb{R}^d \to \mathbb{R}^d$ is diffeomorphic if f^{-1} exists and both f and f^{-1} are continuously differentiable.

Def. 2. [flow]

A time-dependent map $\phi: [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ is a flow if $\forall t \in [0,1]$, $\phi_t = \phi(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is diffeomorphic, and ϕ_t satisfies the ODE:

$$\frac{d}{dt}\phi_t(x) = v_t(\phi_t(x)) \quad (1)$$

$$\phi_0(x) = x \quad (2)$$

Remark 1 :

path $p_t \to v_t \to v_t(x;\theta) \to v_t(x;\hat{\theta})$

$\downarrow$

(i) [Continuous normalizing flow (CNF)] $\quad \hat{v}_t = v_t(x;\hat{\theta})$ $\quad \underbrace{\phi_t(x;\hat{\theta})}_{\text{CNF}}$

A flow $\phi_t(x;\theta)$ generated by a learnable vector field $v_t(x;\theta)$ with $\theta \in \mathbb{R}^d$ as the parameters is called a CNF.

(ii) [Why vector field is needed]

Continuity equation [Villani, 2009] provides a $\underbrace{\text{necessary}}_{\text{(require assumptions)}}$ and sufficient condition for v_t generates p_t:

$$\frac{d}{dt}p_t(x) + \mathrm{div}(p_t(x)\,v_t(x)) = 0 \qquad \mathrm{div}: \mathbb{R}^d \to \mathbb{R}$$

where $\mathrm{div}(\cdot) = \langle \nabla_x, \cdot \rangle$ and $\mathrm{div}(p_t(x)\,v_t(x)) = \sum_{i=1}^{d} \partial_{x_i} p_t(x)\,v_{t,i}(x)$.

(iii) div is a linear operator mapping a vector field to a scalar field, giving the rate that the vector field alters the local volumn.

$\mathrm{div}(v_t) > 0$ $\quad \partial_{x_i} v_{t,i}(x) > 0$ $\qquad \mathrm{div}(v_t) < 0$ $\quad \partial_{x_i} v_{t,i}(x) < 0$

Properties of div :

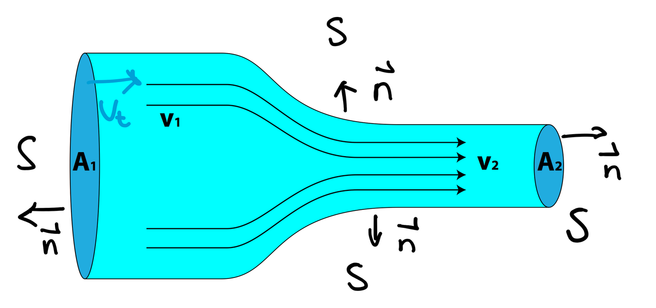
(1) [linearity] $\mathrm{div}(a v_t + b u_t) = a\,\mathrm{div}(v_t) + b\,\mathrm{div}(u_t)$, $\forall a, b \in \mathbb{R}$. v_t, u_t vector fields.

(2) [product rule] $\mathrm{div}(\varphi v_t) = (\nabla_x \varphi)\cdot v_t + \varphi\,\mathrm{div}(v_t)$ if φ scalar-valued.

(3) [curl in $\mathbb{R}^3$] $\nabla\cdot(\nabla \times v_t) = 0$

$$\nabla \times v_t = \begin{vmatrix} e_1 & e_2 & e_3 \\ \partial_{x_1} & \partial_{x_2} & \partial_{x_3} \\ v_{t,1} & v_{t,2} & v_{t,3} \end{vmatrix}.$$

$$\nabla\cdot(v_t \times u_t) = (\nabla \times v_t)\cdot u_t - v_t \cdot (\nabla \times u_t)$$



$\int p_t(x)\,dx = 1$

- mass at time t : $\int_V p_t\,dV$

(i) mass increment rate at time t : $\frac{\partial}{\partial t}\int_V p_t\,dV$ (volume integral)

(ii) mass decrement rate at time t : $\boxed{\oint_S (p_t v_t)\cdot \vec{n}\,dS}$ (surface integral)

mass flux $= \boxed{\frac{\text{Mass}}{\text{Area} \times \text{Time}}}$

mass conservation : $\int p_t(x)\,dx = 1$

(i) + (ii) = 0 $\Rightarrow$ $\frac{\partial}{\partial t}\int_V p_t\,dV + \oint_S (p_t v_t)\cdot\vec{n}\,dS = 0$

(divergence theorem) $= \int_V \mathrm{div}(p_t v_t)\,dV$

Since V is arbitrarily chosen (satisfying conditions of div thm)

$\Rightarrow \frac{\partial}{\partial t}p_t + \mathrm{div}(p_t v_t) = 0$.

> Divergence theorem : Suppose that V is a subset of $\mathbb{R}^n$ which is compact and has a piecewise smooth boundary $S = \partial V$. If F is a continuously differentiable vector field defined on a neighborhood of V, then
> $$\int_V \mathrm{div}(F)\,dV = \oint_S F\cdot\vec{n}\,dS$$
> where $\vec{n}$ is the outward pointing unit vector at almost each point on S.

A method to learn vector field that generates CNF.

Def 3. [Flow matching (FM)]

FM objective is defined as

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{\substack{t \sim \mathrm{Uniform}[0,1] \\ x \sim p_t}}\Big[\|\underbrace{v_t(x;\theta)}_{\text{CNF}} - \overbrace{u_t(x)}^{\text{true vector field}}\|_2^2\Big]$$

Remark 2: FM is intractable due to p_t and u_t.

