

# Deep Generative Models

## Lecture 13

Roman Isachenko

Moscow Institute of Physics and Technology  
Yandex School of Data Analysis

2024, Autumn

# Outline

## 1. Flow Matching

Endpoint conditioning

Pair conditioning

Rectified flows

## Recap of previous lecture

### SDE basics

Let define stochastic process  $\mathbf{x}(t)$  with initial condition  $\mathbf{x}(0) \sim p_0(\mathbf{x})$ :

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w},$$

where  $\mathbf{w}(t)$  is the standard Wiener process (Brownian motion)

$$\mathbf{w}(t) - \mathbf{w}(s) \sim \mathcal{N}(0, (t-s)\mathbf{I}), \quad d\mathbf{w} = \boldsymbol{\epsilon} \cdot \sqrt{dt}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

### Discretization of SDE (Euler method)

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + \mathbf{f}(\mathbf{x}(t), t) \cdot dt + g(t) \cdot \boldsymbol{\epsilon} \cdot \sqrt{dt}$$

- ▶ At each moment  $t$  we have the density  $p_t(\mathbf{x}) = p(\mathbf{x}, t)$ .
- ▶  $p : \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}_+$  is a **probability path** between  $p_0(\mathbf{x})$  and  $p_1(\mathbf{x})$ .

## Recap of previous lecture

### Theorem (Kolmogorov-Fokker-Planck)

Evolution of the distribution  $p_t(\mathbf{x})$  is given by the following equation:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\operatorname{div}(\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) + \frac{1}{2}g^2(t)\Delta_{\mathbf{x}}p_t(\mathbf{x})$$

### Langevin SDE (special case)

$$d\mathbf{x} = \frac{1}{2}\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) dt + \mathbf{1} \cdot d\mathbf{w}$$

The density  $p(\mathbf{x}|\theta)$  is a **stationary** distribution for the SDE.

### Langevin dynamics

Samples from the following dynamics will come from  $p(\mathbf{x}|\theta)$  under mild regularity conditions for small enough  $\eta$ .

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \frac{\eta}{2}\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\theta) + \sqrt{\eta} \cdot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

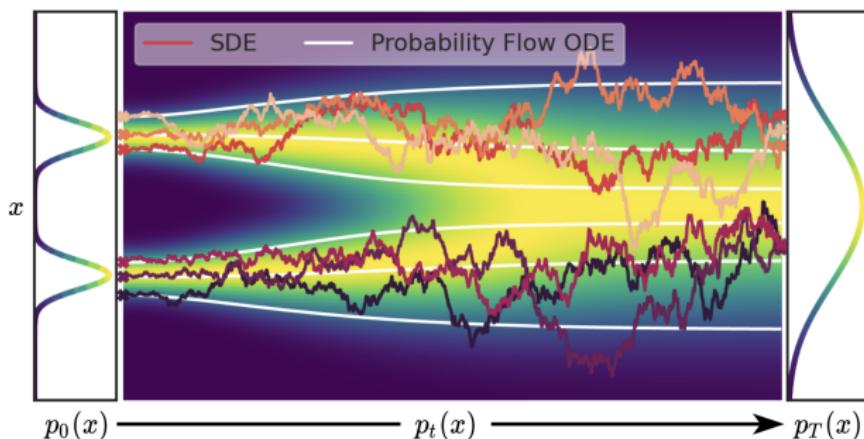
## Recap of previous lecture

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} - \text{SDE with the probability path } p_t(\mathbf{x})$$

## Probability flow ODE

There exists ODE with identical the probability path  $p_t(\mathbf{x})$  of the form

$$d\mathbf{x} = \left[ \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right] dt$$



## Recap of previous lecture

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt, \quad \mathbf{x}(t + dt) = \mathbf{x}(t) + \mathbf{f}(\mathbf{x}, t)dt$$

### Reverse ODE

Let  $\tau = 1 - t$  ( $d\tau = -dt$ ).

$$d\mathbf{x} = -\mathbf{f}(\mathbf{x}, 1 - \tau)d\tau$$

### Reverse SDE

There exists the reverse SDE for the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  that has the following form

$$d\mathbf{x} = \left( \mathbf{f}(\mathbf{x}, t) - g^2(t) \frac{\partial \log p_t(\mathbf{x})}{\partial \mathbf{x}} \right) dt + g(t)d\mathbf{w}, \quad dt < 0$$

### Sketch of the proof

- ▶ Convert initial SDE to probability flow ODE.
- ▶ Revert probability flow ODE.
- ▶ Convert reverse probability flow ODE to reverse SDE.

