Lecture 2

Constructing a Training Target for Flow and Diffusion Models

MIT IAP 2025 | Jan 22, 2025

Peter Holderrieth and Ezra Erives

Sponsor: Tommi Jaakkola

Reminder: Flow and Diffusion Models

Flow Model

Initialize: $X_0 \sim p_{\text{init}},$

ODE: $\mathrm{d}X_t = u_t^\theta(X_t)\mathrm{d}t$

E.g. Gaussian

Neural network vector field

Diffusion coeff.

Diffusion Model

Initialize: $X_0 \sim p_{\text{init}},$

SDE: $\mathrm{d}X_t = u_t^\theta(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$

To get samples, simulate ODE/SDE from t=0 to t=1 and return X_1

Next step: Training the model

Without training, the model produces "non-sense"→ **We need to train** u_t^θ

Training = Finding parameters θ **such that**

$$X_0 \sim p_{\text{init}}, \quad \mathrm{d}X_t = u_t^\theta(X_t)\mathrm{d}t \quad \xrightarrow{\text{Implies}} \quad X_1 \sim p_{\text{data}}$$

Start with initial distribution

Follow along the vector field

The distribution of the final point = data distribution

Goal of lecture 2 (today) and lecture 3 (tomorrow):

Derive training algorithm

Today's goal: Derive a Training Target

- Typically, we train the model by minimizing a **mean squared error**:

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

Training target (→ u_t^{target})

- In regression or classification, the training target is the label.
- Here: No label :(→ We have to **derive a training target**

Today: Derive a formula for the training target: $u_t^{\text{target}}(x)$

Tomorrow: Training algorithm using $u_t^{\text{target}}(x)$

Today: Training target

Tomorrow: Training algorithm

Marginal Vector Field → *Flow Matching*

Marginal Score Function → *Score Matching*

Section 2:

Constructing a Training Target

Goal: Derive a formula for a training target for training our models

Today will be the **technically most challenging lecture!**

The next lectures will **be much much easier!**

You do not have to understand the derivations.

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

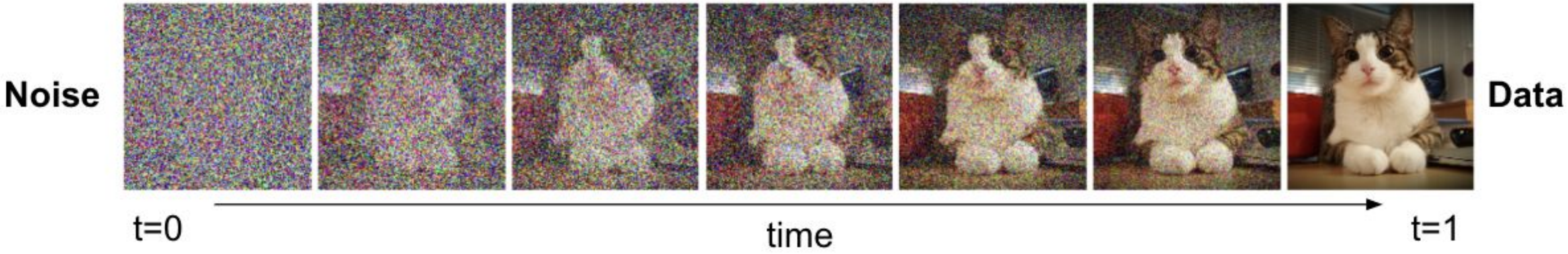
Conditional **Probability Path**	**Conditional** **Vector Field**	**Conditional** **Score Function**
Marginal **Probability Path**	**Marginal** **Vector Field**	**Marginal** **Score Function**

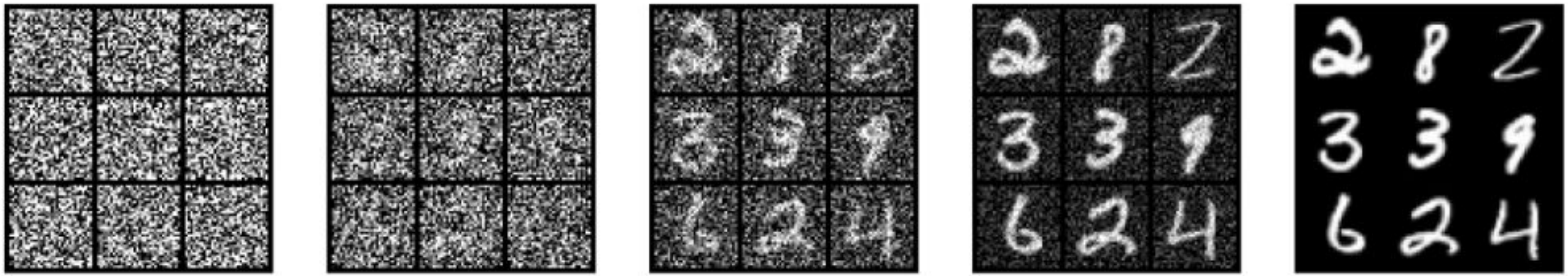
Key terminology:

“Conditional” = “Per single data point”

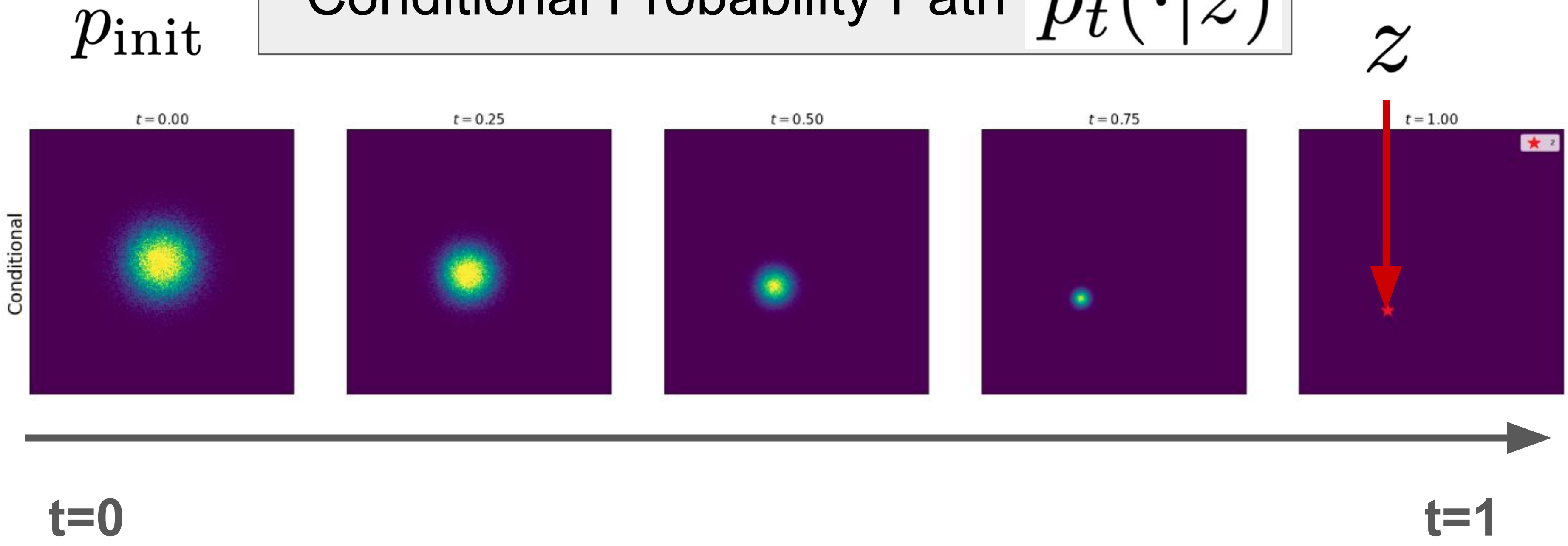
“Marginal” = “Across distribution of data points”

