Constructing Training Target:

Training: Finding Parameters θ

Flow Model:

Initialize:

$X_0 \sim p_{init}$

↓

Gaussian

ODE:

how?

$dX_t = u_t^{\theta}(X_t)\,dt \longrightarrow X_1 \sim p_{data}$

↓

Neural Network vector field.

Goal: Derive Training Target: $u_t^{target}(x)$

↳ we want to train model by minimizing a mean squared error:

$$L(\theta) = \| u_t^{\theta}(x) - u_t^{\theta\, target}(x) \|^2$$

↑

Training Target → what is? In Classification, Labels are training targets → Nevertheless, here we don't have labels

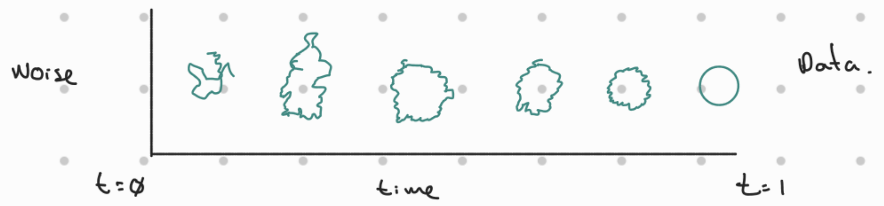
Key terminology:

1) "Conditional": "Per single Data Point" (single)

2) "Marginal": "Across Distribution of Data Points" (many)

Probability Paths:

↳ Interpolation between noise Dist and Data Dist. through time Points.



3.1) Conditional and Marginal Probability Paths

* Dirac Distribution: $z \in \mathbb{R}^d$, δ_z $\quad X \sim \delta_z \Rightarrow X = z$

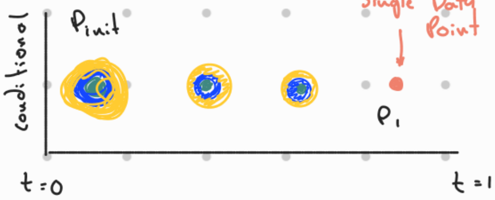
z ↓ Vector; $\mathbb{R}^d$ ← 2u

↳ Always Return Same Point.

↳ Dist. Returns same object

* Conditional Prob. Path:

$p_t(\cdot \mid z)$ — $\cdot$ ↓ Placeholder; z → Data Point



* Dist. Collapses into Data Point *

1) $p_t(\cdot \mid z)$: Distributions are in $\mathbb{R}^d$ → Probability

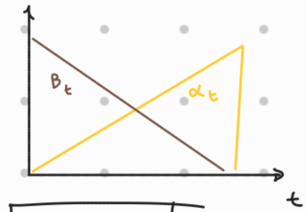
2) $p_0(\cdot \mid z) = p_{init}$, $p_1(\cdot \mid z) = \delta_z$ (Initial Dist; Dirac Dist)

Ex. – Gaussian Probability Path:

$$p_t(\cdot \mid z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$; where α_t, β_t are Noise schedulers

Start: $\alpha_0 = \emptyset$ $\quad \beta_0 = 1$

End: $\alpha_1 = 1$ $\quad \beta_1 = \emptyset$



$\boxed{\alpha_t = t \quad \beta_t = 1 - t}$

$\boxed{\text{* SMP: } p_{init} = \mathcal{N}(0, I_d)}$

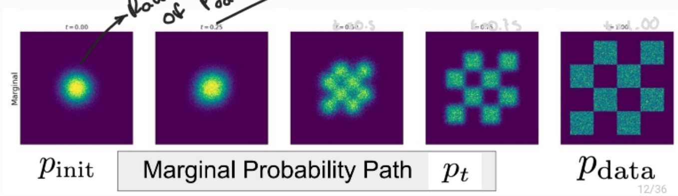
* Marginal Probability Path: $\boxed{p_t}$ ↗ From Data Dist. and CONDITIONAL PROB PATH.

$\underbrace{z \sim p_{data}}_{\text{Random Initial Point}}$, $\underbrace{x \sim p_t(\cdot \mid z)}_{\text{Cond: tional Prob. Path}} \Rightarrow \underbrace{x \sim p_t}_{\text{wo } z}$] how to Sample.