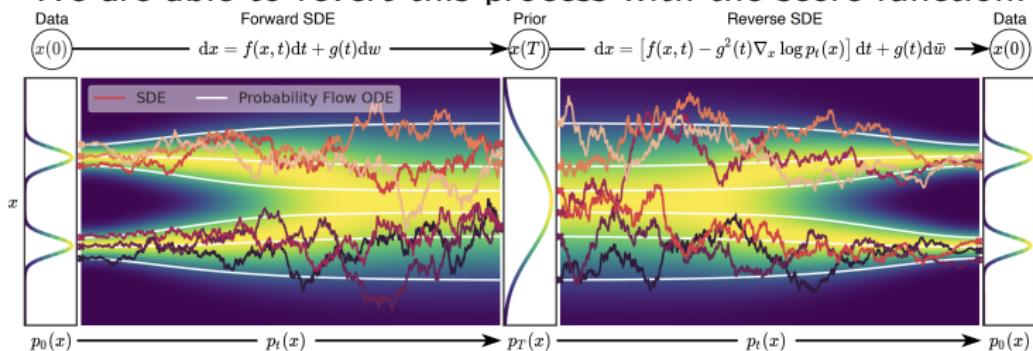
## Recap of previous lecture

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} - \text{SDE}$$

$$d\mathbf{x} = \left[ \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right] dt - \text{probability flow ODE}$$

$$d\mathbf{x} = \left( \mathbf{f}(\mathbf{x}, t) - g^2(t)\frac{\partial \log p_t(\mathbf{x})}{\partial \mathbf{x}} \right) dt + g(t)d\mathbf{w} - \text{reverse SDE}$$

- ▶ We got the way to transform one distribution to another via SDE with some probability path  $p_t(\mathbf{x})$ .
- ▶ We are able to revert this process with the score function.



## Recap of previous lecture

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$$

### Variance Exploding SDE (NCSN)

$$d\mathbf{x} = \sqrt{\frac{d[\sigma^2(t)]}{dt}} \cdot d\mathbf{w}, \quad \mathbf{f}(\mathbf{x}, t) = 0, \quad g(t) = \sqrt{\frac{d[\sigma^2(t)]}{dt}}$$

Variance grows since  $\sigma(t)$  is a monotonically increasing function.

### Variance Preserving SDE (DDPM)

$$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}(t)dt + \sqrt{\beta(t)} \cdot d\mathbf{w}$$

$$\mathbf{f}(\mathbf{x}, t) = -\frac{1}{2}\beta(t)\mathbf{x}(t), \quad g(t) = \sqrt{\beta(t)}$$

Variance is preserved if  $\mathbf{x}(0)$  has a unit variance.

# Outline

## 1. Flow Matching

Endpoint conditioning

Pair conditioning

Rectified flows

# Flow Matching

## Continuous-in-time NF

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{f}_\theta(\mathbf{z}(t), t); \quad \text{with initial condition } \mathbf{z}(0) = \mathbf{z}_0.$$

Theorem (Kolmogorov-Fokker-Planck: special case)

If  $\mathbf{f}$  is uniformly Lipschitz continuous in  $\mathbf{z}$  and continuous in  $t$ , then

$$\frac{d \log p_t(\mathbf{z}(t))}{dt} = -\text{tr} \left( \frac{\partial \mathbf{f}_\theta(\mathbf{z}(t), t)}{\partial \mathbf{z}(t)} \right).$$

$$\log p_1(\mathbf{z}(1)) = \log p_0(\mathbf{z}(0)) - \int_0^1 \text{tr} \left( \frac{\partial \mathbf{f}_\theta(\mathbf{z}(t), t)}{\partial \mathbf{z}(t)} \right) dt.$$

- ▶ Solving KFP using adjoint method is complicated and unstable process.
- ▶ Flow matching generalizes these models and gives the alternative way to solve the continuous dynamics.

## Flow Matching

Let consider ODE dynamic  $\mathbf{x}_t = \mathbf{x}(t)$  in time interval  $t \in [0, 1]$

- ▶  $\mathbf{x}_0 \sim p_0(\mathbf{x}) = p(\mathbf{x})$ ,  $\mathbf{x}_1 \sim p_1(\mathbf{x}) = \pi(\mathbf{x})$ ;
- ▶  $p(\mathbf{x})$  is a base distribution ( $\mathcal{N}(0, \mathbf{I})$ ) and  $\pi(\mathbf{x})$  is a true data distribution.