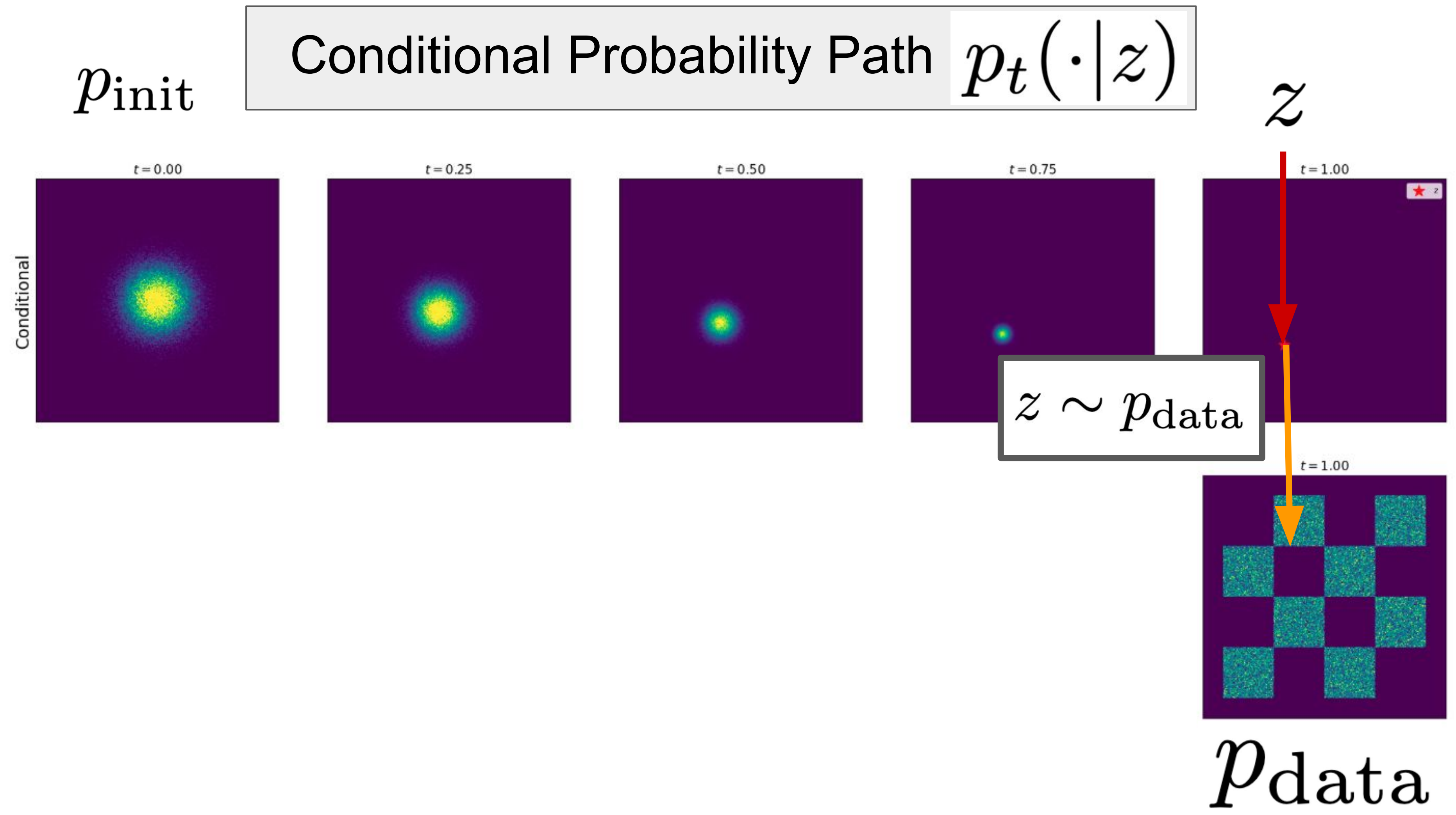
Probability Paths: The Path from Noise to Data

