

Conditional Flow Matching

References

- ① Liu et. al., Flowing straight and fast: Learning to generate and transfer data with rectified flow, ICLR'22
 - ② Lipman et. al., Flow matching for generative modeling, 2022
 - ③ Albergo and Vanden-Eijnden, Building normalizing flows with stochastic interpolants, ICLR'23
 - ④ Lipman et. al., Flow matching guide and code, arXiv: 2412.06264, December 2024.
-

A flow view of generation

Recall our prev. setting: We defined a forward SDE (OU Process), reversed it, and density evolution described through F-P Equation. The drift learned through denoising score matching,

which used Tweedie's formula. We then derived a flow based sampler (DDIM), again using the connection between the F-P Equation and the continuity equation (see Note 9). We also showed the equivalence between a RF based sampler and a stochastic sampler (see Note 15).

In this note, we do the analog starting from a measure transport perspective (see [4] for more details, and the general setting). The outline below provides an alternate approach to directly derive a flow model.

Goal: at $t=0$, $x_0 \sim p_0$ (say $\mathcal{N}(0, I)$).
at $t=1$, $x_1 \sim q_*$ (the data dist.).

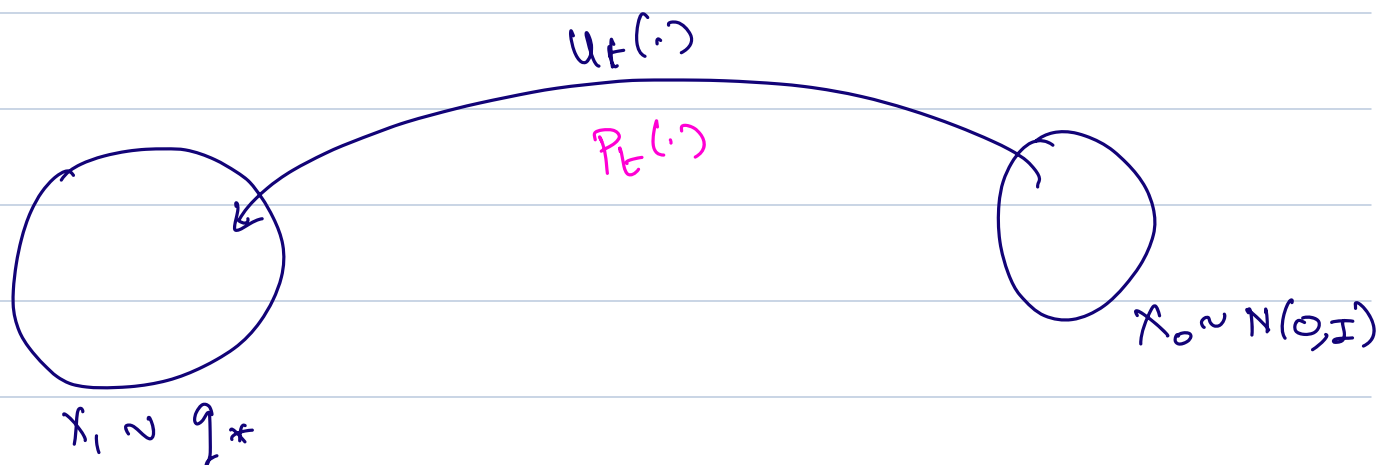
Learn a velocity (vector field) $u_t^\theta(\cdot)$, s.t.

$$\frac{d}{dt} x_t = u_t^\theta(x_t), \quad x_0 \sim p$$

results in $X_1 \sim q_*$.

→ (1)

(This is analogous to the DDIM setting that we studied where the vector field was derived so that the marginal density of X_t matched the density of the SDE reverse process.)



$u_t(\cdot)$ is the true velocity field that satisfies (1).

$p_t(\cdot)$ is the associated density, i.e.,

$$X_t \sim p_t(x), \quad X_0 \sim p(x), \quad X_1 \sim q_*(x), \\ t \in [0, 1].$$

From our prior discussion, we know from the continuity equation that $u_t(\cdot)$ and $P_t(\cdot)$ are closely related:

$$\frac{\partial}{\partial t} P_t(x) = -\nabla \cdot (u_t(x) P_t(x))$$

→ ②

Thus, specifying any one of:

- (a) the sample path: X_t , $t \in [0, 1]$
- (b) the density evolution: $P_t(x)$, $t \in [0, 1]$
- (c) the velocity field: $u_t(x)$, $t \in [0, 1]$.

$$x \in \mathbb{R}^d$$

will result in the derivation of the other processes.

There are several possible paths, e.g., specify $u_t(\cdot)$ to be a "minimum energy" path between source and sink (Optimal Transport), specify

the density evolution of $p_t(\cdot)$ to match some baseline (DDPM $\rightarrow$ DDIM), or specify a sample path property ("straight line") as developed below.

We describe the sample path approach below:

Suppose we want: $x_t = (1-t)x_0 + tx_1$,
 $t \in [0, 1]$.

where $x_1 \sim q^*$, $x_0 \sim N(0, I)$

$$x_0 \perp x_1$$



(Other joint dist. based sampling of (x_0, x_1) are possible, e.g., x_0 is a low-res image and x_1 is the corresponding hi-res image; see [4]).