**Note:** the difference with the diffusion models (and CNF) in the opposite time direction.

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, t),$$

$\mathbf{u}(\mathbf{x}, t) : \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}^m$  is a vector field.

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\text{div}(\mathbf{u}(\mathbf{x}, t)p_t(\mathbf{x}))$$

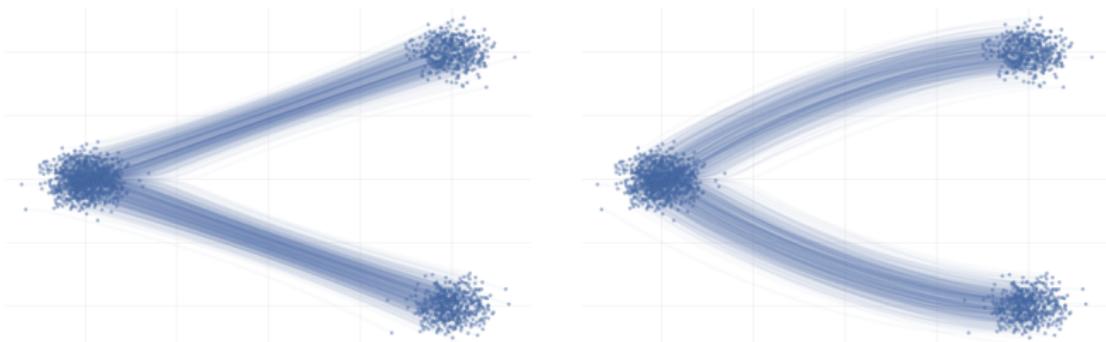
If we know the true vector field  $\mathbf{u}(\mathbf{x}, t)$ , then KFP equation gives us the way to compute the density  $p_t(\mathbf{x})$ .

# Flow Matching

## Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x})} \|\mathbf{u}(\mathbf{x}, t) - \mathbf{u}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

- ▶ There exists infinite number of possible  $\mathbf{u}(\mathbf{x}, t)$  between  $\pi(\mathbf{x})$  and  $p(\mathbf{x})$ .
- ▶ The true vector field  $\mathbf{u}(\mathbf{x}, t)$  is **unknown**.



# Flow Matching

## Latent variable model

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

Here  $p_t(\mathbf{x}|\mathbf{z})$  is a **conditional probability path**.

The conditional probability path  $p_t(\mathbf{x}|\mathbf{z})$  satisfies KFP theorem

$$\frac{\partial p_t(\mathbf{x}|\mathbf{z})}{\partial t} = -\text{div}(\mathbf{u}(\mathbf{x}, \mathbf{z}, t)p_t(\mathbf{x}|\mathbf{z})),$$

where  $\mathbf{u}(\mathbf{x}, \mathbf{z}, t)$  is a **conditional vector field**.

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, t) \quad \Rightarrow \quad \frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, \mathbf{z}, t)$$

# Flow Matching

$$\frac{\partial p_t(\mathbf{x}|\mathbf{z})}{\partial t} = -\operatorname{div}(\mathbf{u}(\mathbf{x}, \mathbf{z}, t) p_t(\mathbf{x}|\mathbf{z})),$$

Theorem

$$\mathbf{u}(\mathbf{x}, t) = \int \mathbf{u}(\mathbf{x}, \mathbf{z}, t) \frac{p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z})}{p_t(\mathbf{x})} d\mathbf{z}$$

Proof

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \frac{\partial}{\partial t} \int p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = \int \left( \frac{\partial p_t(\mathbf{x}|\mathbf{z})}{\partial t} \right) p(\mathbf{z}) d\mathbf{z} = \\ &= \int (-\operatorname{div}(\mathbf{u}(\mathbf{x}, \mathbf{z}, t) p_t(\mathbf{x}|\mathbf{z}))) p(\mathbf{z}) d\mathbf{z} = \\ &= -\operatorname{div} \left( \int \mathbf{u}(\mathbf{x}, \mathbf{z}, t) p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} \right) = -\operatorname{div}(\mathbf{u}(\mathbf{x}, t) p_t(\mathbf{x}))\end{aligned}$$

# Flow Matching

## Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x})} \|\mathbf{u}(\mathbf{x}, t) - \mathbf{u}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

## Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{u}(\mathbf{x}, \mathbf{z}, t) - \mathbf{u}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

## Theorem

If  $\text{supp}(p_t(\mathbf{x})) = \mathbb{R}^m$ , then the optimal value of FM objective is equal to the optimal value of CFM objective.

