CS 180 Discussion 11

Flow Matching II

Agenda

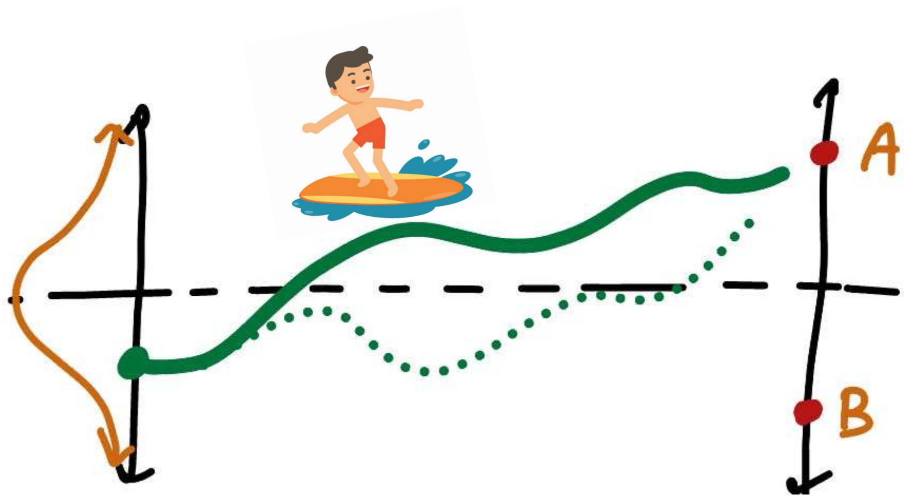
Will post discussion + TA evals form soon™. Please fill it out :)

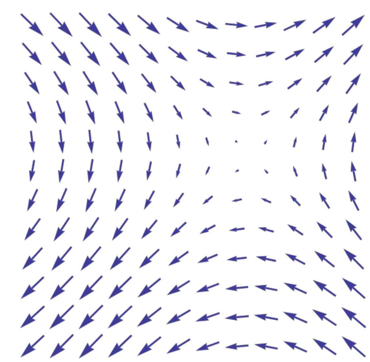
Last short discussion!

1. Flow matching terminology review
2. Sampling
3. AMA if time

Flow matching. Geometrically:

- Trying to go from one distribution to another, by following the flow
- Flow matching model supervises this flow