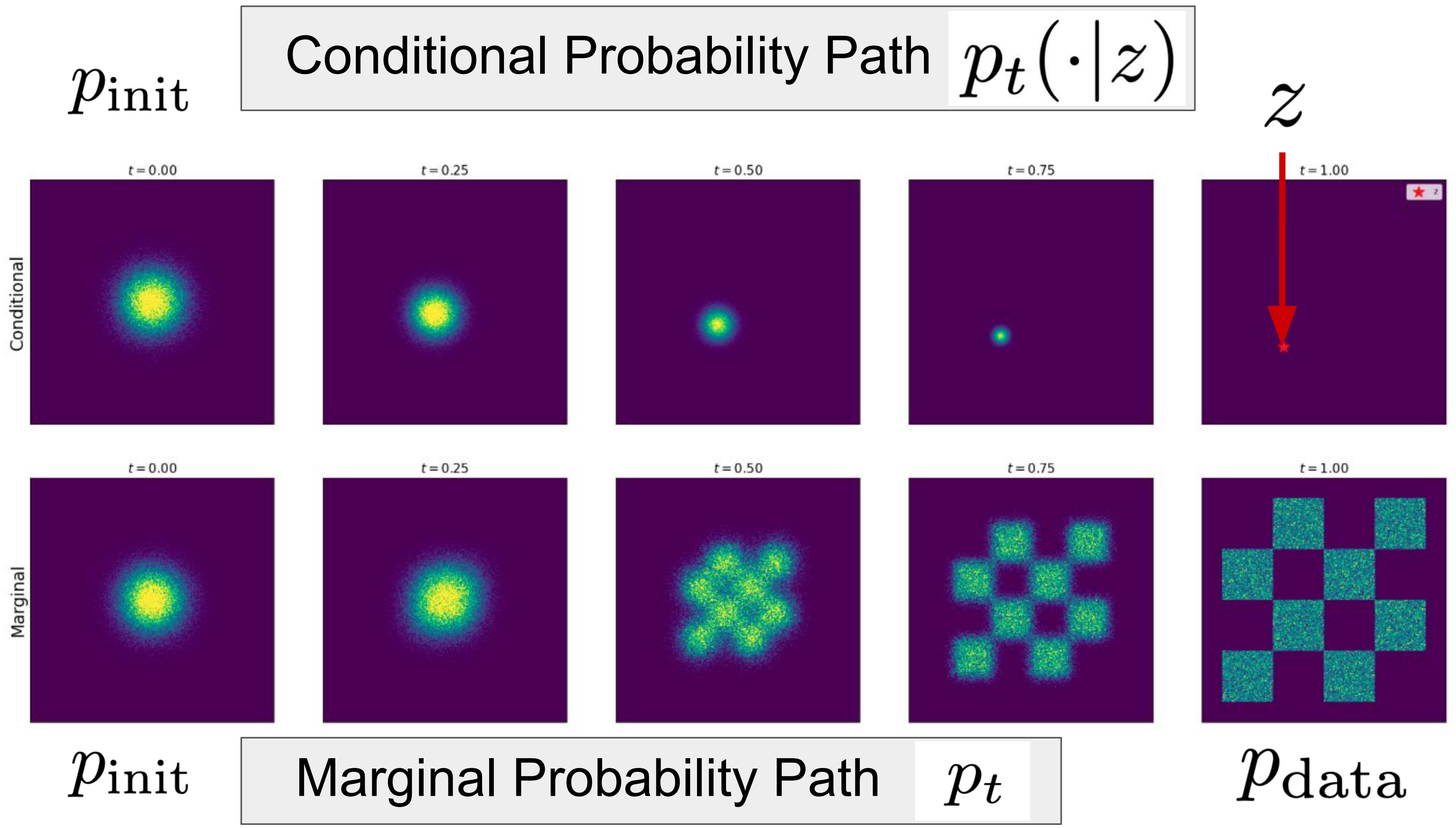




Conditional Probability Path $p_t(\cdot|z)$




Conditional Probability Path $p_t(\cdot|z)$
p_{init}
z
t = 0.00
t = 0.25
t = 0.50
t = 0.75
t = 1.00
Conditional
$z \sim p_{\text{data}}$
t = 1.00
p_{data}


Conditional Probability Path $p_t(\cdot|z)$
p_{init}
z
t = 0.00
t = 0.25
t = 0.50
t = 0.75
t = 1.00
Conditional
Marginal
p_{init}
Marginal Probability Path p_t
p_{data}

Conditional Prob. Path

	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			
Conditional Score Function			