## Proof

It is proved similarly with the denoising score matching theorem.

## Flow Matching

Gaussian conditional probability path

$$p_t(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{z}), \boldsymbol{\sigma}_t^2(\mathbf{z}))$$

- ▶ There is an infinite number of vector fields that generate any particular probability path.
- ▶ Let consider the following dynamics:

$$\mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{z}) + \boldsymbol{\sigma}_t(\mathbf{z}) \odot \mathbf{x}_0$$

# Flow Matching

## Theorem

$$\mathbf{u}(\mathbf{x}, \mathbf{z}, t) = \boldsymbol{\mu}'_t(\mathbf{z}) + \frac{\boldsymbol{\sigma}'_t(\mathbf{z})}{\boldsymbol{\sigma}_t(\mathbf{z})} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{z}))$$

## Proof

$$\mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{z}) + \boldsymbol{\sigma}_t(\mathbf{z}) \odot \mathbf{x}_0 \quad \Rightarrow \quad \mathbf{x}_0 = \frac{1}{\boldsymbol{\sigma}_t(\mathbf{z})} \odot (\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{z}))$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, \mathbf{z}, t)$$

$$\frac{d\mathbf{x}}{dt} = \boldsymbol{\mu}'_t(\mathbf{z}) + \boldsymbol{\sigma}'_t(\mathbf{z}) \odot \mathbf{x}_0 = \boldsymbol{\mu}'_t(\mathbf{z}) + \frac{\boldsymbol{\sigma}'_t(\mathbf{z})}{\boldsymbol{\sigma}_t(\mathbf{z})} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{z}))$$

# Outline

## 1. Flow Matching

Endpoint conditioning

Pair conditioning

Rectified flows

# Flow Matching

## Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{u}(\mathbf{x}, \mathbf{z}, t) - \mathbf{u}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

Let choose  $\mathbf{z} = \mathbf{x}_1$ . Then  $p(\mathbf{z}) = p_1(\mathbf{x}_1)$ .

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_1) p_1(\mathbf{x}_1) d\mathbf{x}_1$$

We need to ensure boundary conditions:

$$\begin{cases} p_0(\mathbf{x}) = p(\mathbf{x}) = \mathcal{N}(0, \mathbf{I}); \\ p_1(\mathbf{x}) = \pi(\mathbf{x}). \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_1) = p_0(\mathbf{x}); \\ p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases}$$

## Gaussian conditional probability path

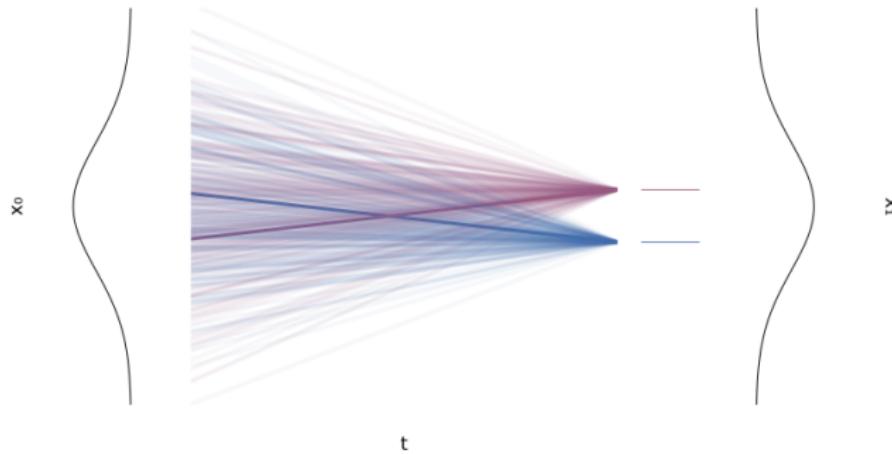
$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_1)); \quad \mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{x}_1) + \boldsymbol{\sigma}_t^2(\mathbf{x}_1) \odot \mathbf{x}_0.$$

# Flow Matching

Gaussian conditional probability path

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_1)); \quad \mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{x}_1) + \boldsymbol{\sigma}_t^2(\mathbf{x}_1)\mathbf{x}_0.$$

$$p_0(\mathbf{x}|\mathbf{x}_1) = p_0(\mathbf{x}) \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1)$$