① $p_t(x) = \int p_t(x \mid z)\, p_{data}(z)\, dz$] Density

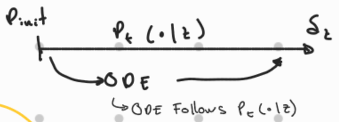
② $p_0 = p_{init}$, $p_1 = p_{data}$] → From Noise to Data

② ⟶ ⊗ ⤴ what we are Doing



3.2 Conditional and marginal Vector Fields.

① Conditional Vector field: $u_t^{target}(x \mid z)$ $\begin{pmatrix} 0 \le t \le 1 \\ x, z \in \mathbb{R}^d \end{pmatrix}$



↳ ODE Follows $p_t(\cdot \mid z)$ with vector field!

Such that

$$\underbrace{x_0 \sim p_{init},}_{\text{Initial Point}} \; \underbrace{\frac{d}{dt} X_t = u_t^{target}(x_t \mid z)}_{\text{we follow vector field}} \Rightarrow \underbrace{X_t \sim p_t(\cdot \mid z)}_{\text{Any Point } X_t \text{ is given by Prob. Path}} \quad 0 \le t \le 1$$

$\boxed{p_0(\cdot \mid z)}$

→ Condition

↓

how construct?

Ex 0.1 - Gaussian:

Last Ex.

Note: $\alpha_t = t$; $\beta_t = 1-t$

↳ Cond. Gaussian Vector Field:

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

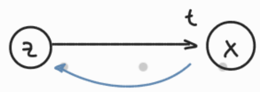
scalar, scalar, vector, ↳ scalar

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

Marginal vector field:

MOST IMP FORMULA!!

$$↳ \; u_t^{target}(x) = \int \underbrace{u_t^{target}(x \mid z)}_{\text{conditional vector field}} \frac{p_t(x \mid z)\, p_{data}(z)}{p_t(x)} \, dz$$



↳ Campo Velocidad marginal como una media Ponderada de campos Velocidad condicionales $u_t^{target}(x \mid z)$, donde z es un vector (x_0, x_1) y Determinan trayectoria interpolación $x(t)$.

Ex.

If I follow this ODE with marginal vector field → then I follow marginal Path!

↳ Satisfies:

$$X_0 \sim p_{init}, \; \frac{d}{dt} X_t = u_t^{target}(x_t) \longrightarrow X_t \sim p_t \quad (0 \le t \le 1)$$

start p_0, ends p_1!

↳ then $X_1 \sim p_{data}$

Conditional Prob. Path, Vector Field, and Score

	Notation	Key property	Gaussian example
Conditional Probability Path	$p_t(\cdot\|z)$	Interpolates p_{init} and a data point z	$\mathcal{N}(\alpha_t z, \beta_t^2 I_d)$
Conditional Vector Field	$u_t^{\text{target}}(x\|z)$	ODE follows conditional path	$\left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$

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	$\int u_t^{\text{target}}(x\|z) \frac{p_t(x\|z) p_{\text{data}}(z)}{p_t(x)} \mathrm{d}z$

Diffusion models

↳ Not necessary but useful to compare

Conditional and marginal score function:

→ Conditional Score: $\underbrace{\nabla_x}_{\text{Prob. Path}} \overset{\text{Gradient}}{\log} \underbrace{p_t(x|z)}_{\text{likelihood}}$ → Marginal Score: $\nabla \log \underbrace{p_t(x)}_{\text{marginal Path.}}$

→ Formula:

↳ $\nabla \log p_t(x) = \dfrac{\nabla p_t(x)}{p_t(x)} = \dfrac{\int \nabla p_t(x|z) p_{data}(z)\, dz}{p_t(x)} = \int \nabla \log p_t(x|z) \dfrac{p_t(x|z) p_{data}(z)}{x} dx$

→ Is above!

↳ trick: $\left(\frac{d}{dx} \log x = \frac{1}{x}\right)$

↳ Implies:

$$\nabla \log p_t(x|z) = \frac{\nabla p_t(x|z)}{p_t(x|z)}$$

→ Density Gaussian Score (Path):

$$\nabla \log p_t(x|z) = - \frac{x - \alpha_t z}{\beta_t^2}$$

why is Imp?

Thm (SDE extension trick): Let $u_t^{target}(x)$ be as before

Then for any $\sigma_t \geq \emptyset$:

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

$$\Rightarrow X_t \sim p_t \quad (\emptyset \leq t \leq 1)$$

↳ $\boxed{X_1 \sim p_{data}}$

Does same thing as Flows

↓

Best!!