Constructing a training target for Flow & Diffusion model.

Goal: Derive training algorithm. (Get a "good" vector field).

- Minimize MSE.

$$\mathcal{L}(\theta) = \| u_t^\theta(x) - \underbrace{u_t^{\text{target}}(x)}_{\text{Training target.}} \|^2$$

In linear reg, training target = label.

But what is it in our setting?

Today: Get a training target.

Some def:

Make sure you **understand the formulas** for:

Conditional Probability Path	**Conditional Vector Field**	**Conditional Score Function**
Marginal Probability Path	**Marginal Vector Field**	**Marginal Score Function**

Conditional: "Per single data pt."
Marginal: "Across distribution of data pts."

Probability Paths. The path from Noise to Data.

Dirac Distribution: $z \in \mathbb{R}^d$. $X \sim \delta_z \Rightarrow X = z$.

Conditional prob path: $\boxed{p_t(\cdot|z) \quad \text{for fixed } z}$

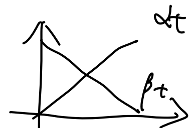
① $p_t(\cdot|z)$ is a distribution over $\mathbb{R}^d$.

② $p_0(\cdot|z) = p_{init}$. $\quad p_1(\cdot|z) = \delta_z$.

Eg. Gaussian prob path.

$$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$

Noise schedulers: α_t, β_t. s.t. $\alpha_0 = 0 \quad \beta_0 = 1$, $\alpha_1 = 1. \quad \beta_1 = 0.$



Eg. $\alpha_t = t, \ \beta_t = 1-t.$

Let's check it fullfill the requirement above.

$$p_0(\cdot|z) = \mathcal{N}(\alpha_0 z, \beta_0^2 I_d) = \mathcal{N}(0, I_d)$$

$$p_1(\cdot|z) = \mathcal{N}(\alpha_1 z, \beta_1^2 I_d) = \mathcal{N}(z, 0) = \delta_z. \quad (\text{Only 1 pt})$$

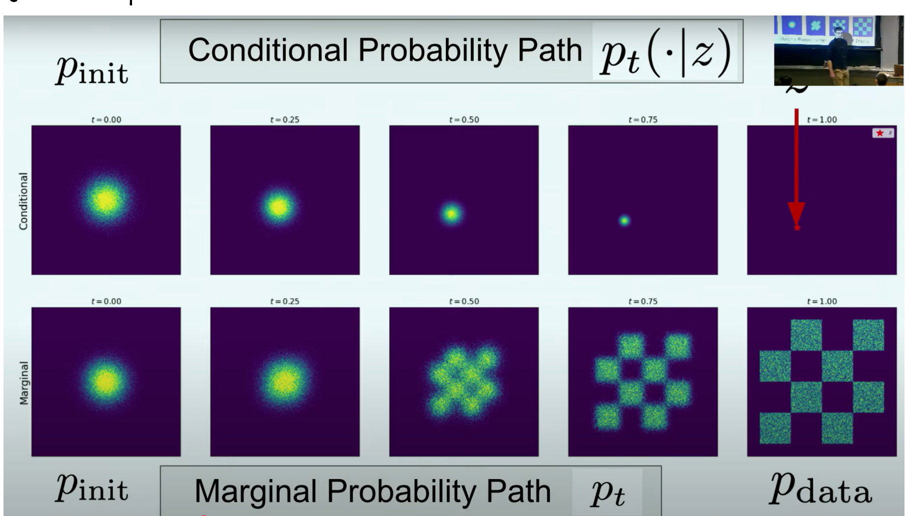
Marginal prob path: $\boxed{p_t(\cdot)}$ (for random $z \sim p_{data}$)

Sample $z \sim p_{data}$, $\quad x \sim p_t(\cdot|z) \Rightarrow x \sim p_t$.

Forget z.

① $p_t(x) = \int p_t(x|z)\, p_{data}(z)\, dz.$

② $p_0 = p_{init}. \quad p_1 = p_{data}.$



3.2 Cond & Marginal Vector Field.

Conditional vector field.

