

# Deep Generative Models

## Lecture 13

Roman Isachenko

Moscow Institute of Physics and Technology  
Yandex School of Data Analysis

2024, Autumn

## 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  $\eta$ .

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

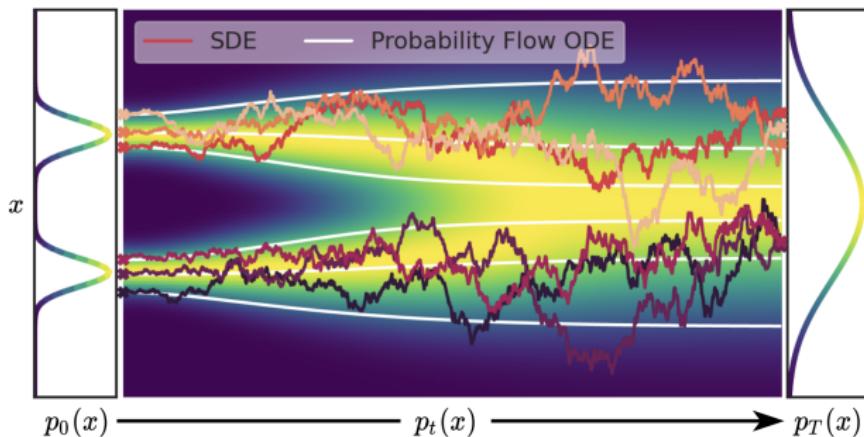
## Recap of previous lecture

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

## 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

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

### Sketch of the proof

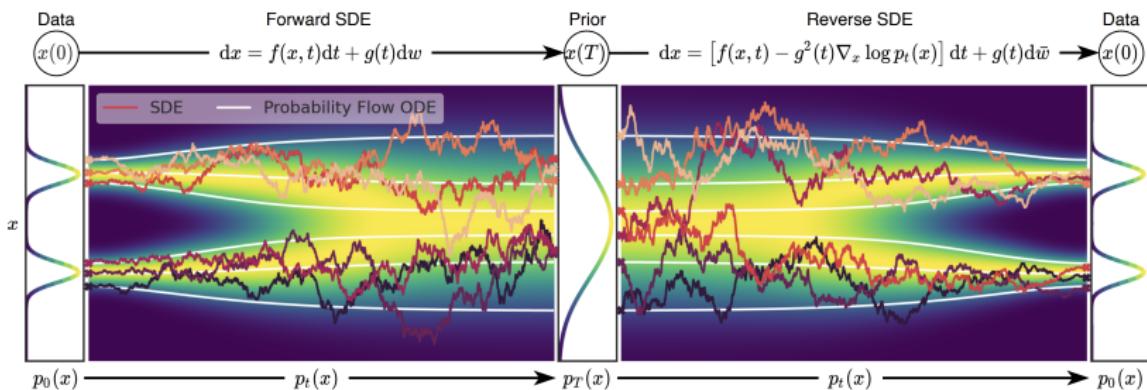
- ▶ Convert initial SDE to probability flow ODE.
- ▶ Revert probability flow ODE.
- ▶ Convert reverse probability flow ODE to reverse SDE.

# Recap of previous lecture

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

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

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



## 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. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

# Outline

1. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

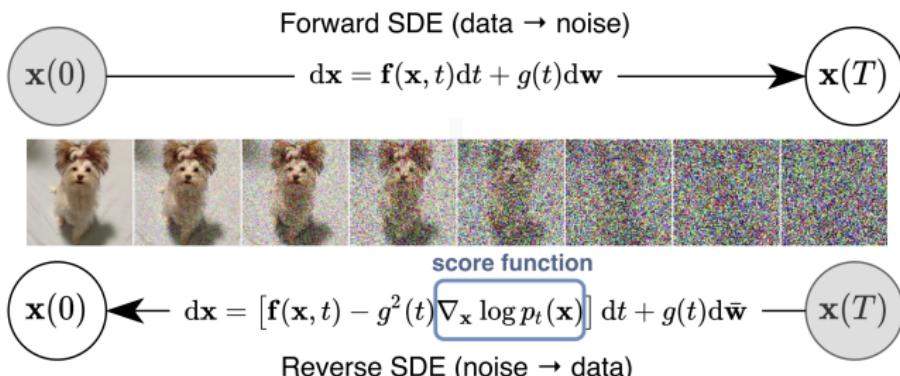
# Score-based generative models through SDEs