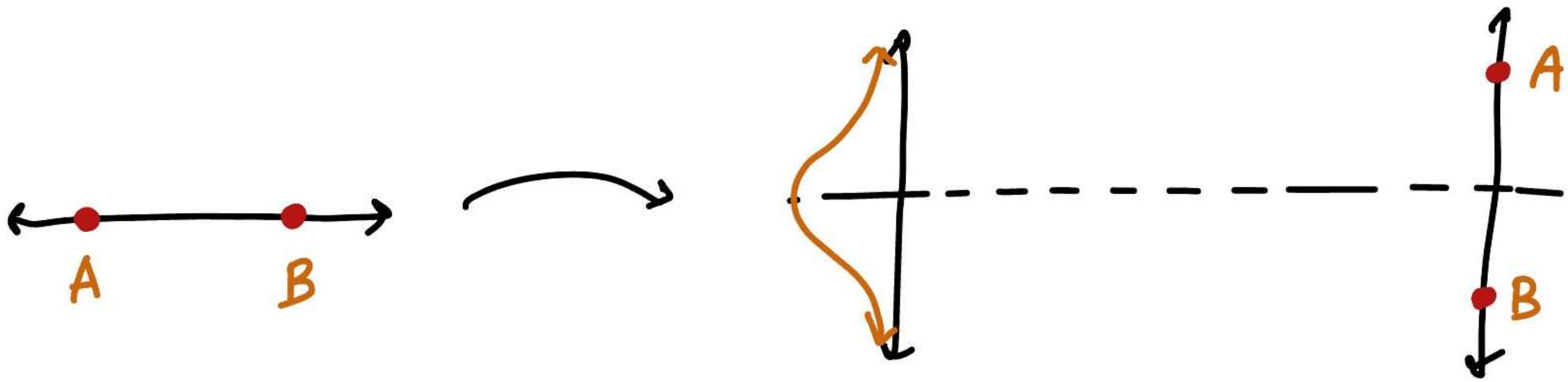




~ Vector field

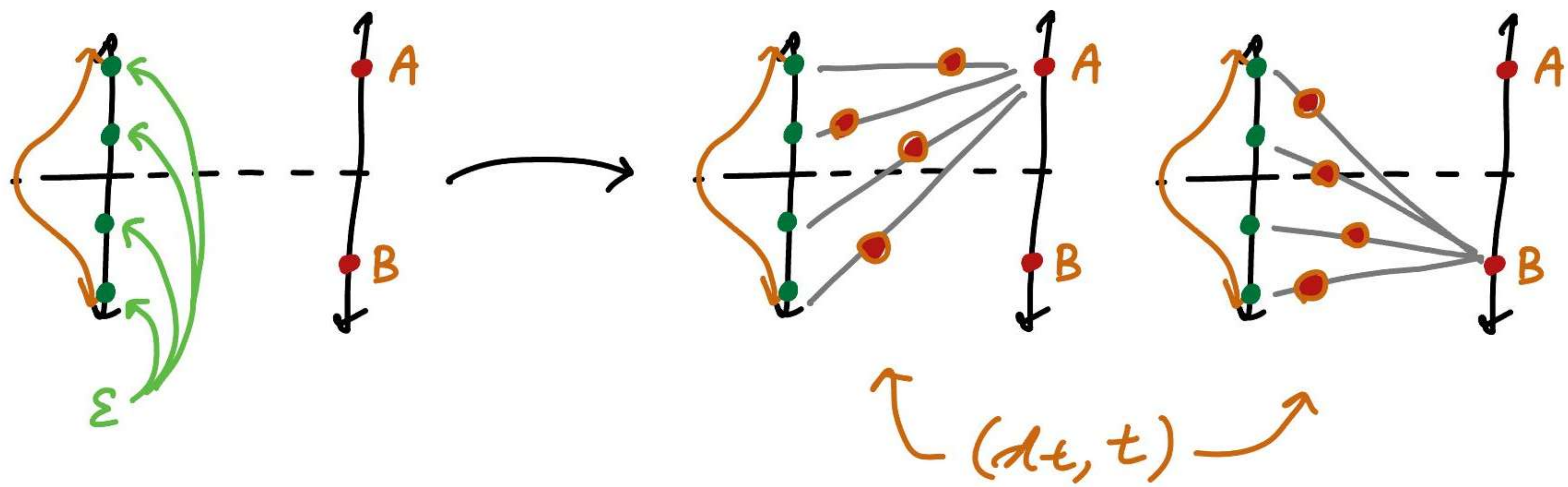
Flow matching. Recall our 1D distribution:

- **x: (clean or noised) data points. Specify the timestep!**
  - For FM: noise is t=0, clean data is t =1
- Last time, we modeled a 1D distribution
- The axes here are for: (1) data, and (2) diffusion timesteps