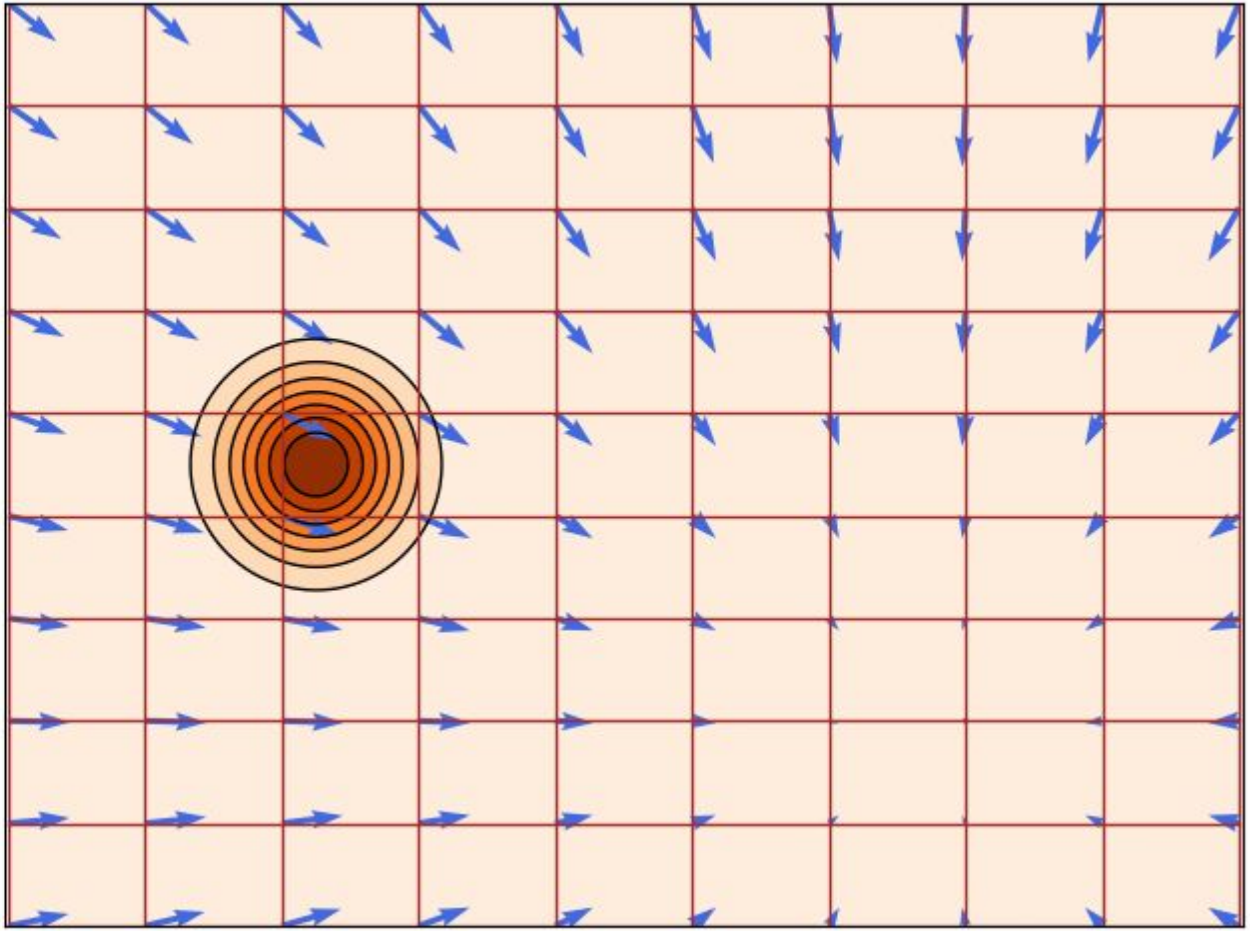
Marginal Prob. Path

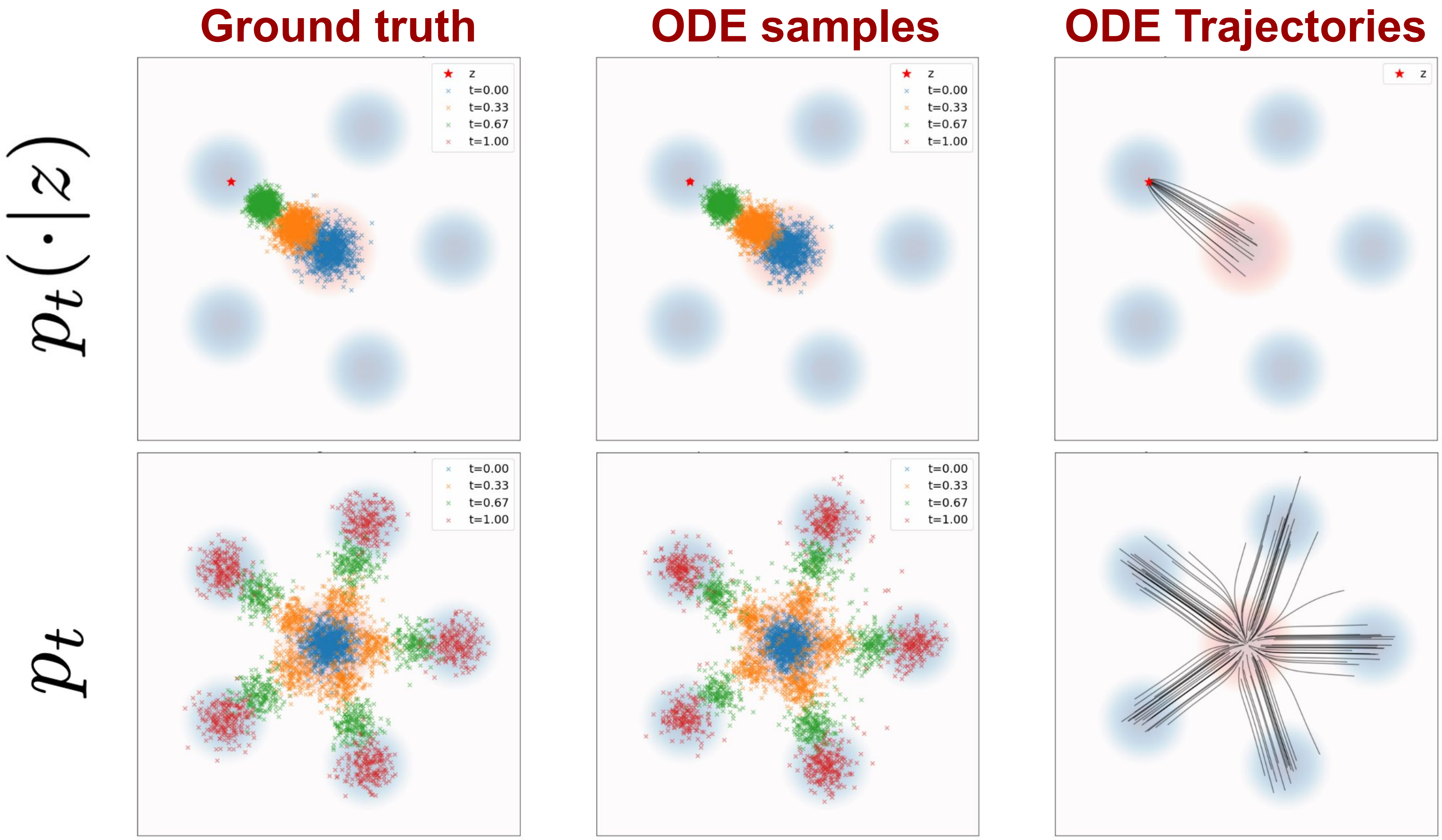
	Notation	Key property	Formula
Marginal Probability Path	p_t	Interpolates p_{init} and p_{data}	$\int p_t(x\|z)p_{\text{data}}(z)\mathrm{d}z$
Marginal Vector Field			
Marginal Score Function			

Simulating ODE with Conditional Vector Field for Conditional Probability Path

NOTE: This is an animated gif and is static in a PDF

Figure credit: Yaron Lipman




Ground truth
ODE samples
ODE Trajectories
$p_t(\cdot|z)$
p_t
z
t=0.00
t=0.33
t=0.67
t=1.00

Continuity Equation

Randomly initialized ODE

Given: $X_0 \sim p_{\text{init}}, \quad \frac{\mathrm{d}}{\mathrm{d}t} X_t = u_t(X_t)$

Follow probability path:

$$X_t \sim p_t \quad (0 \leq t \leq 1)$$

Marginals are p_t

equivalent

Continuity equation holds

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

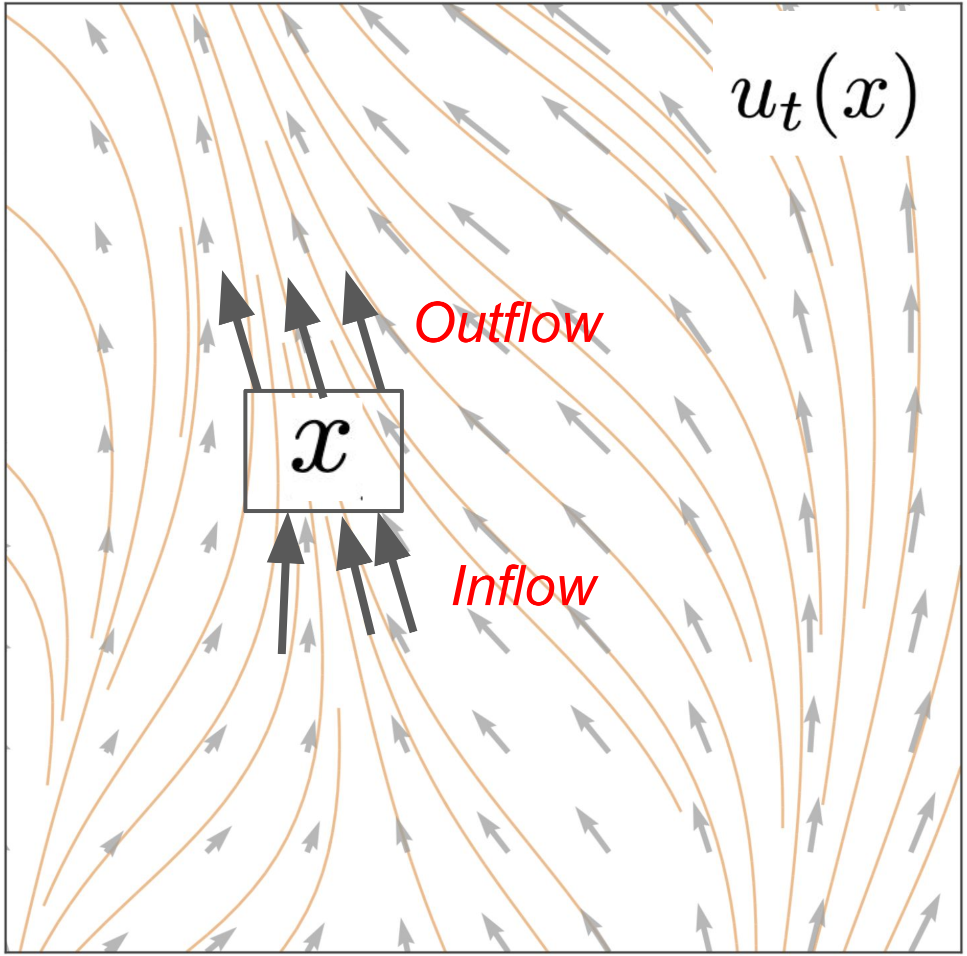
PDE holds

Continuity Equation

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

Change of probability mass at x

Outflow - inflow of probability mass from u



Gaussian Conditional Probability Path And Conditional Vector Field

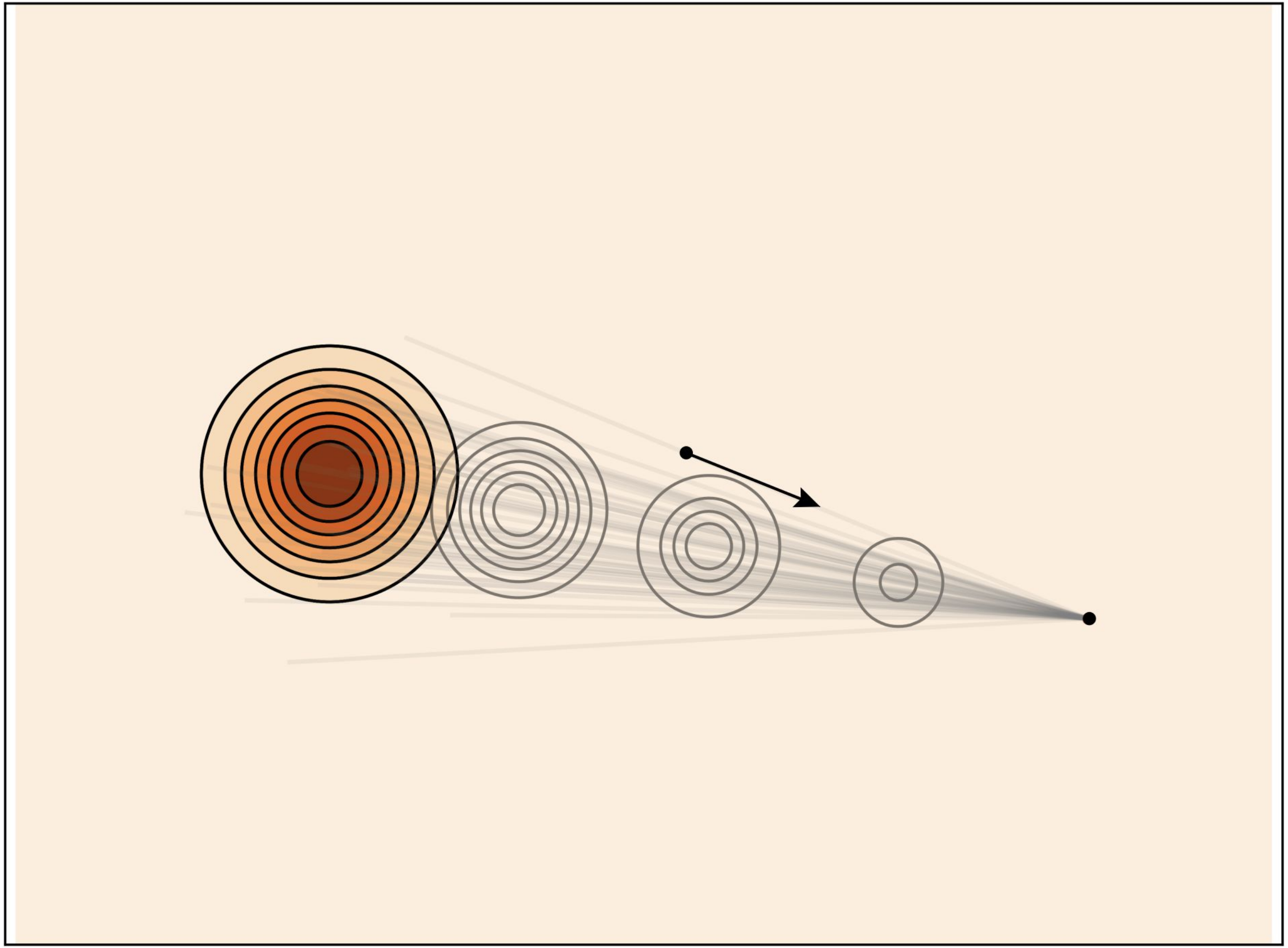


Figure credit: Yaron Lipman

Toy example

NOTE: This is an animated gif and is static in a PDF

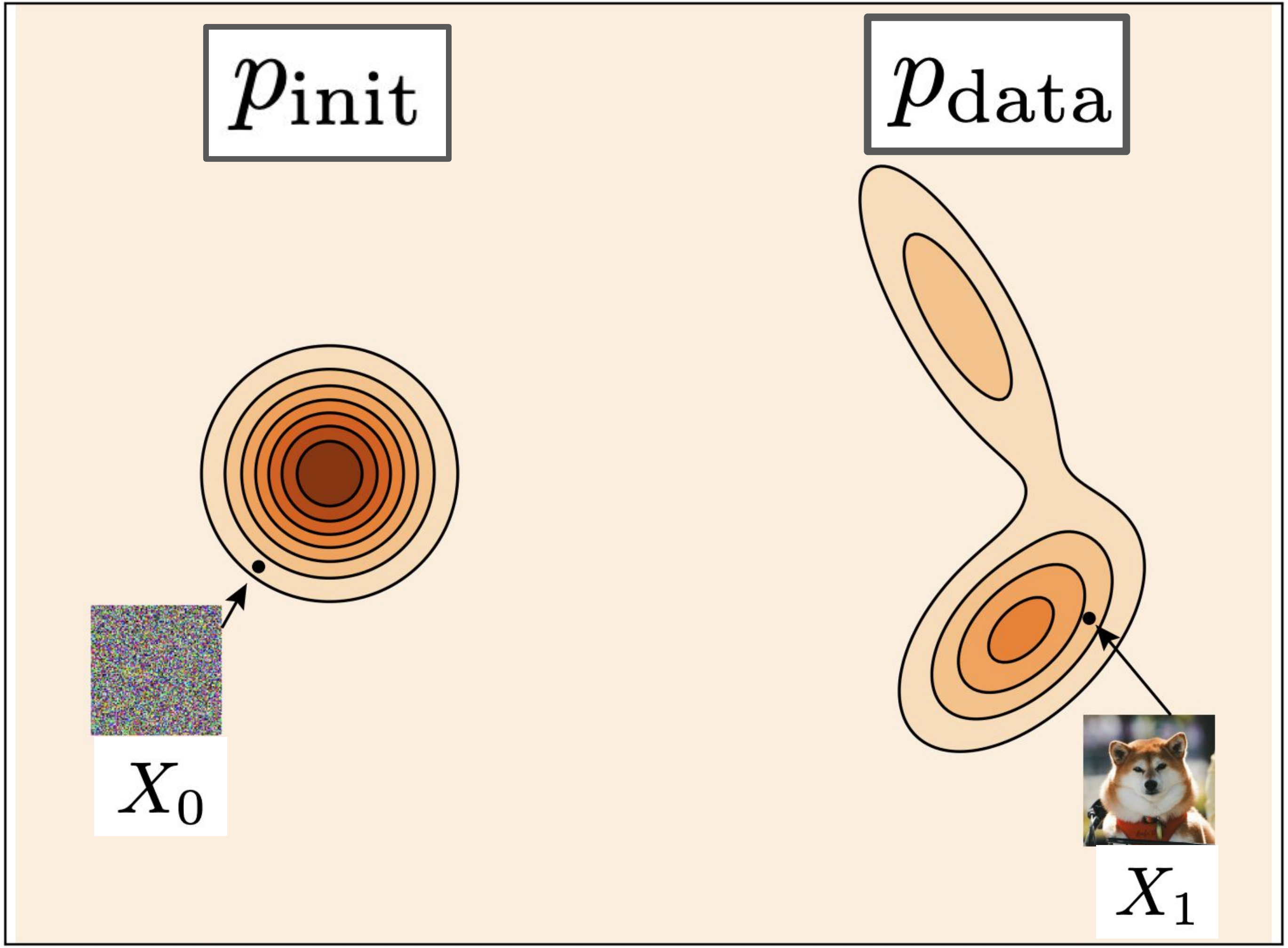


Figure credit: Yaron Lipman

Simulating ODE with Marginal Vector Field for Gaussian Probability Path

