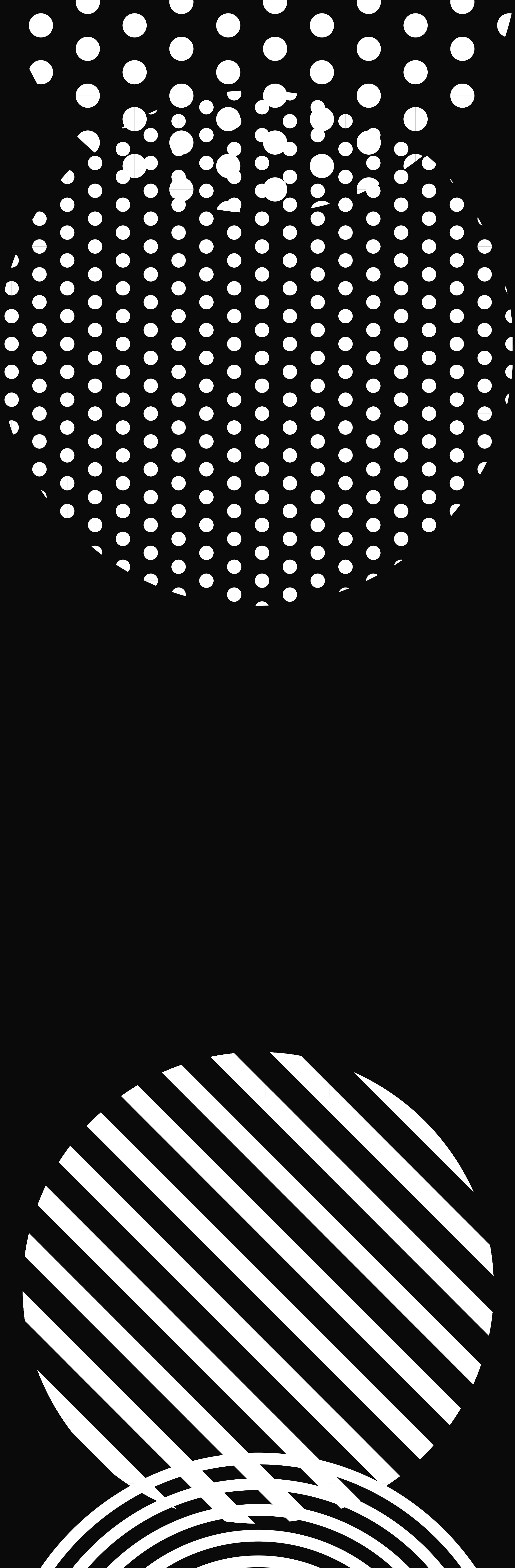


FLOW MATCHING

If x_0 is noise and x_1 is data, the easiest way to go from $x_0 \to x_1$ is linear interpolation:

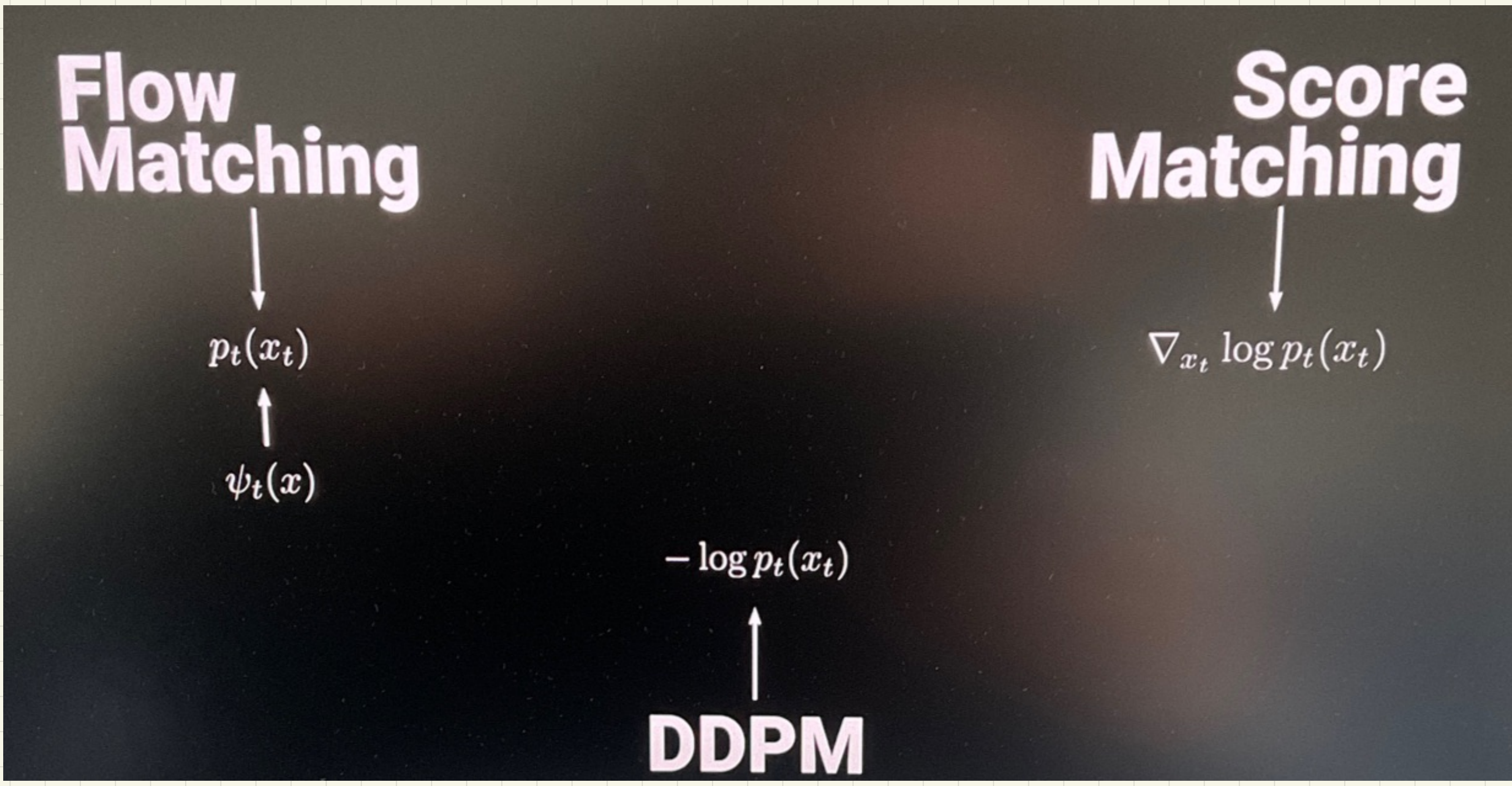
$$x_t = t x_1 + (1-t) x_0 \qquad t \in [0, 1]$$

We are interested in:

$$\frac{dx_t}{dt} \Rightarrow \text{we understand how } x_t \text{ changes varying } t$$

$$\frac{dx_t}{dt} = x_1 - x_0$$

So we can try a NN that given x_t predicts $x_1 - x_0$ and this is the WHOLE idea of flow matching



$\psi_t(x)$ = flow and $x_t = \psi_t(x)$

$$\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x)) = u_t(x_t)$$

So the way x_t changes is determined by u_t which is a vector field

"u_t points into the direction you have to move x_t to get closer to data"

So flow matching learns u_t via

$$v_t \Rightarrow \mathbb{E}_{x_t \sim p_t(x_t)}\left[\| v_t(x_t) - u_t(x_t)\|^2\right]$$

So we have:

$$= \mathbb{E}_{x_t \sim p_t(x_t)} \left[v_t^2(x_t) + u_t^2(x_t) - 2 v_t(x_t) u_t(x_t) \right]$$

we know that $\mathbb{E}_{p(x)} = \int p(x)\, dx$ so:

$$\mathbb{E}_{x_t \sim p_t(x_t)} \left[2 v_t(x_t) u_t(x_t) \right] = 2 \int p_t(x_t) v_t(x_t) u_t(x_t) dx_t$$

we apply marginalisation:

$$u_t(x_t) = \int u_t(x_t | x_1) \underbrace{\frac{p_t(x_t | x_1)\, q(x_1)}{p_t(x_t)}}_{} dx_1$$

density value that tells us if x_t is likely to move towards x_1

weighting term which tells us the influence each data points has.