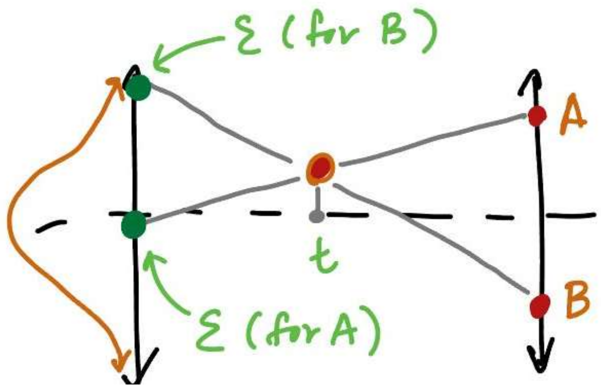
Flow matching. Generating noisy samples

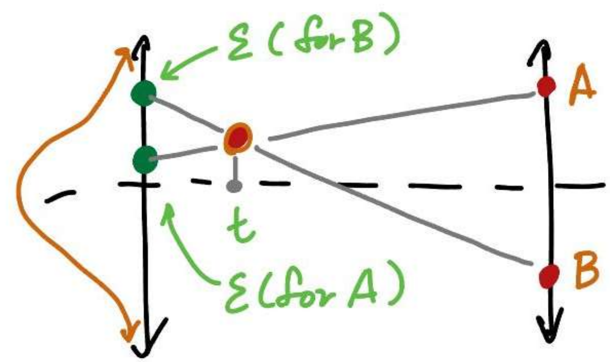
- **epsilon: noise**
- Sample (epsilon, t), forward noising process from x_1 to get (x_t, t)
- Why linear? Using flow matching model, $x_t = (1-t)\epsilon + (t)x_1$



Side note: (x_t, t) can be sampled from multiple points!

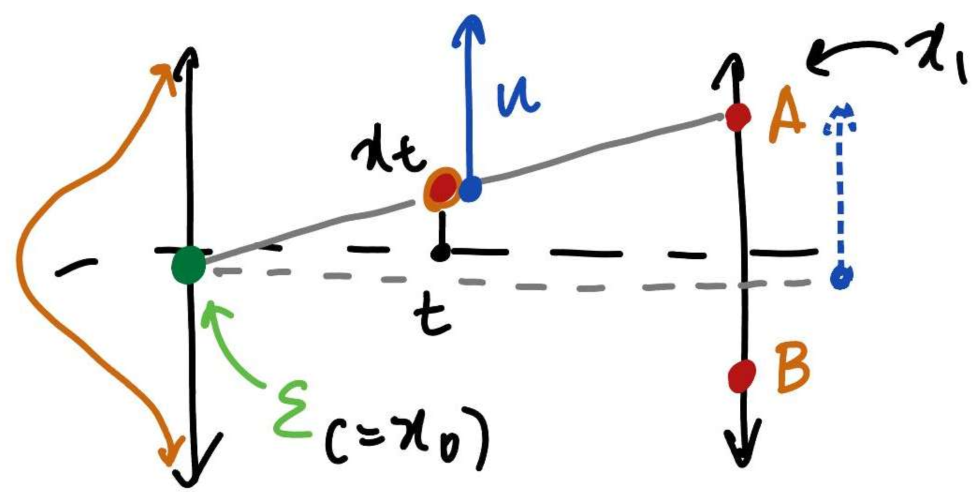
- Just a matter of which epsilon sampled
- Each epsilon has a different likelihood, though (normal distribution)





Flow matching. velocity

- **u: velocity [towards the clean data sample]** $u = x_1 - \epsilon$
- Note the magnitude! (~ velocity != displacement, …)



Flow matching. Term overview

- x: data
    - x_0: fully noised data
    - x_1: clean data (e.g., dataset)
    - x_t: somewhere in between
- epsilon: noise
- u: velocity [towards the data]

$$
\begin{aligned}
u &= x_1 - \epsilon \\
x_t &= (1-t)\epsilon + (t)x_1 \\
&= x_1 - (1-t)u
\end{aligned}
$$

$u(x_t, t)$ or $\epsilon(x_t, t)$ if GT ... $u_\theta(x_t, t)$ or $\epsilon_\theta(x_t, t)$ if predicted from network.