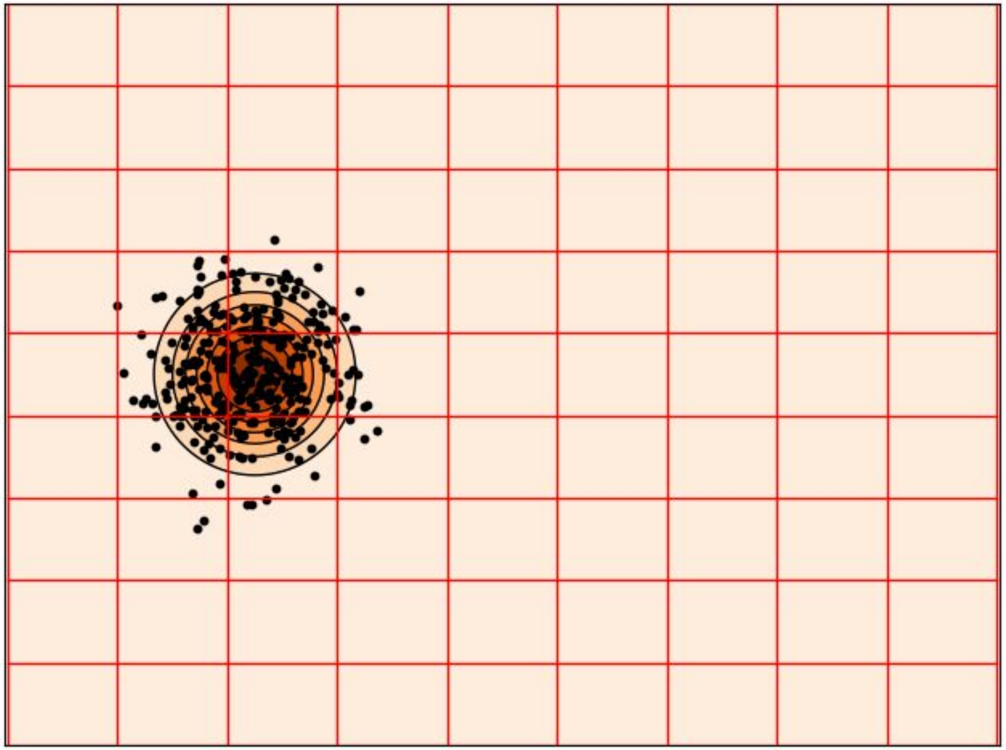


Figure credit: Yaron Lipman

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	$\left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$
Conditional Score Function			

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			

Outlook (Next class) - Flow Matching Loss

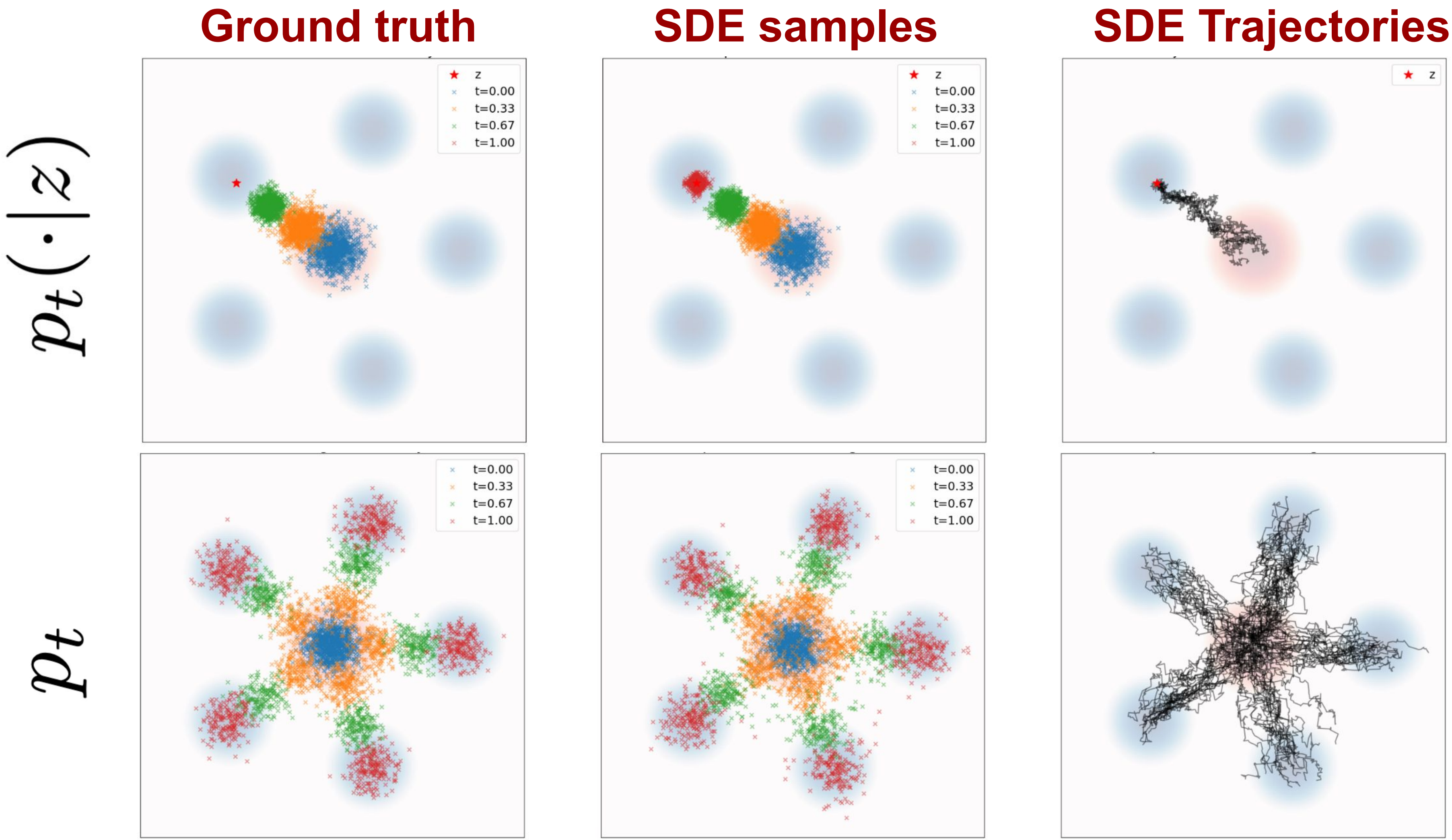
The Flow Matching loss is a mean squared error between the neural network and the marginal vector field:

$$L_{\mathrm{fm}}(\theta) = \mathbb{E}_{t\sim\mathrm{Unif},x\sim p_t}[\|u_t^\theta(x) - u_t^{\mathrm{target}}(x)\|^2]$$

Training a Flow Model Consists of Learning the Marginal Vector Field (How? Next lecture!)