## Discrete-in-time objective

$$\mathbb{E}_{\pi(\mathbf{x}_0)} \mathbb{E}_{t \sim U\{1, T\}} \mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} \| \mathbf{s}_{\theta, t}(\mathbf{x}_t) - \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t | \mathbf{x}_0) \|_2^2$$

## Continuous-in-time objective

$$\mathbb{E}_{\pi(\mathbf{x}(0))} \mathbb{E}_{t \sim U[0, 1]} \mathbb{E}_{q(\mathbf{x}(t) | \mathbf{x}(0))} \| \mathbf{s}_{\theta}(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t) | \mathbf{x}(0)) \|_2^2$$



# Score-based generative models through SDEs

## Continuous-in-time objective

$$\mathbb{E}_{\pi(\mathbf{x}(0))} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{q(\mathbf{x}(t)|\mathbf{x}(0))} \| \mathbf{s}_\theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2$$

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}\left(\boldsymbol{\mu}(\mathbf{x}(t), \mathbf{x}(0)), \boldsymbol{\Sigma}(\mathbf{x}(t), \mathbf{x}(0))\right)$$

## Theorem

Moments of the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  satisfies the equations

$$\frac{d\boldsymbol{\mu}(\mathbf{x}(t), \mathbf{x}(0))}{dt} = \mathbb{E} [\mathbf{f}(\mathbf{x}(t), t)|\mathbf{x}(0)]$$

$$\frac{d\boldsymbol{\Sigma}(\mathbf{x}(t), \mathbf{x}(0))}{dt} = \mathbb{E} \left[ \mathbf{f} \cdot (\mathbf{x}(t) - \boldsymbol{\mu})^T + (\mathbf{x}(t) - \boldsymbol{\mu}) \cdot \mathbf{f}^T | \mathbf{x}(0) \right] + g^2(t) \cdot \mathbf{I}$$

Let prove the first one.

# Score-based generative models through SDEs

Theorem

$$\frac{d\mu(\mathbf{x}(t), \mathbf{x}(0))}{dt} = \mathbb{E} [\mathbf{f}(\mathbf{x}(t), t) | \mathbf{x}(0)]$$

Proof

$$\begin{aligned}\mathbb{E} [d\mathbf{x} | \mathbf{x}(0)] &= \mathbb{E} [\mathbf{f}(\mathbf{x}, t) dt | \mathbf{x}(0)] + \mathbb{E} [g(t) d\mathbf{w} | \mathbf{x}(0)] \\ &= \mathbb{E} [\mathbf{f}(\mathbf{x}, t) | \mathbf{x}(0)] dt + g(t) \mathbb{E} [d\mathbf{w} | \mathbf{x}(0)] \\ &= \mathbb{E} [\mathbf{f}(\mathbf{x}, t) | \mathbf{x}(0)] dt\end{aligned}$$

$$\frac{d\mathbb{E} [\mathbf{x} | \mathbf{x}(0)]}{dt} = \frac{d\mu(\mathbf{x}(t), \mathbf{x}(0))}{dt} = \mathbb{E} [\mathbf{f}(\mathbf{x}, t) | \mathbf{x}(0)]$$

Examples

**NCSN:**  $\mathbf{f}(\mathbf{x}, t) = 0 \quad \Rightarrow \quad \mu = \mathbf{x}(0)$

**DDPM:**  $\mathbf{f}(\mathbf{x}, t) = -\frac{1}{2}\beta(t)\mathbf{x}(t) \quad \Rightarrow \quad \mu = \mathbf{x}(0) \exp\left(-\frac{1}{2} \int_0^t \beta(s) ds\right)$

# Score-based generative models through SDEs

## Training

$$\mathbb{E}_{\pi(\mathbf{x}(0))} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{q(\mathbf{x}(t)|\mathbf{x}(0))} \| \mathbf{s}_\theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2$$

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}\left(\boldsymbol{\mu}(\mathbf{x}(t), \mathbf{x}(0)), \boldsymbol{\Sigma}(\mathbf{x}(t), \mathbf{x}(0))\right)$$

## NCSN

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}\left(\mathbf{x}(0), [\sigma^2(t) - \sigma^2(0)] \cdot \mathbf{I}\right)$$