Sampling. Let's briefly talk about training:

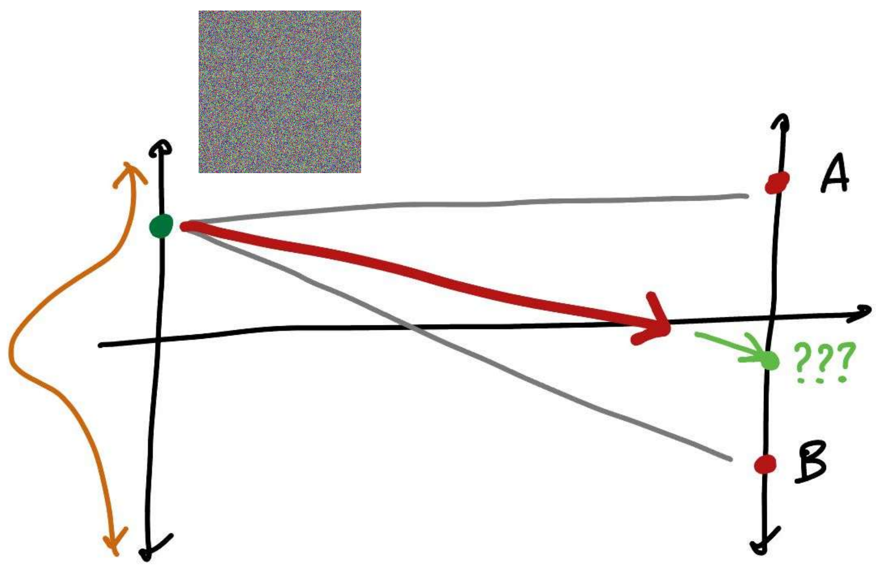
- Model can predict either: u or epsilon
  - Possible to solve for x_1 given (x_t, t) and u OR epsilon
    - Equations:

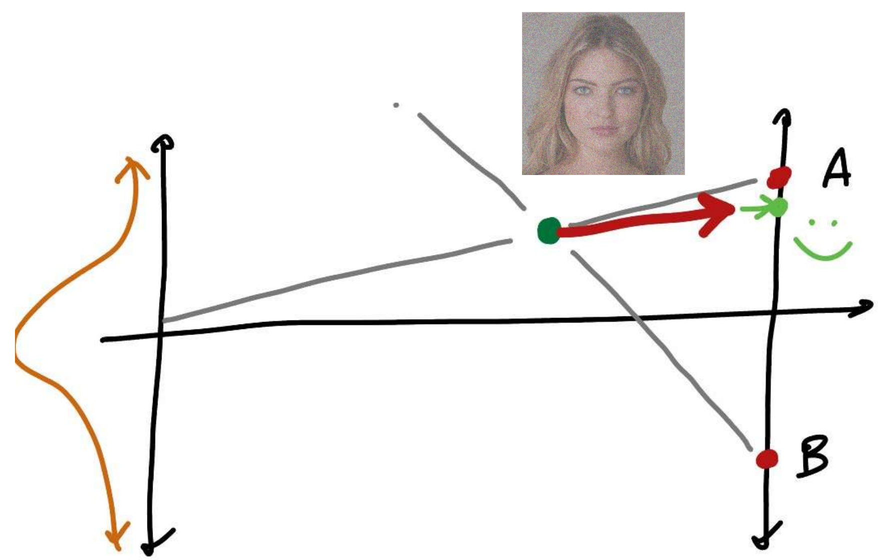
$$\hat{x}_1 = x_t + (1-t)u_\theta(x_t, t) \qquad \hat{x}_1 = \frac{1}{t}(x_t + (1-t)\epsilon_\theta(x_t, t))$$

  - Do you see where things could be unstable?
- Training target: point back towards the data sample

Sampling. Why sample?

- Why? Predictions at small *t* are more ambiguous.





Sampling.

- Many options. One of them is euler integration.

  $$x_{t+\Delta t} = x_t + \Delta t \cdot u_\theta(x_t, t)$$

- What if your network predicts epsilon?

→ Q2!

Q1 + Q2 :-)