So we have:

$$= 2 \int v_t(x_t) \frac{\int u_t(x_t | x_1)\, p_t(x_t | x_1)\, q(x_1)\, dx_1}{\cancel{p_t(x_t)}} \cancel{p_t(x_t)}\, dx_t$$

$$= 2 \iint v_t(x_t)\, u_t(x_t|x_1)\, p_t(x_t|x_1)\, q(x_1)\, dx_1\, dx_t$$

Fubini Tonelli theorem

$$= 2\, \mathbb{E}_{x_t \sim p_t(x_t|x_1),\, x_1 \sim q(x_1)} \left[v_t(x) \cdot u_t(x_t|x_1) \right]$$

So we have

$$\mathbb{E}_{x_t \sim p_t(x_t)} \left[\| v_t(x_t) - u_t(x_t) \|^2 \right] =$$

$$= \mathbb{E}_{x_t \sim p_t(x_t),\, x_1 \sim q(x_1)} \left[\| v_t(x_t) \|^2 - 2\, v_t(x_t)\, u_t(x_t|x_1) + \| u_t(x_t) \|^2 \right]$$

Sum and substract $\| u_t(x_t|x_1) \|^2$

$$= \mathbb{E} \left[\| v_t(x_t) \|^2 - 2\, v_t(x_t)\, u_t(x_t|x_1) + \| u_t(x_t) \|^2 + - \| u_t(x_t|x_1) \|^2 + \| u_t(x_t|x_1) \|^2 \right] =$$

$$= \mathbb{E} \left[\| v_t(x_t) - u_t(x_t|x_1) \|^2 + \| u_t(x_t) \|^2 - \| u_t(x_t|x_1) \|^2 \right]$$

$= \mathbb{E}\left[\| v_t(x_t) - u_t(x_t | x_1) \|^2 \right] +$

$+ \underbrace{\mathbb{E}\left[\| \cancel{u_t(x_t)} \|^2 \right] + \mathbb{E}\left[\| \cancel{u_t(x_t | x_1)} \|^2 \right]}$

Do not depend on v_t so they are constants if we want to minimize

$$= \mathbb{E}_{x_t \sim p_t(x_t | x_0),\, x_1 \sim q(x_1)} \left[\| v_t(x_t) - u_t(x_t | x_1) \|^2 \right]$$

$u_t(x_t | x_1)$ is MUCH simpler than $u_t(x_t)$ because it only depends on x_1

THIS IS called CONDITIONAL FLOW MATCHING OBJECTIVE

We remember

$$\frac{d\, \psi_t(x)}{dt} = u_t(\psi_t(x)) = u_t(x_t)$$

We define $\psi_t(x) = \sigma_t(x_1)\, x + \mu_t(x_1)$

They say $x \sim N(0, \sigma I)$ so

$$\left[\psi_t(x_0) = \sigma_t(x_1)\, x_0 + \mu_t(x_1)\right]$$

CANONICAL TRANSFORMATION

[DDPM is non linear interpolation,
FLOW is linear]

Then we have:

$$\mu_t(x_1) = t\, x_1 \qquad \sigma(x_1) = 1 - t$$

So we have

$$\psi_t(x_0) = (1-t)\, x_0 + t\, x_1$$

So we have

$$\mathbb{E}_{x_t \sim p_t(x_t),\, x_1 \sim q(x_1)} \left[\| v_t(x_t) - u_t(x_t \mid x_1) \|^2 \right] =$$

$$= \mathbb{E}\left[\| v_t(x_t) - \frac{d}{dt} \psi_t(x_0) \|^2 \right] = \text{F}$$

but $\frac{d}{dt} = x_1 - x_0$

$x_0 \sim$ random noise $x_1 \sim$ data distribution

$t \sim [0, 1]$

then we give it in input to the network and we solve F