To derive the velocity field, we use a conditioning idea, as described below.

Let us define: $p_{t|1}(x|x_1)$ as the conditional

density of X_t given $X_1 = x_1$. Then, there is a corresponding conditional velocity field $u_t(x|x_1)$ s.t.

$$\frac{\partial}{\partial t} P_{t|1}(x|x_1) = -\nabla \cdot (u_t(x|x_1) P_{t|1}(x|x_1))$$

→ (4)

Also, from Total prob. theorem,

$$P_t(x) = \int P_{t|1}(x|x_1) q_*(x_1) dx_1$$

$\frac{\partial}{\partial t}$ ↗

$$\therefore \partial_t P_t(x) = \int \partial_t P_{t|1}(x|x_1) q_*(x_1) dx_1$$

$$= \int -\nabla \cdot (u_t(x|x_1) P_{t|1}(x|x_1)) q_*(x_1) dx_1$$

$$= -\nabla \cdot \left(\int u_t(x|x_1) P_{t|1}(x|x_1) q_*(x_1) dx_1 \right)$$

From Baye's Rule:

$$P_{i|t}(x_i|x) = \frac{P_{t|i}(x|x_i) q_*(x_i)}{P_t(x)}$$

$$\Rightarrow \therefore \partial_t P_t(x) =$$

$$- \nabla \cdot \left(\underbrace{P_t(x)}_{\text{dummy variable } z} \cdot \int u_t(x|x_i) \underbrace{\frac{P_{t|i}(x|x_i) q_*(x_i)}{P_t(x)}}_{= P_{i|t}(x_i|x)} dx_i \right)$$

$$= - \nabla \cdot \left(P_t(x) E_{z \sim P_{i|t}(\cdot|x)} [u_t(x|z)] \right)$$

Also, from continuity equation,

$$\partial_t P_t(x) = - \nabla \cdot (P_t(x) u_t(x)).$$

$$\Rightarrow u_t(x) = E_{z \sim P_{i|t}(\cdot|x)} [u_t(x|z)].$$

→ (5).

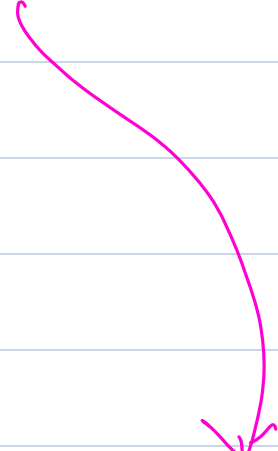
Thus, for the sample path trajectory in (3), the corresponding velocity field is given by (5).

Finally, we can derive $u_t(x|x_1)$ in closed form. From (3), $X_t = (1-t)X_0 + tX_1$,

With $X_1 = x_1$, we thus have:

$$\frac{dX_t}{dt} = -X_0 + X_1,$$

$$\frac{dX_t}{dt} = u_t(X_t|x_1).$$


$$X_0 = \frac{X_t - tX_1}{1-t}$$

$$\Rightarrow u_t(X_t|x_1) = x_1 - X_0 = \frac{x_1(1-t) - X_t + t x_1}{1-t}$$

$$= \left(\frac{x_1 - X_t}{1-t} \right). \quad \rightarrow (6)$$

$\therefore$ Combining (5), (6):

$$u_t(x) = E_{Z \sim p_{1|t}(\cdot|x)} \left[\left(\frac{Z-x}{1-t} \right) \right].$$

$$= \frac{E_{Z \sim p_{1|t}(\cdot|x)} [Z] - x}{1-t}.$$

→ ⑦.

Aside: We see that the velocity field is affine in the conditional expectation, and thus, is also affine in the score (from Tweedie's formula).

Thus, this approach also results in an annealed score based generation analogous to the DDPM/DDIM discussion; see also Note 15 for deriving the explicit stochastic sampler.

Learning $u_t^\theta(\cdot)$ to approximate $u_t(\cdot)$:

The flow matching loss is given by:

$$\mathcal{L}_{\text{FM}}(\theta) = E_{\substack{T \sim \text{Unif}[0,1] \\ X_t \sim P_t(\cdot)}} \left[\|u_T(X_t) - u_T^\theta(X_t)\|^2 \right]$$

→ ⑧

Note that above square loss is generalized to any Bregman Divergence loss; the square loss is a special case.

However, since we do not have a closed-form expression for $u_t(\cdot)$, we instead look at the conditional flow matching loss:

$$\mathcal{L}_{\text{CFM}}(\theta) = E_{\substack{T \sim \text{Unif}[0,1] \\ Z \sim q_Z(\cdot) \\ X_t \sim P_{t|Z}(\cdot|Z)}} \left[\|u_t(X_t|Z) - u_t^\theta(X_t)\|^2 \right]$$

→ ⑨

The squared-loss is a Bregman Divergence. Using the fact that for any Bregman f -Divergence $D_f(\cdot, \cdot)$, we have:

↙ convex f ,

Bregman Divergence: For a convex fn f ,

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

$$E [\nabla_y D_f(z, y)] = \nabla_y D_f(E[z], y) \quad \longrightarrow \quad \text{xx}$$

(see sec 4.5 in [4]), it follows that

$$\nabla_{\theta} d_{FM}(\theta) = \nabla_{\theta} d_{CFM}(\theta).$$

$\longrightarrow$ 10.