Example marginal vector field - Meta MovieGen

These videos are generated by simulating the ODE with the (learnt) marginal vector field



Ground truth
SDE samples
SDE Trajectories
$p_t(\cdot|z)$
p_t
z
t=0.00
t=0.33
t=0.67
t=1.00

Fokker-Planck equation

Randomly initialized SDE

Given: $X_0 \sim p_{\text{init}}, \quad \mathrm{d}X_t = u_t(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$

Follow probability path:

$$X_t \sim p_t \quad (0 \leq t \leq 1)$$

Marginals are p_t

equivalent

Fokker-Planck equation holds

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(x) = -\mathrm{div}(p_t u_t)(x) + \frac{\sigma_t^2}{2} \Delta p_t(x)$$

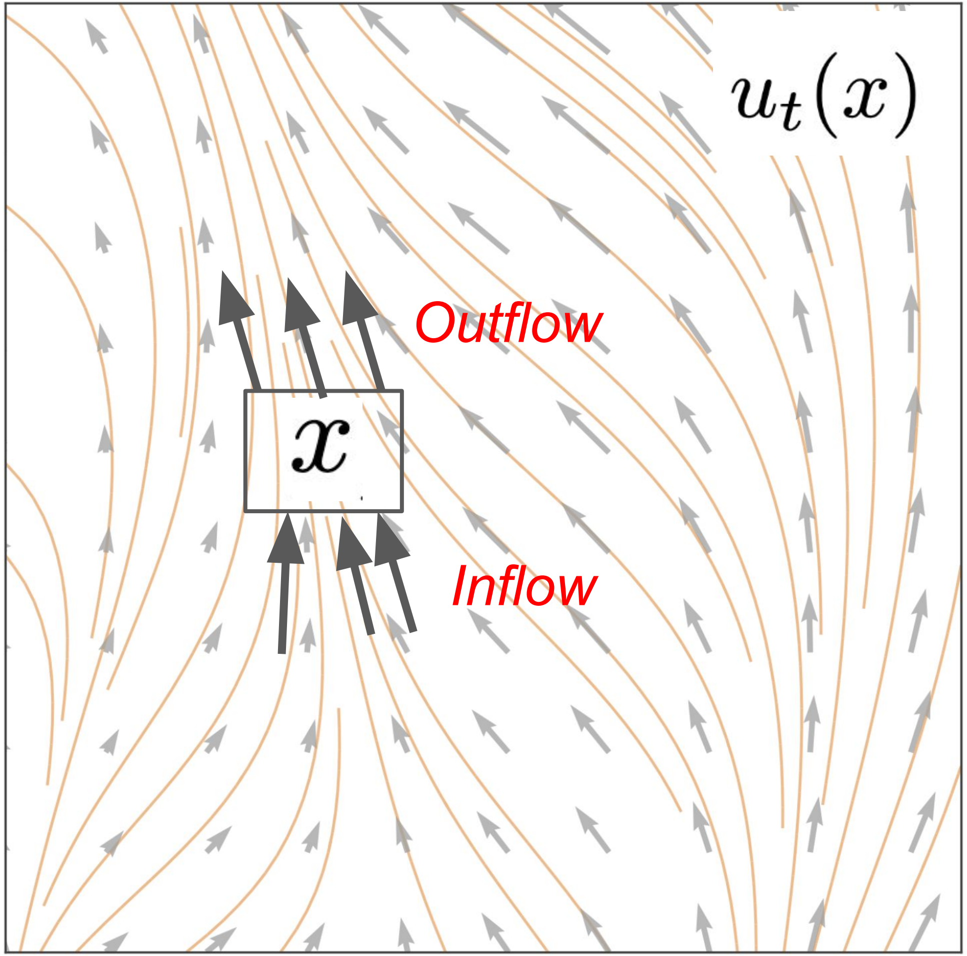
Continuity equ. *Heat equ.*

Continuity Equation

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

Change of probability mass at x

Outflow - inflow of probability mass from u

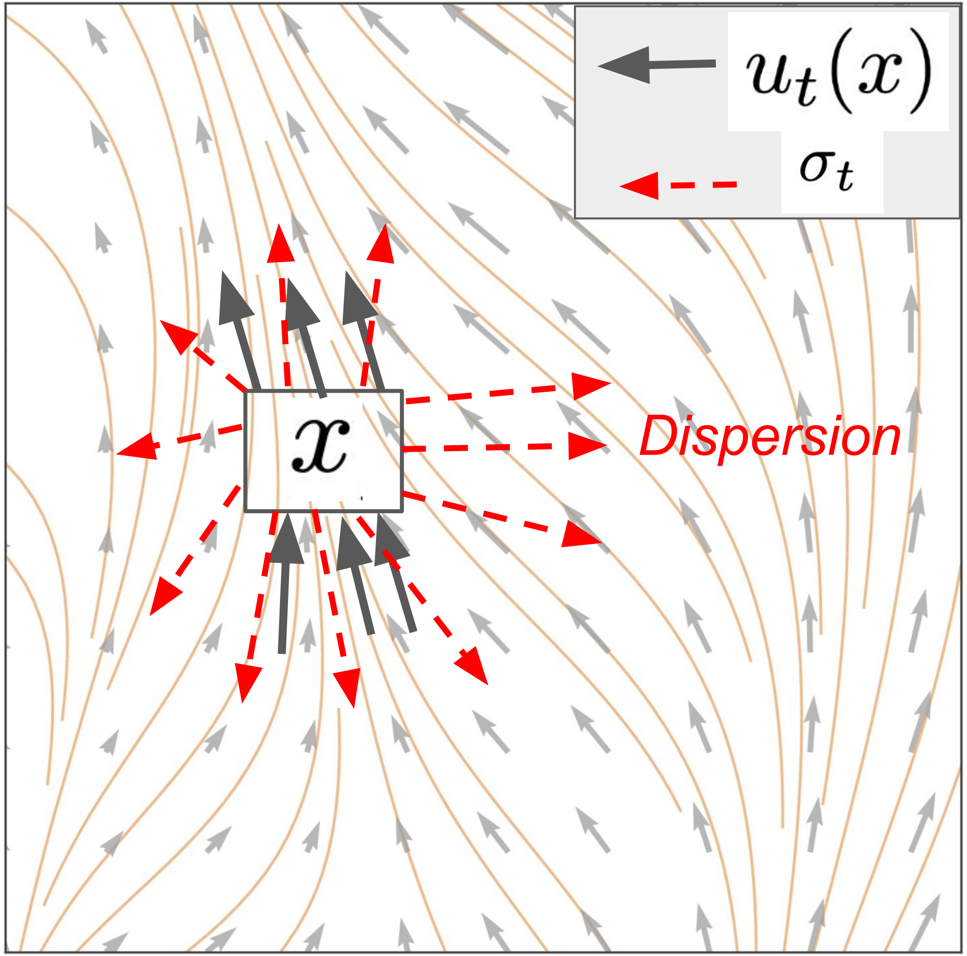


Fokker-Planck Equation

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(x) = -\mathrm{div}(p_t u_t)(x) + \frac{\sigma_t^2}{2}\Delta p_t(x)$$