$u_t^{target}(x|z) \qquad 0 \le t \le 1, \; x, z \in \mathbb{R}^d.$

$$p_{init} \xrightarrow{p_t(\cdot|z)} \delta_z$$

ODE.

s.t. $\boxed{X_0 \sim p_{init}. \quad \frac{d}{dt} X_t = u_t^{target}(X_t | z) \Rightarrow X_t \sim p_t(\cdot|z) \qquad (0 \le t \le 1)}$

$\| \; p_0(\cdot|z).$

Follow the simulated ODE, the we get it.

• Z. $\quad p(x) = \sum_z p(x|z)\, p(z).$

Eg. For Gaussian

$$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$

<u>Cond Gaussian VF</u> :

$$\underbrace{u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x}_{\text{A comb of } x \,\&\, z.}$$

$$\dot{\alpha}_t = \frac{d}{dt}\alpha_t. \qquad \dot{\beta}_t = \frac{d}{dt}\beta_t.$$

<u>Thm. Marginalization Trick</u>:

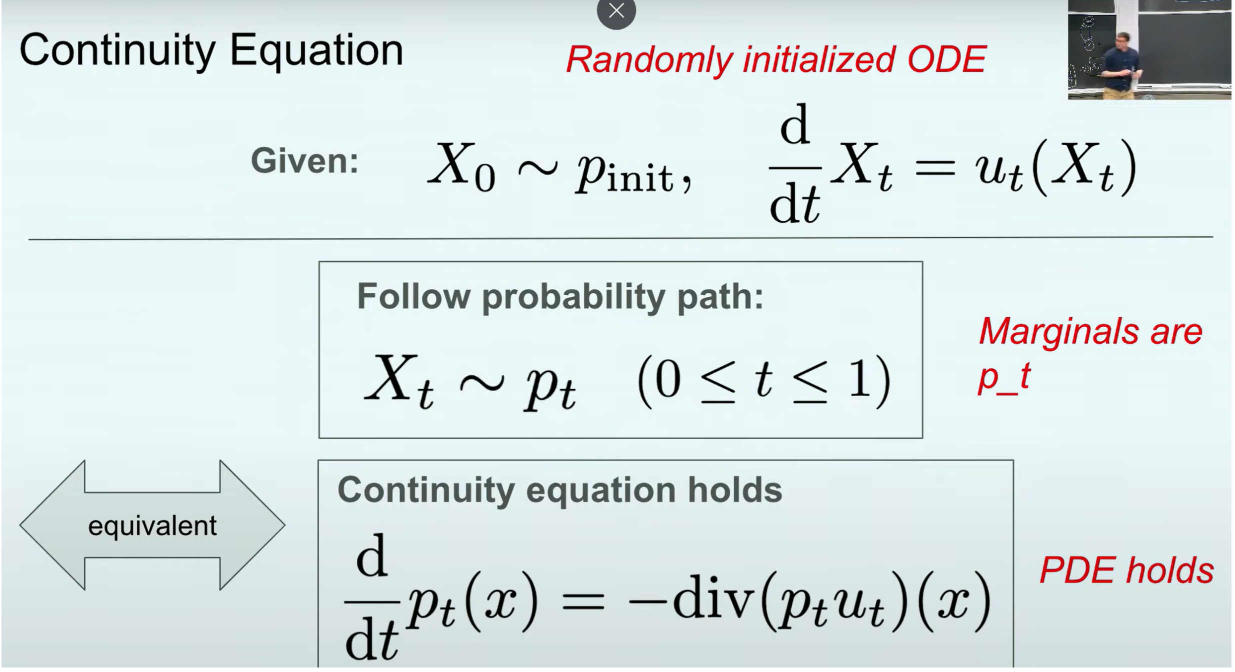
The marginal Vector field by

$$\boxed{u_t^{target}(x) = \int u_t^{target}(x|z) \frac{p_t(x|z)\, p_{data}(z)}{p_t(x)} \, dz.}$$

satisfies

$$\boxed{X_0 \sim p_{init}, \quad \frac{d}{dt} X_t = u_t^{target}(X_t) \Rightarrow X_t \sim p_t. \; (0 \le t \le 1)} \Rightarrow X_1 \sim p_{data}.$$

Just for knowledge:



Random in'it ODE $\Longleftrightarrow$ PDE

$$\mathrm{div}(V_t)(x) = \sum_{i=1}^{d} \frac{\partial}{\partial x_i} V_t(x) \qquad V_t : \mathbb{R}^d \to \mathbb{R}^d.$$

$$\frac{d}{dt} p_t(x) = -\mathrm{div}(p_t u_t)$$

Outflow − inflow.

