# Flow Matching Notes

allenuma@usc.com

July 2025

# 1 Introduction

## 1.1 Key idea

Train a generative model to learn a pseudo-time-dependent vector field $\boldsymbol{v}(\boldsymbol{\theta}; \boldsymbol{x}_t, t)$ that approximates the the true flow $\boldsymbol{u}(\boldsymbol{x}, t)$ that transports samples of some prior $p_0$ to the target $p_1$.

## 1.2 Flows

Consider the simple ODE for the desired probability transport

$$\frac{d\psi_t(\boldsymbol{x}_t)}{dt} = \boldsymbol{u}_t(\boldsymbol{x}_t), \quad t \in [0,1] \quad \text{with} \quad \boldsymbol{x}_0 \sim p_0(x) \quad \boldsymbol{x}_1 \sim p_1(x) \tag{1}$$

Given the boundaries, is straightforward to integrate and see the flow solution

$$\boldsymbol{\psi}_t = \boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t(\boldsymbol{x}_1) \tag{2}$$

The sample trajectory is a linear interpolation in pseudo-time from the prior to the target distribution, and the sample velocity is given by

$$\dot{\boldsymbol{\psi}}_t = \boldsymbol{u}_t = \boldsymbol{x}_1 - \boldsymbol{x}_0 \tag{3}$$

## 1.3 Loss Function

In the vanilla case where the posterior is not conditioned on labels, to learn $\boldsymbol{v}_t(\boldsymbol{\theta}; \boldsymbol{x}, t)$ we can regress on the true velocity field $\boldsymbol{u}_t(x, t)$

$$\boldsymbol{v}_t^*(x) := \arg\min_v \mathbb{E}_{x\sim p_t}\left[\|\boldsymbol{v}_t(\boldsymbol{x}) - \boldsymbol{u}_t(\boldsymbol{x})\|^2\right] \tag{4}$$

Lipman provides proof of the conditional flow matching objective (conditional flow i.e. $\boldsymbol{x}_t|\boldsymbol{x}_1$)

$$\mathcal{L}_{\text{CFM}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}_1\sim p_1(\boldsymbol{x}_1), x\sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_1)}[\|\boldsymbol{v}_t(\boldsymbol{\theta}; \boldsymbol{x}_t, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|^2] \tag{5}$$

For *semantic* conditional flow matching, where the posterior is conditioned on labels to generate samples of $p(\boldsymbol{x}|\boldsymbol{y})$ (where $\boldsymbol{y}$ is the label), we can extend the objective to include the expectation over the conditional distribution $p(\boldsymbol{x}|\boldsymbol{y})$ by constructing and training on the joint distribution

$$\mathcal{L}_{\text{SCFM}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}_1\sim p_1(\boldsymbol{x}_1|\boldsymbol{y}_1), \boldsymbol{x}_t\sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_1,\boldsymbol{y}_1)}[\|\boldsymbol{v}_t(\boldsymbol{\theta}; \boldsymbol{x}_t, \boldsymbol{y}_1, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|^2] \tag{6}$$

# 2 Methodology

## 2.1 Training Approach

To construct the training dataset, concatenate the data:

$$\begin{bmatrix} \boldsymbol{x}_0^{(1)} & \boldsymbol{x}_1^{(1)} & \boldsymbol{y}_1^{(1)} & t^{(1)} & \boldsymbol{x}_t^{(1,1)} \\ \boldsymbol{x}_0^{(1)} & \boldsymbol{x}_1^{(1)} & \boldsymbol{y}_1^{(1)} & t^{(2)} & \boldsymbol{x}_t^{(1,2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_0^{(i)} & \boldsymbol{x}_1^{(i)} & \boldsymbol{y}_1^{(i)} & t^{(j)} & \boldsymbol{x}_t^{(i,j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_0^{(N)} & \boldsymbol{x}_1^{(N)} & \boldsymbol{y}_1^{(N)} & t^{(T)} & \boldsymbol{x}_t^{(N,T)} \end{bmatrix} \tag{7}$$

$N$ samples are drawn from each distribution $\boldsymbol{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (standard Gaussian) $\boldsymbol{x}_1 \sim q(\boldsymbol{x}_1)$ (target distribution) and simply paired by index, then expanded along with the intermediate values of $\boldsymbol{x}_t = \psi_t(\boldsymbol{x}_0, \boldsymbol{x}_1, t)$ computed using (2), including the timestep $t \in [0, T]$. $\boldsymbol{y}^{(i)}$, $t^{(j)}$, and $\boldsymbol{x}_t^{(i,j)}$ are passed to the network.

