Flow Matching

Discussion #10

Written by:
Justin, Konpat

1 Warmup: Noise Shapes

At their core, diffusion and flow matching models sample from a base probability *distribution* (usually Gaussian noise) and transport it to a target data distribution.

Note: In this sheet we use the "flow" formulation. Don't confuse it with the "diffusion" one!

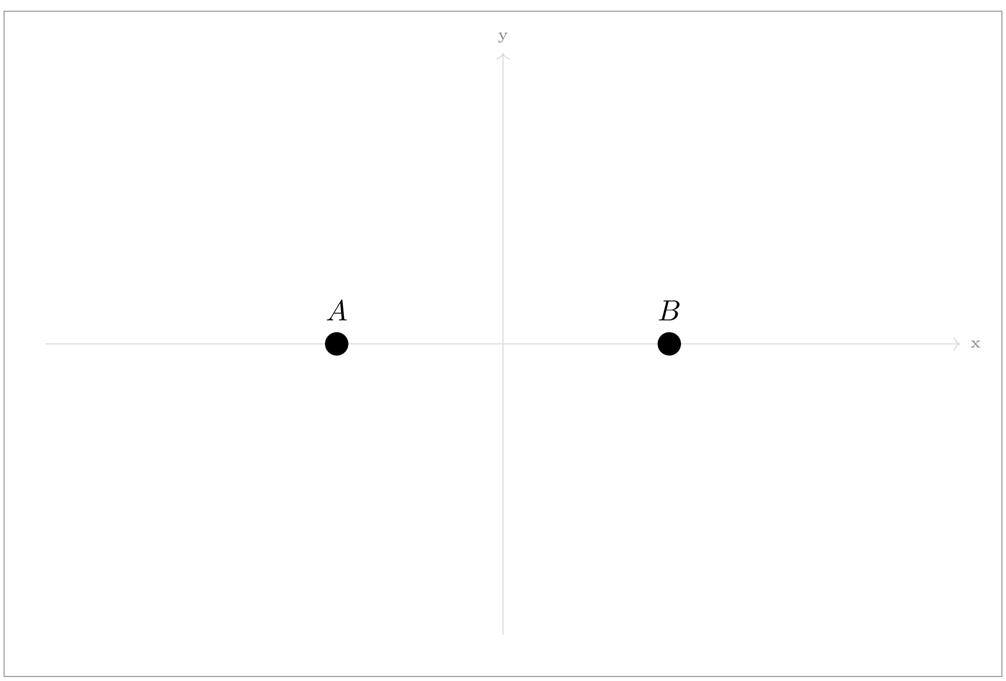
For each task below, state the shape of the input noise x_0 (a.k.a. x_{noise}).

0) Generating a grayscale, 1920×1080 image

1) Generating a random point (x, y) within a square

2 Visualizing Flow in 2D

We want to train a flow matching model to generate samples from a data distribution consisting of exactly two points, $A = (-1, 0)$ and $B = (1, 0)$ (50% probability each).



1.0) Define the noise region. Sketch in the plot above the contour of $\mathcal{N}(0, 1)$ at 1-std deviation (to roughly denote the high probability region of the noise).

Conditional Flow During each iteration of diffusion model training, we select a data point d (either A or B) at random and combine it with random noise r according to timestep $t \in [0, 1]$ like $x_t = d \cdot t + (1 - t) \cdot r$.