Constructions of $p_t(x)$, $u_t(x)$ (that generates $p_t(x)$)

$p_t(x|x_1)$ $\quad p_t(x) = \int p_t(x|x_1)\,q(x_1)\,dx_1$

(1) [Lipman et. al., 2023] conditional probability path $p_t(x|x_1)$ and conditional vector field $u_t(x|x_1)$.

(2) [Albergo et. al., 2024] stochastic interpolant with coupling.

$I_t = f_t(x_0) + g_t(x_1) + h_t(z)$, $z \perp (x_0, x_1)$

p_t is density of I_t

Intuition : Suppose $x_1 \sim q(x)$ [the target distribution]

For (1) Choose appropriate $p_t(x|x_1)$ s.t.

(i) $p_0(x|x_1) = \underbrace{p(x)}_{N(0,1)}$ independent of x_1

$\boxed{\text{(ii)}}$ $p_1(x|x_1) \approx \underbrace{\delta_{x_1}(x)}_{N(x_1, \sigma_{min})}$ concentrated at x_1

Let $p_t(x) = \int p_t(x|x_1)\,q(x_1)\,dx_1$ which satisfies $\quad p_t(x_1|x) = \frac{p_t(x|x_1)\,q(x_1)}{p_t(x)}$

(a) $\boxed{p_1(x)} \approx \int \delta_{x_1}(x)\,q(x_1)\,dx_1 = \boxed{q(x_1)}$.

(b) $\boxed{u_t(x)} = \int u_t(x|x_1)\boxed{\frac{p_t(x|x_1)\,q(x_1)}{p_t(x)}}dx_1 =: \mathbb{E}_{x_1 \sim p_t(x_1|x)}[u_t(x|x_1)]$ is the vector field that generates $p_t(x)$, where $u_t(x|x_1)$ is the vector field that generates $p_t(x|x_1)$. [Theorem 1]

vector field

(c) Conditional flow matching (CFM) objective is tractable:

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{\substack{t \sim \mathrm{Uniform}[0,1] \\ x_1 \sim q, \\ x \sim p_t(x|x_1).}}\left[\|v_t(x;\theta) - u_t(x|x_1)\|_2^2\right]$$

$$\nabla \mathcal{L}_{CFM}(\theta) = \nabla \mathcal{L}_{FM}(\theta).$$

[Proof of thm 1] By Continuity Equation, it suffices to show $\frac{d}{dt}p_t(x) = -\mathrm{div}(p_t(x)\,u_t(x))$

where $p_t(x) = \int p_t(x|x_1)\,q(x_1)\,dx_1$, $u_t(x) = \int u_t(x|x_1)\frac{p_t(x|x_1)\,q(x_1)}{p_t(x)}\,dx_1$.

$$\begin{aligned}\frac{d}{dt}p_t(x) &= \int \left(\frac{d}{dt}p_t(x|x_1)\right)\cdot q(x_1)\,dx_1 \\ &= \int \left(-\mathrm{div}(p_t(x|x_1)\,u_t(x|x_1))\cdot q(x_1)\,dx_1\right. \\ &= -\mathrm{div}\left(\int p_t(x|x_1)\,u_t(x|x_1)\cdot q(x_1)\,dx_1\right) \\ &= -\mathrm{div}\left(p_t(x)\cdot\int u_t(x|x_1)\,\frac{p_t(x|x_1)\cdot q(x_1)}{p_t(x)}\,dx_1\right) \\ &= -\mathrm{div}(p_t(x)\,u_t(x))\end{aligned}$$

[Example] $p_t(x|x_1) = \mathcal{N}(x \mid \mu_t(x_1), \sigma_t(x_1)^2 I)$

where $\mu_0(x_1) = 0$, $\sigma_0(x_1) = 1$,

$\mu_1(x_1) = x_1$, $\sigma_1(x_1) = \sigma_{min}$ which is sufficient small.

For (2), choose a joint density $p(x_0, x_1)$ from coupling $\Pi(p_0, p_1)$

Let $I_t = \alpha_t x_0 + \beta_t x_1 + \gamma_t \underbrace{z}_{\sim N(0,1)}$, $t \in [0,1]$ where

(i) $\alpha_t, \beta_t, \gamma_t^2$ differentiable s.t. $\alpha_0 = \beta_1 = 1$, $\alpha_1 = \beta_0 = \gamma_0 = \gamma_1 = 0$,

$\alpha_t^2 + \beta_t^2 + \gamma_t^2 > 0 \quad \forall t \in [0,1]$.

(ii) $z \sim N(0, I) \perp (x_0, x_1)$

(iii) $\int p(x_0, x_1)\,dx_1 = p_0(x_0)$, $\int p(x_0, x_1)\,dx_0 = q(x_1)$.

Let $p_t(x)$ be the density of I_t. Then

$$b_t(x) = \mathbb{E}\left[\frac{d}{dt}I_t \,\middle|\, I_t = x\right]$$
$$= \mathbb{E}[\dot{\alpha}_t x_0 + \dot{\beta}_t x_1 + \dot{\gamma}_t z \mid I_t = x]$$

is the vector field for $\underbrace{p_t(x)}_{\text{probability path of } I_t.}$, which can be learned by minimizing

$$L_b(\hat{b}_t) = \int_0^1 \mathbb{E}_{I_t \sim p_t}\left[|\hat{b}_t(I_t)|^2 - 2\left(\frac{d}{dt}I_t\right)\cdot \hat{b}_t(I_t)\right]dt.$$

$$\operatorname{argmin} L_b(\hat{b}_t) = \operatorname{argmin} \int_0^1 \mathbb{E}_{I_t \sim p_t}\left[\|\hat{b}_t(I_t) - \tfrac{d}{dt}I_t\|_2^2\right]dt.$$