## Flow Matching

$$\begin{cases} p_0(\mathbf{x}|\mathbf{x}_1) = p_0(\mathbf{x}); \\ p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases} \Rightarrow \begin{cases} \mu_0(\mathbf{x}_1) = 0, & \sigma_0(\mathbf{x}_1) = 1; \\ \mu_1(\mathbf{x}_1) = \mathbf{x}_1, & \sigma_1(\mathbf{x}_1) = 0. \end{cases}$$

Let consider straight conditional paths

$$\begin{cases} \mu_t(\mathbf{x}_1) = t\mathbf{x}_1; \\ \sigma_t(\mathbf{x}_1) = (1-t). \end{cases} \Rightarrow \begin{cases} p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2); \\ \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0. \end{cases}$$

$$\mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = \mu'_t(\mathbf{x}_1) + \frac{\sigma'_t(\mathbf{x}_1)}{\sigma_t(\mathbf{x}_1)} \odot (\mathbf{x} - \mu_t(\mathbf{x}_1))$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, \mathbf{z}, t); \quad \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\mathbf{x}_1 - \mathbf{x}}{1-t}; \quad \frac{d\mathbf{x}}{dt} = \mathbf{x}_1 - \mathbf{x}_0.$$

# Flow Matching

## Conditional Flow Matching

$$\begin{aligned} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \| \mathbf{u}(\mathbf{x}, \mathbf{z}, t) - \mathbf{u}_\theta(\mathbf{x}, t) \|^2 = \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x})} \| (\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{u}_\theta(\mathbf{x}, t) \|^2 \rightarrow \min_{\theta} \end{aligned}$$

We fit straight lines between noise distribution  $p(\mathbf{x})$  and the data distribution  $\pi(\mathbf{x})$ .

## Flow Matching

- ▶ The conditional probability path  $p_t(\mathbf{x}|\mathbf{z})$  is an optimal transport path from  $p_0(\mathbf{x}|\mathbf{z})$  to  $p_1(\mathbf{x}|\mathbf{z})$ .
- ▶ The marginal path  $p_t(\mathbf{x})$  is not in general an optimal transport path from the standard normal  $p_0(\mathbf{x})$  to the data distribution  $p_1(\mathbf{x})$ .



## Flow Matching

### Flow matching probability path

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2\mathbf{I}); \quad \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\mathbf{x}_1 - \mathbf{x}}{1-t}$$

### Variance Exploding SDE probability path

$$d\mathbf{x} = \sqrt{\frac{d[\sigma^2(t)]}{dt}} \cdot d\mathbf{w}; \quad \Rightarrow \quad \begin{cases} p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1, \sigma_{1-t}^2 \mathbf{I}) \\ \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = -\frac{\sigma'_{1-t}}{\sigma_{1-t}} \cdot (\mathbf{x} - \mathbf{x}_1) \end{cases}$$

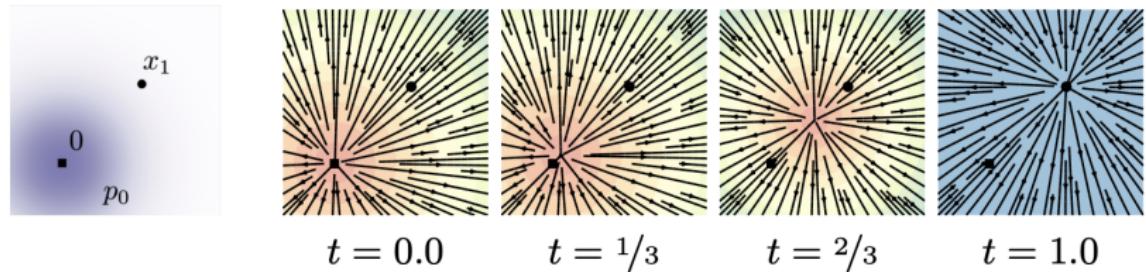
### Variance Preserving SDE probability path

$$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}(t)dt + \sqrt{\beta(t)}d\mathbf{w}; \quad \Rightarrow \quad \begin{cases} p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\alpha_{1-t}\mathbf{x}_1, (1-\alpha_{1-t}^2)\mathbf{I}) \\ \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\alpha'_{1-t}}{1-\alpha_{1-t}^2} \cdot (\alpha_{1-t}\mathbf{x} - \mathbf{x}_1) \end{cases}$$