Change of probability mass at x

Heat dispersion



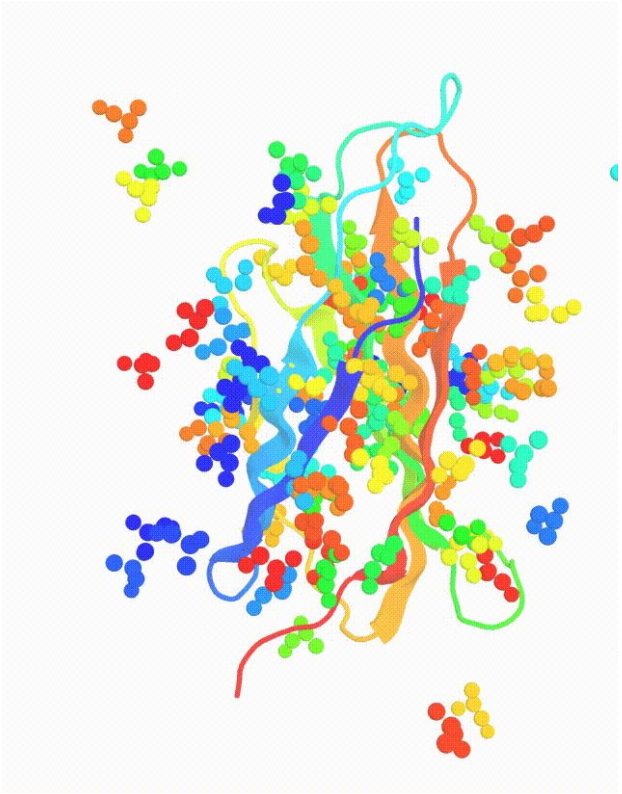
Outlook (Next class) - Score Matching Loss

The Score Matching loss is a mean squared error between the neural network and the marginal score function:

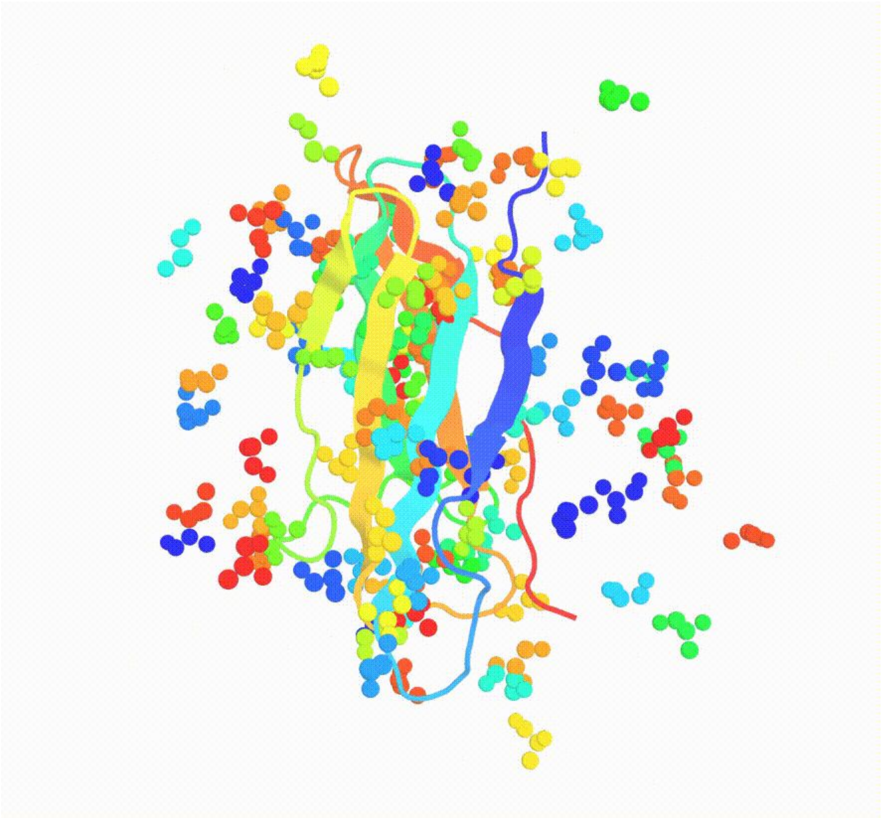
$$L_{\mathrm{sm}}(\theta) = \mathbb{E}_{z \sim p_{\mathrm{data}}, x \sim p_t(\cdot|z)}[||s_t^\theta(x) - \nabla \log p_t(x)||^2]$$

To train a diffusion model, we need to train the score network by minimizing the score matching loss (How? Next class!)

Marginal VF

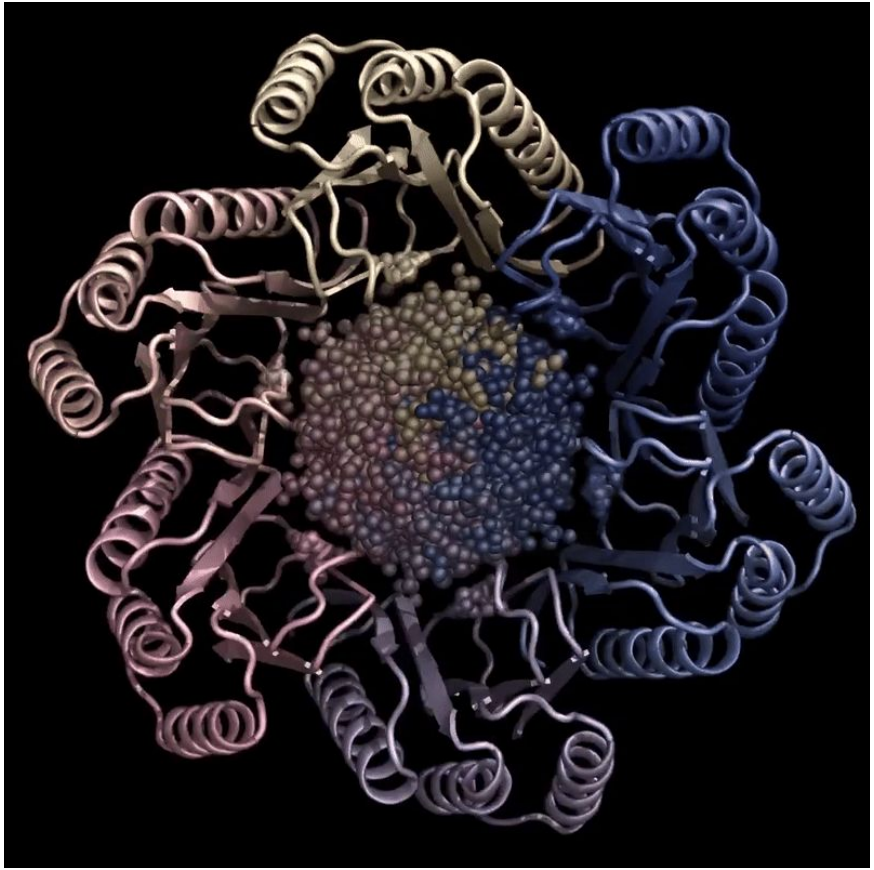


Marginal VF + Score



Training a Diffusion Model = Learning the Score Function

Conversion of of noise into protein structure by marginal vector field



NOTE: This is an animated gif and is static in a PDF

Slide credit: Jason Yim

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	$\left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$
Conditional Score Function	$\nabla \log p_t(x\|z)$	Gradient of log-likelihood	$-\frac{x - \alpha_t z}{\beta_t^2}$

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	$\int \nabla \log p_t(x\|z)\frac{p_t(x\|z)p_{\text{data}}(z)}{p_t(x)}\mathrm{d}z$

Today was the **technically most challenging lecture**!

The next lectures will **be much much easier**!

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**

These 6 formulas is all we need for training!

Next class:

Thursday (Tomorrow), 11am-12:30pm

Training algorithm!

E25-111 (same room)

Office hours: Today, 3pm-4:30pm in 37-212