

The Analytical Optimal Velocity in Flow Matching

Wenhan Gao

Department of Applied Mathematics and Statistics, Stony Brook University, NY 11794, USA
`wenhan.gao@stonybrook.edu`

1 The Flow-matching Setup

We are trying to learn a time-dependent velocity field $v(x, t)$ that transports samples from an initial distribution $p_0(x_0)$ to a target distribution $p_1(x_1)$. A sample trajectory is created by interpolating between $x_0 \sim p_0(x_0)$ and $x_1 \sim p_1(x_1)$ through some schedule:

$$x_t = \alpha_t x_0 + \beta_t x_1, \quad (1.1)$$

where $t \in [0, 1]$, $\alpha_0 = \beta_1 = 1$, and $\alpha_1 = \beta_0 = 0$. A common choice is $\alpha_t = 1 - t$ and $\beta_t = t$.

2 The Sample Velocity

The sample velocity of a sample trajectory is the derivative of x_t :

$$v_s(x_0, x_1, t) = \frac{dx_t}{dt} = \alpha'_t x_0 + \beta'_t x_1. \quad (2.1)$$

v_s is sometimes referred to as the conditional velocity given the endpoints x_0, x_1 , and (2.1) is called the flow ODE.

3 The Optimal Velocity Field

We want a function $v(x, t)$, which is the velocity field, that depends only on the current position in space-time, not on which particular x_0, x_1 it starts from.

The training objective is to minimize the MSE:

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_{x_0, x_1, t} \left[\|v(x, t) - v_s(x_0, x_1, t)\|^2 \right] \\ &= \mathbb{E}_{x_t, t} \left[\mathbb{E}_{x_0, x_1 | x_t, t} \left[\|v(x, t) - v_s(x_0, x_1, t)\|^2 \right] \right]. \end{aligned} \quad (3.1)$$

Since the expectation operator is linear, the optimal velocity field is:

$$v^*(x, t) = \mathbb{E}_{x_0, x_1} [v_s(x_0, x_1, t) | x_t = x]. \quad (3.2)$$

Theorem 3.1 (Continuity of Time-marginals). *Under fairly general regularity conditions¹, the optimal velocity field given in (3.2) makes the time-marginals $\{p_t(x)\}$ to satisfy the continuity equation:*

$$\partial_t p_t(x) = -\nabla_x \cdot (p_t(x)v(x, t)). \quad (3.3)$$

¹This write-up is intended as an introduction rather than a fully mathematically rigorous treatment.

Proof. The marginal is:

$$\begin{aligned} p_t(x) &= \iint p_t(x | x_0, x_1) p(x_0, x_1) dx_0 dx_1 \\ &= \iint \delta(x - x_t) p(x_0, x_1) dx_0 dx_1, \end{aligned}$$

where $x_t = \alpha_t x_0 + \beta_t x_1$. Differentiate w.r.t. time t :

$$\begin{aligned} \partial_t p_t(x) &= \iint \partial_t \delta(x - x_t) p(x_0, x_1) dx_0 dx_1 \\ &= \iint \nabla_x \delta(x - x_t) \frac{d}{dt}(x - x_t) p(x_0, x_1) dx_0 dx_1 && \text{by chain rule} \\ &= -\nabla_x \cdot \left(\iint \delta(x - x_t) v_s(x_0, x_1, t) p(x_0, x_1) dx_0 dx_1 \right) && \text{by definition} \\ &= -\nabla_x \cdot \left(\iint v_s(x_0, x_1, t) p(x_0, x_1 | x_t = x) p_t(x) dx_0 dx_1 \right) && \text{by Bayes} \\ &= -\nabla_x \cdot \left(p_t(x) \cdot \iint v_s(x_0, x_1, t) p(x_0, x_1 | x_t = x) dx_0 dx_1 \right) \\ &= -\nabla_x \cdot (p_t(x) \cdot \mathbb{E}_{x_0, x_1} [v_s(x_0, x_1, t) | x_t = x]) \\ &= -\nabla_x \cdot (p_t(x) \cdot v^*(x, t)). \end{aligned}$$

□

As a result, if we start from p_0 , the velocity field defines a valid transport that conserves probability and evolves p_0 exactly into p_1 .