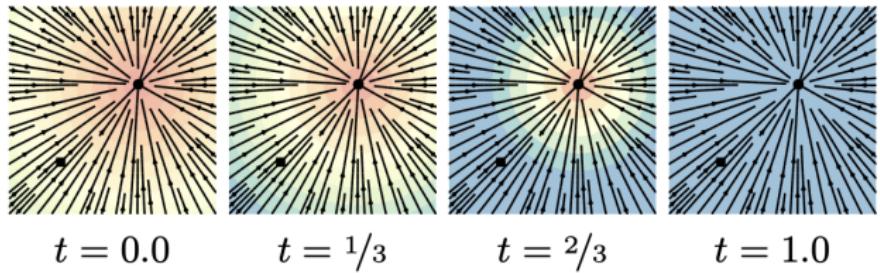
Here  $\alpha_t = \exp\left(-\frac{1}{2}\int_0^t \beta(s)ds\right)$ .

# Flow Matching

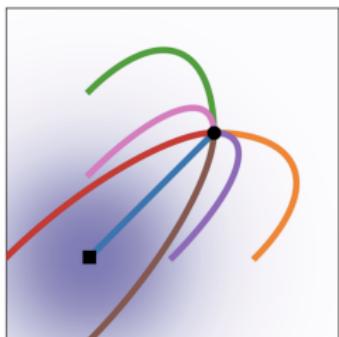
## Diffusion vector field



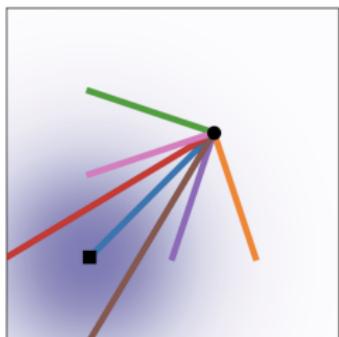
## Flow matching vector field



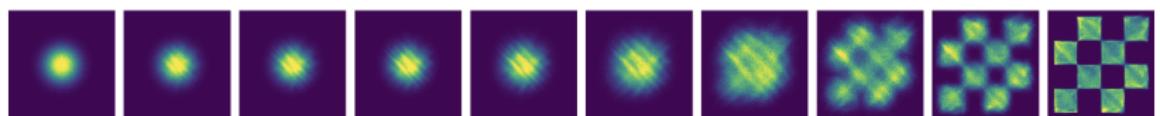
# Flow Matching



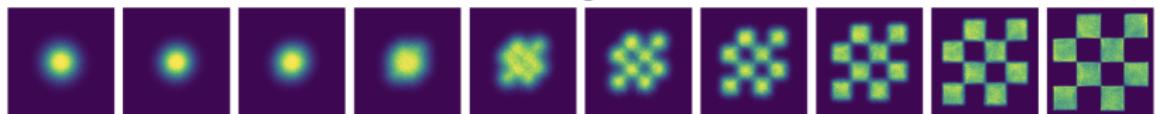
Diffusion



OT



Score matching w/ Diffusion



Flow Matching w/ OT

# Outline

## 1. Flow Matching

Endpoint conditioning

Pair conditioning

Rectified flows

## Flow Matching

### Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{u}(\mathbf{x}, \mathbf{z}, t) - \mathbf{u}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

Let choose  $\mathbf{z} = (\mathbf{x}_0, \mathbf{x}_1)$ . Then  $p(\mathbf{z}) = p_0(\mathbf{x}_0)p_1(\mathbf{x}_1)$ .

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) p_0(\mathbf{x}_0) p_1(\mathbf{x}_1) d\mathbf{x}_0 d\mathbf{x}_1$$

We need to ensure boundary conditions:

$$\begin{cases} p_0(\mathbf{x}) = p(\mathbf{x}) = \mathcal{N}(0, \mathbf{I}); \\ p_1(\mathbf{x}) = \pi(\mathbf{x}). \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases}$$

### Gaussian conditional probability path

$$p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_0, \mathbf{x}_1)); \quad \mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1) + \boldsymbol{\sigma}_t^2(\mathbf{x}_0, \mathbf{x}_1) \odot \mathbf{x}_0$$