## 2.2 Flow MLP Model

The flow is modeled with a simple MLP as shown below

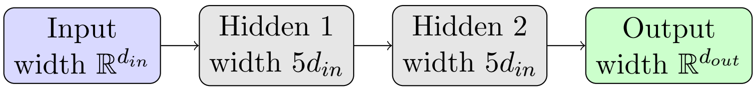


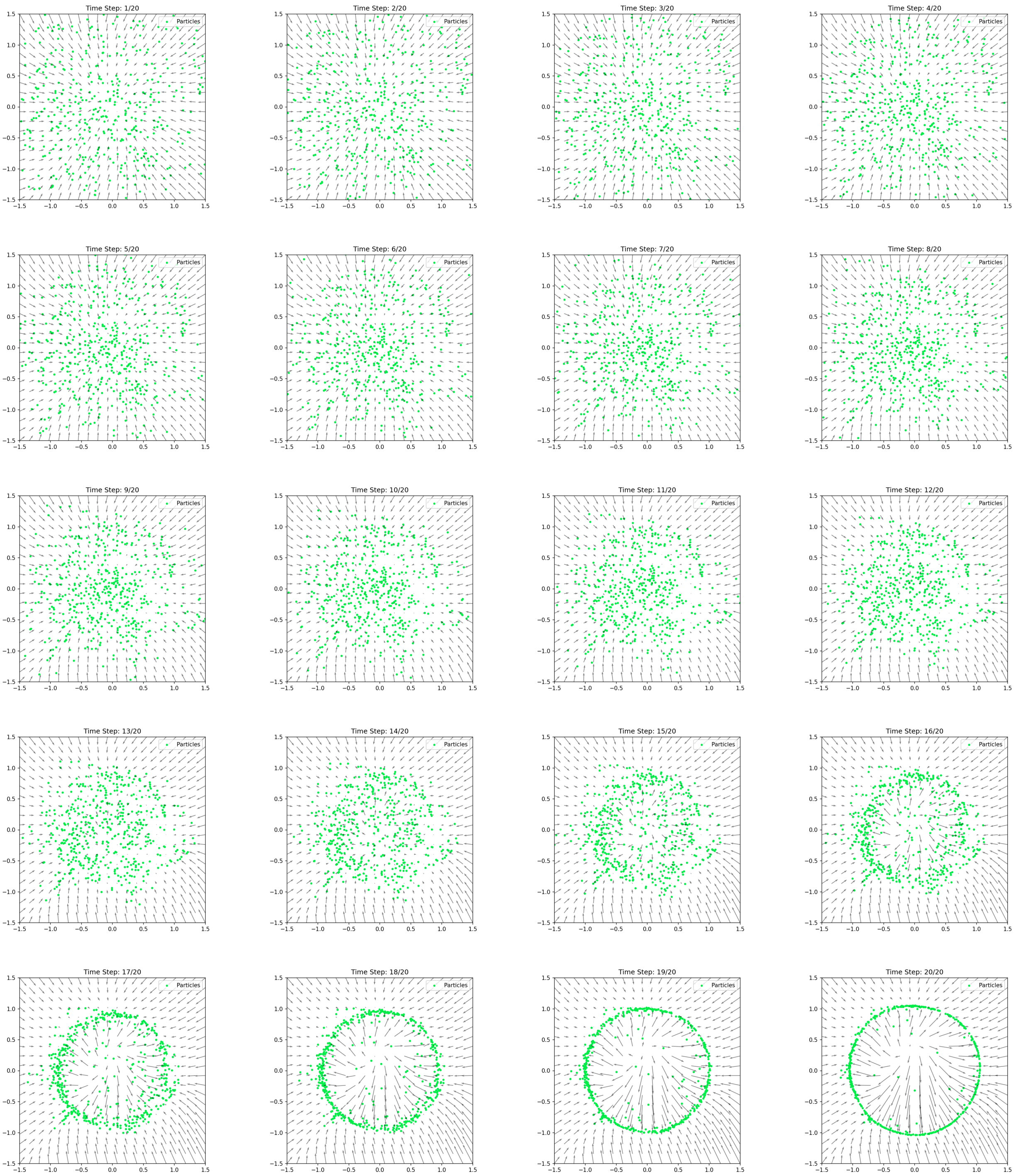

Figure 1: A simple MLP architecture with input size $d_{in}$, two hidden layers, and output size $d_{out}$.

## 2.3 Inference

During inference, for each pseudo-timestep, $\boldsymbol{y}^{(i)}$, $t^{(j)}$, and $\boldsymbol{x}_t^{(i,j)}$ are passed to the model. The network outputs the predicted velocity field $\boldsymbol{v}_{t^{(j)}} = f_{\boldsymbol{\theta}}\left(\boldsymbol{y}_1^{(i)},\ t^{(j)},\ \boldsymbol{x}_t^{(i,j)}\right)$ evaluated at each sample $\boldsymbol{x}_0^{(j)}$. Each sample $\boldsymbol{x}_0^{(j)}$ can then be transported towards the target distribution via forward-Euler integration of the ODE (1).

# 3 Results - Unconditional

**Example results for unconditional flow matching.**
Target distribution samples $\boldsymbol{x}_1 \sim p_1(\boldsymbol{x}_1) = [\cos(2\pi\theta),\ \sin(2\pi\theta)]$, where $\theta \sim \mathcal{U}[0,1]$.
Posterior initialized as $\boldsymbol{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with 512 samples (green scatterplot).
Input vector $\in \mathbb{R}^5$, output vector $\in \mathbb{R}^2$.
Trained on 512 samples for 200 epochs (not plotted).



# 4 Results - Conditioned

**Example results for label-conditioned flow matching.**
Target distribution samples $\boldsymbol{x}_{1|1} \sim p_{X_1|Y_1}(\boldsymbol{x}_1|\boldsymbol{y}_1)$. Marginal prior $\boldsymbol{x}_1 \sim p_1(\boldsymbol{x}_1) = [\cos(2\pi\theta),\ \sin(2\pi\theta)]$, where $\theta \sim \mathcal{U}[0,1]$ (green). Measurement model $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x_1}, \mathbf{I})$ (yellow). Posterior initialized as $\boldsymbol{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (blue). Input vector $\in \mathbb{R}^7$, output vector $\in \mathbb{R}^2$. Trained on 512 samples for 200 epochs.