Proof:

$$d_{FM}(\theta) = E_{\substack{t \sim \text{Unif}[0,1] \\ X_t \sim P_t(\cdot)}} [\|u_t(X_t) - u_t^{\theta}(X_t)\|^2]$$

$\therefore$

$$\nabla_{\theta} d_{FM}(\theta) = E_{\substack{t \sim \text{Unif}[0,1] \\ X_t \sim P_t(\cdot)}} \left[\nabla_{\theta} \|u_t(X_t) - \underbrace{u_t^{\theta}(X_t)}_y\|^2 \right]$$

$$= E_{\substack{t \sim \text{Unif}[0,1] \\ X_t \sim P_t(\cdot)}} \left[\nabla_y \|u_t(X_t) - \underbrace{u_t^{\theta}(X_t)}_y\|^2 \nabla_{\theta} u_t^{\theta}(X_t) \right]$$

$$u_t(x) = E_{z \sim p_t(\cdot|x)} [u_t(x|z)]. \quad \text{From (5)}$$

$$= E_{\substack{t \sim \text{Unif}[0,1] \\ x_t \sim p_t(\cdot)}} \left[\nabla_y \| E_{z \sim p_t(\cdot|x_t)} [u_t(x_t|z)] - \overbrace{u_t^\theta(x_t)}^y \|^2 \cdot \nabla_\theta u_t^\theta(x_t) \right]$$

$\therefore \|\cdot\|^2$ is a Bregman Div.,
see (xx) above.

$$= E_{\substack{t \sim \text{Unif}[0,1] \\ x_t \sim p_t(\cdot)}} \left[E_{z \sim p_t(\cdot|x_t)} \left[\nabla_y \| u_t(x_t|z) - u_t^\theta(x_t) \|^2 \right] \cdot \nabla_\theta u_t^\theta(x_t) \right]$$

collapse chain rule

$$= E_{\substack{t \sim \text{Unif}[0,1] \\ x_t \sim p_t(\cdot)}} \left[E_{z \sim p_t(\cdot|x_t)} \left[\nabla_\theta \| u_t(x_t|z) - u_t^\theta(x_t) \|^2 \right] \right]$$

$$= E_{\substack{t \sim \text{Unif}[0,1] \\ x_t \sim p_t(\cdot) \\ z \sim p_t(\cdot|x_t)}} \left[\nabla_\theta \| u_t(x_t|z) - u_t^\theta(x_t) \|^2 \right]$$

linearity of expectation

$$= \nabla_\theta E_{\substack{t \sim \text{Unif}[0,1] \\ x_t \sim p_t(\cdot) \\ z \sim p_t(\cdot|x_t)}} \left[\| u_t(x_t|z) - u_t^\theta(x_t) \|^2 \right]$$

$$= \nabla_{\theta} \mathcal{L}_{CFM}(\theta). \quad \square$$

$$\therefore \theta^* \in \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{FM}(\theta)$$

$$\iff \theta^* \in \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{CFM}(\theta)$$

Thus, it suffices to train using (9) to minimize the standard flow matching loss (8).

Finally,

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{\substack{T \sim \text{Unif}[0,1] \\ Z \sim q_{\star}(\cdot) \\ X_t \sim p_t(\cdot|Z)}} \left[\|u_t(X_t|Z) - u_t^{\theta}(X_t)\|^2 \right]$$

Also recall, $X_t = (1-t)X_0 + tX_1$

Therefore, specifying X_t at any two time instances defines the sample path.

Further, from ⑥, we have:

$$u_t(x_t|x_1) = \left(\frac{x_1 - x_t}{1-t} \right).$$

$$(1-\tau)x_0 + \tau x_1$$

$$\therefore \mathcal{L}_{CFM}(\theta) = E_{\substack{\tau \sim \text{Unif}[0,1] \\ x_0 \sim N(0, \Sigma) \\ x_1 \sim q_{\theta}(\cdot)}} \left[\left\| \frac{x_1 - x_{\tau}}{1-\tau} - u_{\theta}(x_{\tau}) \right\|^2 \right]$$

i.e., equivalently,

$$\begin{aligned} (x_1 - x_{\tau}) &= x_1 - x_0 \\ &\quad + \tau x_0 - \tau x_1 \\ &= (1-\tau)(x_1 - x_0). \end{aligned}$$

$$\mathcal{L}_{CFM}(\theta) = E_{\tau, x_0, x_1} \left[\left\| u_{\tau}^{\theta}(x_{\tau}) - (x_1 - x_0) \right\|^2 \right]$$

which is the loss we discussed in Note 15.

⑦