## Flow Matching

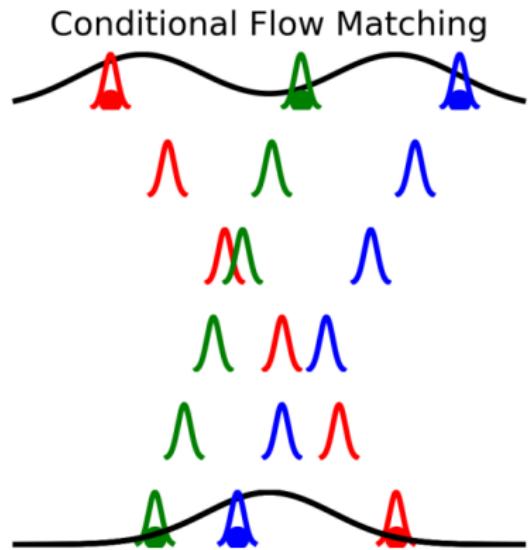
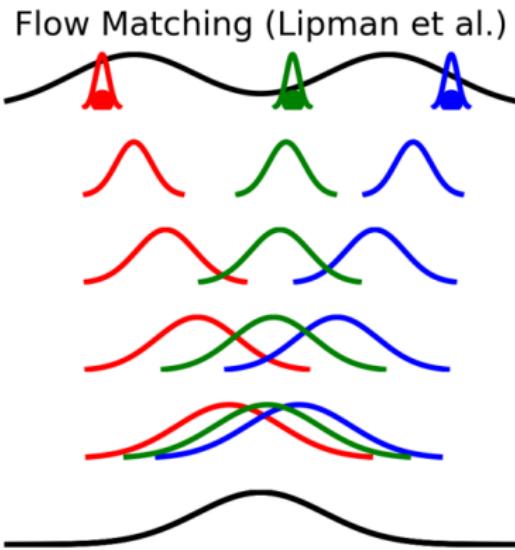
$$\begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases} \Rightarrow \begin{cases} \mu_0(\mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_0, & \sigma_0(\mathbf{x}_0, \mathbf{x}_1) = 0 \\ \mu_1(\mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_1, & \sigma_1(\mathbf{x}_0, \mathbf{x}_1) = 0 \end{cases}$$

Let consider straight conditional paths

$$\begin{cases} \mu_t(\mathbf{x}_0, \mathbf{x}_1) = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0; \\ \sigma_t(\mathbf{x}_0, \mathbf{x}_1) = 0. \end{cases} \Rightarrow \begin{cases} p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1 + (1 - t)\mathbf{x}_0, 0) \\ \mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0. \end{cases}$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{x}_1 - \mathbf{x}_0.$$

# Flow Matching



# Outline

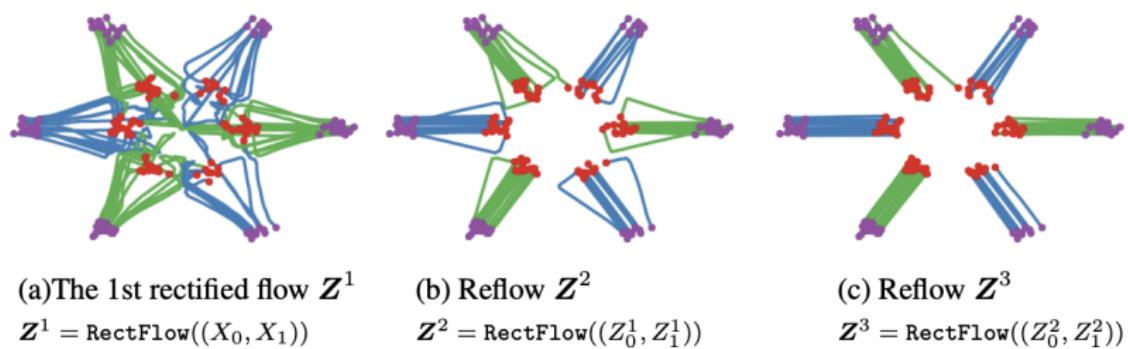
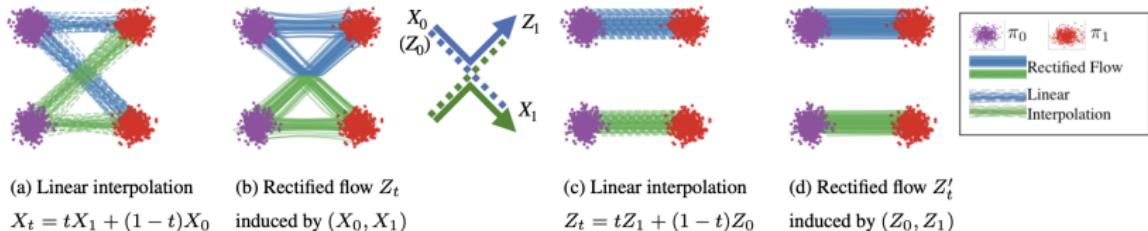
## 1. Flow Matching