1.1 Pick **one** "random" x_0 (full noise) and **plot** them in the chart.

1.2 Say $x_1 = A$, **plot** $x_{0.9}$.

1.3 For all point x_t's, **draw** the velocity vector ($x_{\text{clean}} - x_{\text{noise}}$) at each point.

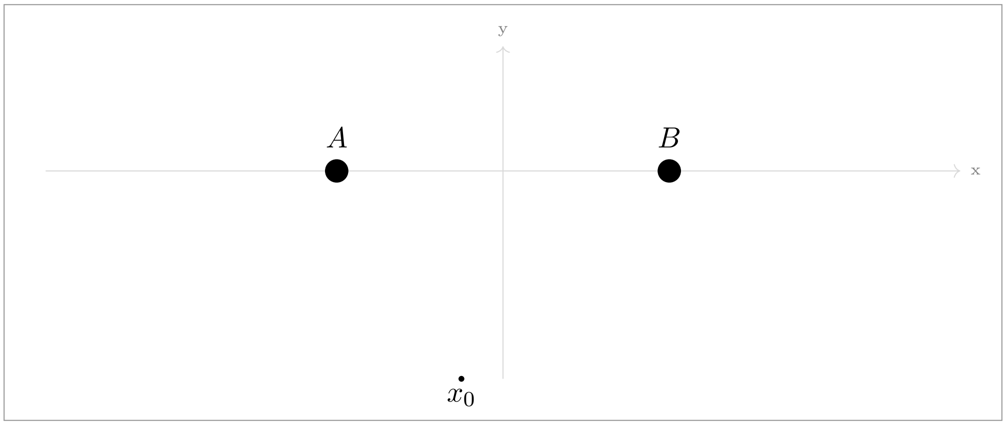
Marginal Flow. Our diffusion model models marginal flow, which for a given x_t is the **weighted average** over all conditional flows from the datapoints we train on.

1.4 For the point x_0, **plot** marginal flows (roughly).

1.5 For the point $x_{0.9}$, **plot** marginal flows (roughly).

1.6 In English, explain the direction of the marginal flow at $t = 0$ and $t = 0.9$.

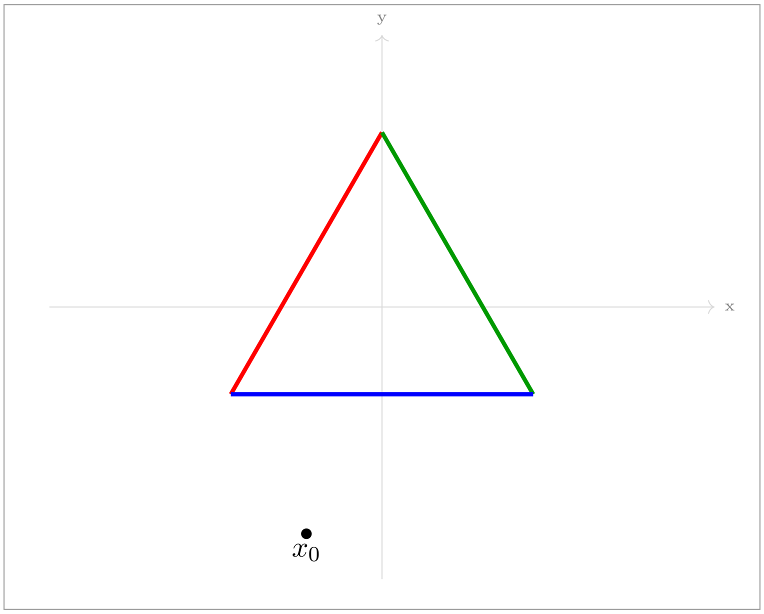
1.7 Start from x_0, draw out the flow sampling procedure with 4 sampling steps. (Hint: follow the marginal flow at each step)



1.8 What if I wanted to generate samples of point A twice as frequently as point B? How could I modify the standard training pipeline to achieve this?

3 Conditioning and Guidance

Now suppose we want to draw from a distribution corresponding to sides of a triangle (think of this like our little image manifold). In addition, we want to *control* what type of point our model generates: red, green, or blue. (not the color, but the location)



3.1 Adding control. The most common way to control diffusion models is by additional *conditioning* input, like language, to the model.

Explain how one could sample from a specific color without additional model input by changing training.

Classifier-Free Guidance (CFG) modifies the vector field to nudge samples more towards red compared to an unconditional sample:

$$\tilde{u_\theta}(x, t, c) = u_\theta(x, t, \varnothing) + w \cdot (u_\theta(x, t, y) - u_\theta(x, t, \varnothing))$$

where u_θ is our neural network model, which learns marginal flow.

3.2 Prompted Flow: Draw in the diagram where $u_\theta(x, t, \textit{red})$ and $u_\theta(x, t, \textit{green})$ point.

3.3 Unprompted Flow: Draw in the diagram where $v(x_0, t, \varnothing)$ points

3.4 Guidance: Draw the CFG vector $\tilde{u_\theta}$ for a large w. How does it differ from the standard prompted vector?