Change of probability mass at x

pf of Marginalization:

$$\frac{d}{dt} p_t(x) = \frac{d}{dt} \int p_t(x|z)\, p_{data}(z)\, dz = \int \frac{d}{dt} p_t(x|z)\, p_{data}(z)\, dz$$

$$= \int -\mathrm{div}\left(p_t(\cdot|z)\, u_t^{target}(\cdot|z)\right)(x)\, p_{data}(z)\, dz$$

$$= -\mathrm{div}\left(\int p_t(x|z)\, u_t^{target}(x|z)\, p_{data}(z)\, dz\right)$$

(Continuity Equation is a prior knowledge).

$$= -\mathrm{div}\left(p_t(x) \int u_t^{target}(x|z) \frac{p_t(x|z)\, p_{data}(z)}{p_t(x)}\, dz\right)$$

$$= -\mathrm{div}(p_t\, u_t^{target})(x) \quad \text{(As we know)}$$

$$\Rightarrow u_t^{target}(x) =$$

For Cond Gaussian VF

$$u_t^{\text{target}}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$$

Plug this in, $u_t^{\text{target}}(z) = \int - - - \; u_t^{\text{target}}(x|z)\, dz$.

3.3. Cond & Marginal Score.

Conditional Score: $\nabla_x \log p_t(x|z)$

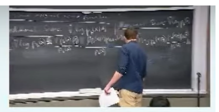
Marginal Score: $\nabla_x \log p_t(x)$

Formula: $$\nabla_x \log p_t(x) = \frac{\nabla_x p_t(x)}{p_t(x)} = \frac{\nabla_x \int p_t(x|z)\, p_{\text{data}}(z)\, dz}{p_t(x)} = \frac{\int \nabla_x p_t(x|z)\, p_{\text{data}}(z)\, dz}{p_t(x)}$$

$$= \int \nabla_x \log p_t(x|z) \frac{p_t(x|z)\, p_{\text{data}}(z)}{p_t(z)}\, dz$$

↑ same trick.

Marginal Prob. Path, Vector Field, and Score



	Notation	Key property	Formula
Marginal Probability Path	p_t	Interpolates p_{init} and p_{data}	$\int p_t(x|z)p_{\text{data}}(z)\mathrm{d}z$
Marginal Vector Field	$u_t^{\text{target}}(x)$	ODE follows marginal path	$\int u_t^{\text{target}}(x|z)\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}\mathrm{d}z$
Marginal Score Function	$\nabla \log p_t(x)$	Can be used to convert ODE target to SDE	$\int \nabla \log p_t(x|z)\frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}\mathrm{d}z$

Eg. For Gaussian Path, $p_t(x|z) = \frac{1}{\sqrt{2\pi}\beta_t} \exp\left(-\frac{(x-\alpha_t z)^2}{2\beta_t^2}\right)$ $\quad \alpha_t z$, $\beta_t^2 I_d$

Cond score $= \nabla_x \log p_t(x|z) = \nabla_x\left(-\frac{(x-\alpha_t z)^2}{2\beta_t^2} - \log(\sqrt{2\pi}\beta_t)\right) = \boxed{-\frac{x-\alpha_t z}{\beta_t^2}}$ Gaussian Score.

Why score is Useful ?

$\boxed{\text{Thm. SDE extension trick.}}$

Let $u_t^{target}(x)$ be as before. Then for any $\sigma_t \geq 0$:

$$X_0 \sim p_{init}, \quad dX_t = \Big[u_t^{target}(X_t) + \underbrace{\frac{\sigma_t^2}{2}\nabla \log p_t(X_t)}\Big]dt + \underbrace{\sigma_t \, dW_t}$$

$\Rightarrow X_t \sim p_t.$

diffusion term

Add correction term.