Endpoint conditioning

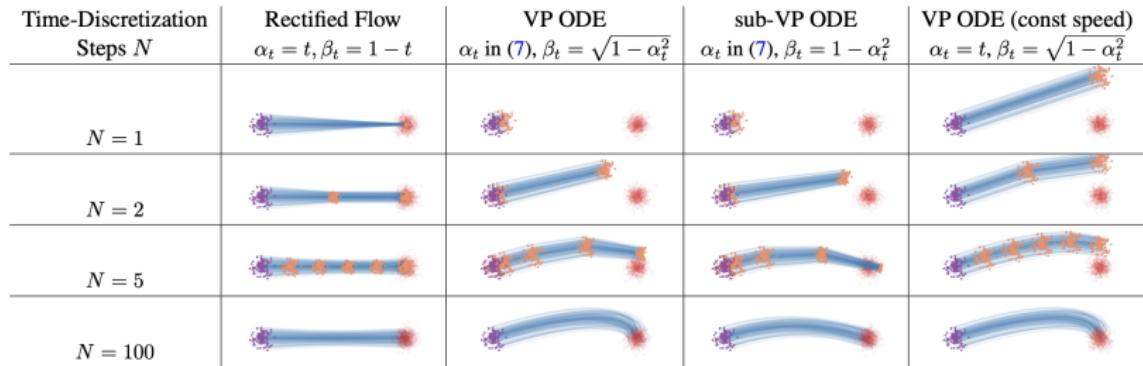
Pair conditioning

Rectified flows

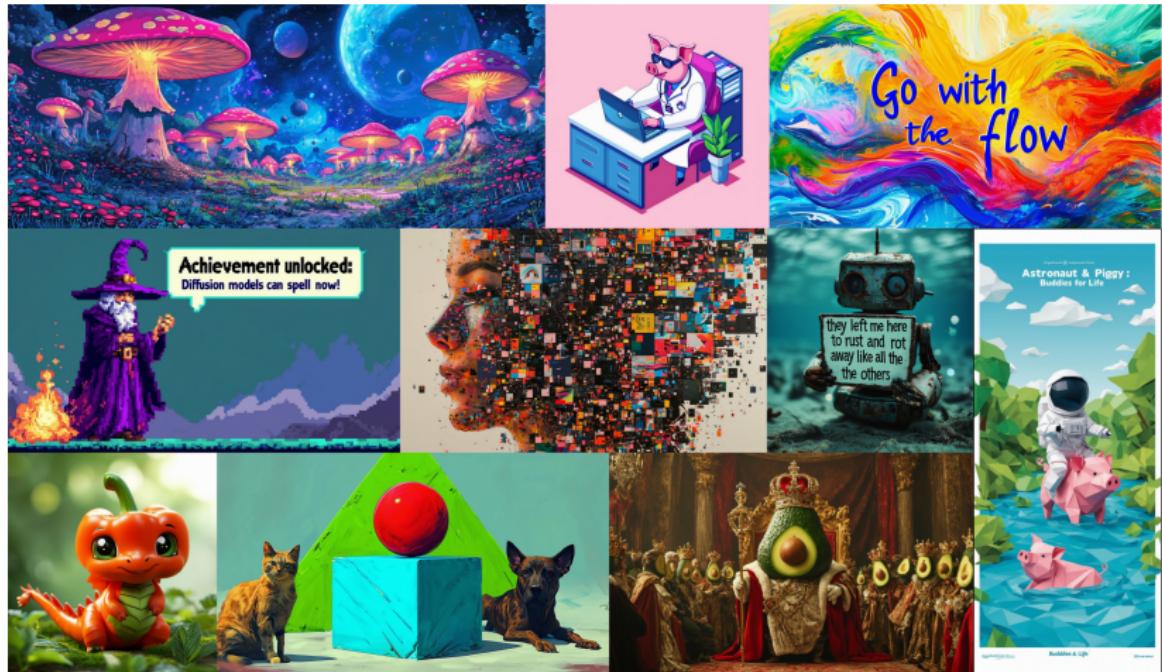
# Flow Matching



# Flow Matching



# Stable Diffusion 3



# Summary