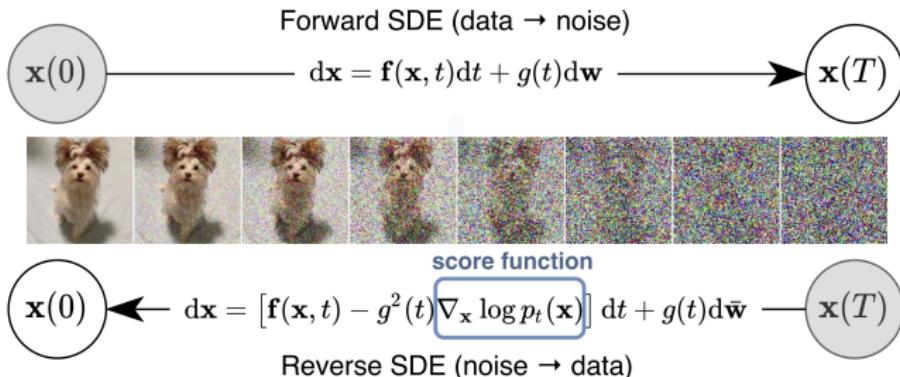
## DDPM

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}\left(\mathbf{x}(0)e^{-\frac{1}{2} \int_0^t \beta(s) ds}, \left(1 - e^{-\int_0^t \beta(s) ds}\right) \cdot \mathbf{I}\right)$$

# Score-based generative models through SDEs

## Sampling

Solve reverse SDE using numerical solvers (ODESolve).



- ▶ Discretization of the reverse SDE gives us the ancestral sampling.
- ▶ If we use probability flow instead of SDE than the reverse ODE gives us the DDIM sampling.

# Outline

1. Score-based generative models through SDEs

2. Flow Matching

3. Conditional Flow Matching

Conical gaussian paths

Linear interpolation

4. Rectified flows

## Continuous-in-time NF

Let consider ODE dynamic  $\mathbf{x}(t)$  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 time direction is the same as for CNF (opposite to score-based SDE models).

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

## KFP theorem (continuity equation)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\operatorname{div}(\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) \Leftrightarrow \frac{d \log p_t(\mathbf{x}(t))}{dt} = -\operatorname{tr}\left(\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right)$$

# Continuous-in-time NF

## KFP theorem (continuity equation)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\text{div}(\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) \Leftrightarrow \frac{d \log p_t(\mathbf{x}(t))}{dt} = -\text{tr}\left(\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right)$$

- ▶ If we know the true vector field  $\mathbf{f}(\mathbf{x}, t)$ , then KFP equation (or continuity equation) gives us the way to compute the density  $p_t(\mathbf{x})$ .
- ▶ Solving the continuity equation using the adjoint method is complicated and unstable process.
- ▶ **Flow matching** generalizes these models and gives the alternative way to solve the Neural ODE.

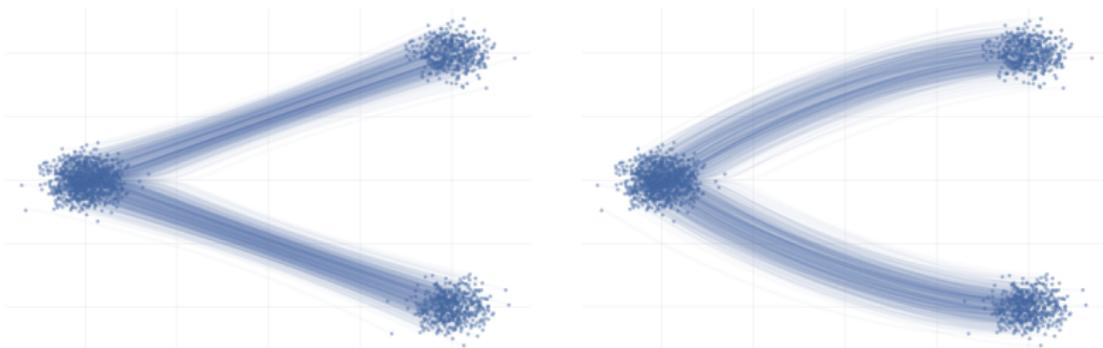
## Flow Matching

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

# Flow Matching

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

- ▶ There exists infinite number of possible  $\mathbf{f}(\mathbf{x}, t)$  between  $\pi(\mathbf{x})$  and  $p(\mathbf{x})$ .
- ▶ The true vector field  $\mathbf{f}(\mathbf{x}, t)$  is **unknown**.
- ▶ We need to select the "best"  $\mathbf{f}(\mathbf{x}, t)$  and makes the objective tractable.