Remark 3.2 (Non-uniqueness of Velocity Fields). There can be infinitely many velocity fields that induce the same density evolution, differing by divergence-free (solenoidal) components. For any v that satisfies the continuity equation, v' also satisfies the continuity equation:

$$v' = v + u, \quad \text{where } \nabla \cdot (p_t u) = 0. \quad (3.4)$$

We can call the optimal velocity field given in (3.2) the conditional expectation velocity field. All different v' produces the same evolving marginals p_t , they differ in how individual particles are coupled between p_0 and p_1 (how $p(\mathbf{x}_0, \mathbf{x}_1)$ is defined). In practice, people usually choose, as in our setup as well, the independent coupling:

$$p(\mathbf{x}_0, \mathbf{x}_1) = p_0(\mathbf{x}_0) p_1(\mathbf{x}_1). \quad (3.5)$$

4 The Analytical Optimal Velocity

The optimal velocity can be expressed analytically involving only p_0 and p_1 . Recall,

$$\begin{aligned} v^*(x, t) &= \mathbb{E}_{x_0, x_1} [v_s(x_0, x_1, t) | x_t = x] \\ &= \int v_s(x_0, x_1, t) p(x_0, x_1 | x_t = x) dx_0 dx_1 \\ &= \int v_s(x_0, x_1, t) \frac{p(x_0, x_1) \delta(x - x_t)}{p_t(x)} dx_0 dx_1 \\ &= \frac{1}{p_t(x)} \int v_s(x_0, x_1, t) p(x_0, x_1) \delta(x - x_t) dx_0 dx_1, \end{aligned} \quad (4.1)$$

where $p_t(x) = \iint \delta(x - x_t) p(x_0, x_1) dx_0 dx_1$ and $x_t = \alpha_t x_0 + \beta_t x_1$.

4.1 An Example of the Analytical Optimal Velocity

Assume the following:

- Gaussian base distribution: $p_0(x_0) = \mathcal{N}(0, I)$.
- A discrete distribution of N data samples s_j : $p_1(x_1) = \sum_{j=1}^N \frac{1}{N} \delta(x_1 - s_j)$.
- Linear schedule: $\alpha = 1 - t, \beta = t \Rightarrow v_s(x_0, x_1, t) = x_1 - x_0$.

Consider $p_t(x)$ first:

$$\begin{aligned} p_t(x) &= \sum_{s_j} \frac{1}{N} \delta(x_1 - s_j) \int p_0(x_0) \delta(x - (1-t)x_0 + tx_1) dx_0 \\ &= \frac{1}{N} \sum_{s_j} \int p_0(x_0) \delta(x - (1-t)x_0 + ts_j) dx_0. \end{aligned} \quad (4.2)$$

The root of the delta constraint $x - (1-t)x_0 + ts_j$ is $x_0^* = \frac{x-ts_j}{1-t}$, we perform change of variable to the delta function:

$$\begin{aligned} p_t(x) &= \frac{1}{N} \sum_{s_j} \int p_0(x_0) \frac{\delta(x_0 - x_0^*)}{\det\left(\frac{\partial}{\partial x_0}(x - (1-t)x_0 + ts_j)\right)} dx_0 \\ &= \frac{1}{(1-t)^d} \frac{1}{N} \sum_{s_j} \int p_0(x_0) \delta(x_0 - x_0^*) dx_0 \quad \text{evaluate det} \\ &= \frac{1}{(1-t)^d} \frac{1}{N} \sum_{s_j} p_0(x_0^*) \quad \text{integrate out } x_0. \end{aligned} \quad (4.3)$$

Similarly,

$$\begin{aligned} v^*(x, t) &= \frac{1}{p_t(x)} \sum_{s_j} \left[\frac{1}{N} \delta(x_1 - s_j) \int (x_1 - x_0) p_0(x_0) \delta(x - (1-t)x_0 + tx_1) dx_0 \right] \\ &= \frac{1}{p_t(x)} \frac{1}{N} \sum_{s_j} \int (s_j - x_0) p_0(x_0) \delta(x - (1-t)x_0 + ts_j) dx_0 \\ &= \frac{1}{p_t(x)} \frac{1}{N} \sum_{s_j} \int (s_j - x_0) p_0(x_0) \frac{\delta(x_0 - x_0^*)}{\det\left(\frac{\partial}{\partial x_0}(x - (1-t)x_0 + ts_j)\right)} dx_0 \\ &= \frac{1}{p_t(x)} \frac{1}{(1-t)^d} \frac{1}{N} \sum_{s_j} \int (s_j - x_0) p_0(x_0) \delta(x - x_0) dx_0 \\ &= \frac{1}{p_t(x)} \frac{1}{(1-t)^d} \frac{1}{N} \sum_{s_j} \left[\left(s_j - \frac{x-ts_j}{1-t} \right) p_0(x_0^*) \right] \\ &= \frac{1}{1-t} \frac{1}{p_t(x)} \frac{1}{(1-t)^d} \frac{1}{N} \sum_{s_j} [(s_j - x) p_0(x_0^*)] \\ &= \frac{1}{1-t} \frac{1}{p_t(x)} \frac{1}{(1-t)^d} \frac{1}{N} \left(\sum_{s_j} (s_j p_0(x_0^*)) - x \sum_{s_j} p_0(x_0^*) \right). \end{aligned} \quad (4.4)$$