$\Rightarrow X_1 \sim p_{data}.$

Verify that $u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right)z + \frac{\dot{\beta}_t}{\beta_t}x$ is a solution (a valid cond VF)

to $\boxed{\begin{cases} X_0 \sim \mathcal{N}(0, I_d) = p_{init} \\ \frac{d}{dt}X_t = u_t^{target}(X_t|z) \end{cases} \Rightarrow X_t \sim \mathcal{N}(\alpha_t z, \beta_t^2 I_d)}$

(Gaussian)

Example 11 (Target ODE for Gaussian probability paths)
As before, let $p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$ for noise schedulers α_t, β_t (see eq. (16)). Let $\dot{\alpha}_t = \partial_t \alpha_t$ and $\dot{\beta}_t = \partial_t \beta_t$ denote respective time derivatives of α_t and β_t. Here, we want to show that the **conditional Gaussian vector field** given by

$$u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right)z + \frac{\dot{\beta}_t}{\beta_t}x \tag{21}$$

is a valid conditional vector field model in the sense of theorem 10: its ODE trajectories X_t satisfy $X_t \sim p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$ if $X_0 \sim \mathcal{N}(0, I_d)$. In fig. 6, we confirm this visually by comparing samples from the conditional probability path (ground truth) to samples from simulated ODE trajectories of this flow. As you can see, the distribution match. We will now prove this.

Proof. Let us construct a conditional flow model $\psi_t^{target}(x|z)$ first by defining

Directly construct a solution for $\begin{cases} X_0 \sim \mathcal{N}(0, I_d) \\ X_t \sim \mathcal{N}(\alpha_t z, \beta_t^2 I) \end{cases}$ s.t. $\alpha_0 = 0 \quad \beta_0 = 1$, $\alpha_1 = 1 \quad \beta_1 = 0.$

$$\psi_t^{target}(x|z) = \alpha_t z + \beta_t x. \tag{22}$$

(This is not a ODE, for now).

If X_t is the ODE trajectory of $\psi_t^{target}(\cdot|z)$ with $X_0 \sim p_{init} = \mathcal{N}(0, I_d)$, then by definition

$$X_t = \psi_t^{target}(X_0|z) = \alpha_t z + \beta_t X_0 \sim \mathcal{N}(\alpha_t z, \beta^2 I_d) = p_t(\cdot|z).$$

We conclude that the trajectories are distributed like the conditional probability path (i.e, eq. (18) is fulfilled). It remains to extract the vector field $u_t^{target}(x|z)$ from $\psi_t^{target}(x|z)$. By the definition of a flow (eq. (2b)), it holds

Then, we want it to be a solution of ODE:

$$\frac{d}{dt}\psi_t^{target}(x|z) = u_t^{target}(\psi_t^{target}(x|z)|z) \quad \text{for all } x, z \in \mathbb{R}^d$$
$$\overset{(i)}{\Leftrightarrow} \quad \dot{\alpha}_t z + \dot{\beta}_t x = u_t^{target}(\alpha_t z + \beta_t x|z) \quad \text{for all } x, z \in \mathbb{R}^d$$
$$\overset{(ii)}{\Leftrightarrow} \quad \dot{\alpha}_t z + \dot{\beta}_t\left(\frac{x - \alpha_t z}{\beta_t}\right) = u_t^{target}(x|z) \quad \text{for all } x, z \in \mathbb{R}^d$$
$$\overset{(iii)}{\Leftrightarrow} \quad \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right)z + \frac{\dot{\beta}_t}{\beta_t}x = u_t^{target}(x|z) \quad \text{for all } x, z \in \mathbb{R}^d$$

where in (i) we used the definition of $\psi_t^{target}(x|z)$ (eq. (22)), in (ii) we reparameterized $x \to (x - \alpha_t z)/\beta_t$, and in (iii) we just did some algebra. Note that the last equation is the conditional Gaussian vector field as we defined in eq. (21). This proves the statement.[a] □

[a]One can also double check this by plugging it into the continuity equation introduced later in this section.