# Outline

1. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

# Flow Matching

## Latent variable model

Let introduce the latent variable  $\mathbf{z}$ :

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

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

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

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

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

What is the relationship between  $\mathbf{f}(\mathbf{x}, t)$  and  $\mathbf{f}(\mathbf{x}, \mathbf{z}, t)$ ?

# Flow Matching

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

## Theorem

The following vector field generates the probability path  $p_t(\mathbf{x})$ .

$$\mathbf{f}(\mathbf{x}, t) = \mathbb{E}_{p_t(\mathbf{z}|\mathbf{x})} \mathbf{f}(\mathbf{x}, \mathbf{z}, t) = \int \mathbf{f}(\mathbf{x}, \mathbf{z}, t) \frac{p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z})}{p_t(\mathbf{x})} d\mathbf{z}$$

## Proof

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

# Flow Matching

## Flow Matching (FM)

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

## Conditional Flow Matching (CFM)

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

### 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.

# Conditional Flow Matching

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

## Idea of CFM

- ▶ We do not want to model  $p_t(\mathbf{x})$  because it is complex.
- ▶ We showed that it is possible to solve CFM task instead of FM task.
- ▶ Let choose the convenient conditioning latent variable  $\mathbf{z}$ .
- ▶ Let parametrize  $p_t(\mathbf{x}|\mathbf{z})$  instead of  $p_t(\mathbf{x})$ . It must satisfy the following constraints

$$p(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_0(\mathbf{x}|\mathbf{z}); \quad \pi(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_1(\mathbf{x}|\mathbf{z})$$

## What is left?

- ▶ How to choose the conditioning latent variable  $\mathbf{z}$ ?
- ▶ How to define  $p_t(\mathbf{x}|\mathbf{z})$  which follows the constraints?

# 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:

$$\mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{z}) + \boldsymbol{\sigma}_t(\mathbf{z}) \odot \mathbf{x}_0, \quad \mathbf{x}_0 \sim p_0(\mathbf{x}) = \mathcal{N}(0, \mathbf{I})$$

## Theorem

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

## Proof

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{z}, t) \quad \mathbf{x}_0 = \frac{1}{\boldsymbol{\sigma}_t(\mathbf{z})} \odot (\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{z}))$$

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

Tong et al. Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport, 2023

# Outline

1. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

# Flow Matching

## Conditional Flow Matching

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

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

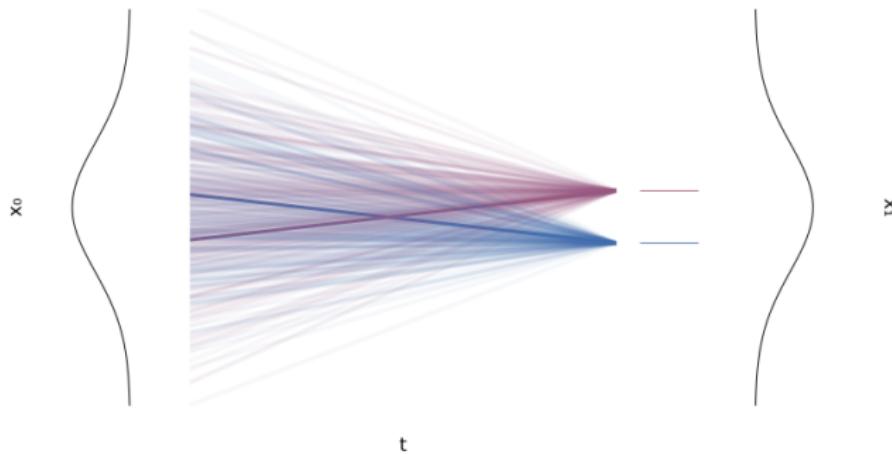
$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_1)); \quad \mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{x}_1) + \boldsymbol{\sigma}_t^2(\mathbf{x}_1) \odot \mathbf{x}_0.$$