Substitute $p_t(x)$ in (4.3) into (4.4):

$$\begin{aligned}
v^*(x, t) &= \frac{1}{1-t} \frac{\sum_{s_j} (s_j p_0(x_0^*)) - x \sum_{s_j} p_0(x_0^*)}{\sum_{s_k} p_0(x_0^*)} \\
&= \frac{1}{1-t} \left(\sum_{s_j} s_j \frac{p_0(x_0^*)}{\sum_{s_k} p_0(x_0^*)} - x \right) && \text{simplify} \\
&= \frac{1}{1-t} \left(\sum_{s_j} s_j \frac{\exp\left(-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2}\right)}{\sum_{s_k} \exp\left(-\frac{1}{2}\frac{(x-ts_k)^2}{(1-t)^2}\right)} - x \right) && p_0 = \mathcal{N}(0, I) \\
&= \frac{1}{1-t} \left(\sum_{s_j} s_j \text{softmax}_{s_j}\left(-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2}\right) - x \right) \\
&= \frac{1}{1-t} \left(\sum_{s_j} w_j(x, t) s_j - x \right),
\end{aligned} \tag{4.5}$$

where $w_j(x, t) = \text{softmax}_{s_j}\left(-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2}\right)$.

Interpretation of this analytical solution.

- $w_j(x, t)$: calculates all pairwise L_2 distances multiplied by a time-dependent factor and turns them into a distribution, representing the posterior weight of the data point s_j given that your sample is observed at position $x_t = x$ at time t . Closer
- $\sum_j s_j w_j(x, t)$: the posterior mean of data samples consistent with $x_t = x$.
- $(\sum_{s_j} w_j(x, t) s_j - x)$: the vector that points x to the posterior mean.
- $\frac{1}{1-t}$: in a single time step integration to solve the ODE, the end point will be the posterior mean.

Notable limits of $w_j(x, t)$.

1. Limit as $t \rightarrow 0$:

$$\lim_{t \rightarrow 0} w_j(x, t) = \lim_{t \rightarrow 0} \text{softmax}_{s_j}\left(-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2}\right). \tag{4.6}$$

Inside the softmax:

$$-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2} \xrightarrow[t \rightarrow 0]{} -\frac{1}{2}x^2. \tag{4.7}$$

That is constant in j , so all components are equal. Hence the softmax gives a uniform distribution:

$$\lim_{t \rightarrow 0} w_j(x, t) = \frac{1}{N},$$

where N is the number of data samples.

2. Limit as $t \rightarrow 1$:

$$\lim_{t \rightarrow 1} w_j(x, t) = \lim_{t \rightarrow 1} \text{softmax}_{s_j}\left(-\frac{1}{2}\frac{(x-ts_j)^2}{(1-t)^2}\right). \tag{4.8}$$

Inside the softmax, let $h = 1 - t \rightarrow 0$:

$$-\frac{1}{2}\frac{(x-(1-h)s_j)^2}{h^2} = -\frac{1}{2}\frac{(x-s_j+hs_j)^2}{h^2}. \tag{4.9}$$

Dominant term as $h \rightarrow 0$:

$$-\frac{1}{2}\frac{(x-s_j)^2}{h^2}. \tag{4.10}$$

In the softmax, the term with the largest exponent (least negative quadratic) dominates, i.e., where $|x - s_j|$ is smallest:

$$\lim_{t \rightarrow 1} w_j(x, t) = \begin{cases} 1, & s_j = \arg \min_k |x - s_k| \\ 0, & \text{otherwise.} \end{cases} \quad (4.11)$$

When $t \rightarrow 0$, the weights are uniform, and the velocity moves toward the mean of the data. When $t \rightarrow 1$, the velocity moves toward the nearest data sample.

Final Note: The analytical optimal solution only generates samples that already exist in the dataset; it does not create novel data. In contrast, a trained velocity field with neural networks can generalize beyond the observed samples.