# 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

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

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

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

# Flow Matching

## Conditional Flow Matching

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

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

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

Variance Exploding SDE probability path

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

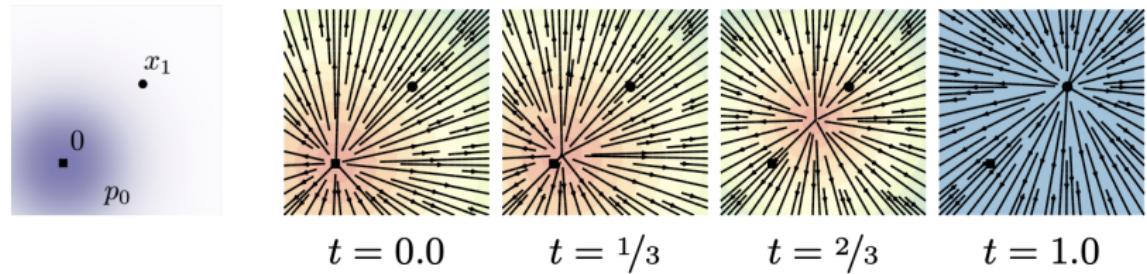
Variance Preserving SDE probability path

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

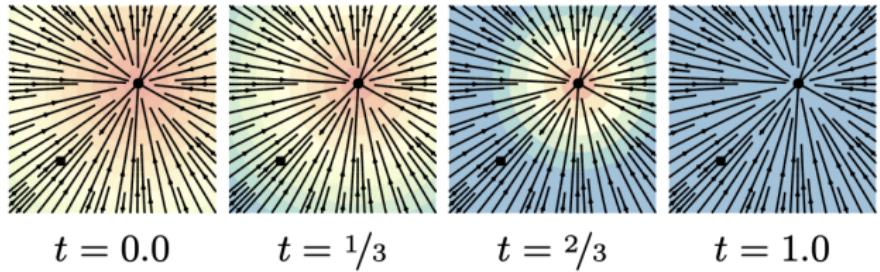
Here  $\alpha_t = \exp\left(-\frac{1}{2} \int_0^t \beta(s)ds\right)$ .

# Flow Matching

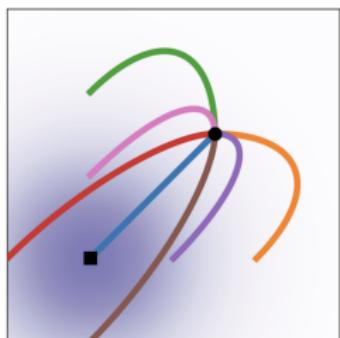
## Diffusion vector field



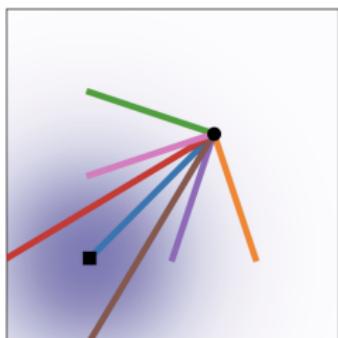
## Flow matching vector field



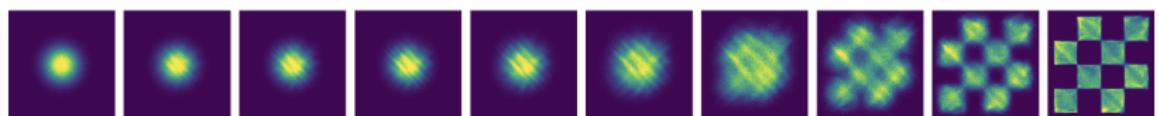
# Flow Matching



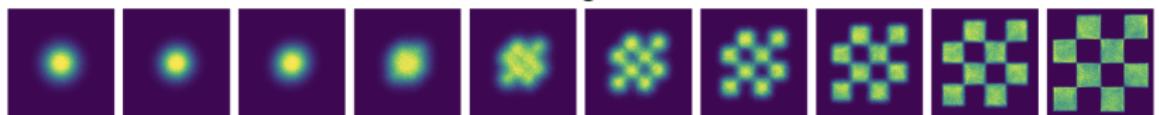
Diffusion



OT



Score matching w/ Diffusion



Flow Matching w/ OT

# Outline

1. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

# Flow Matching

## Conditional Flow Matching

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

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

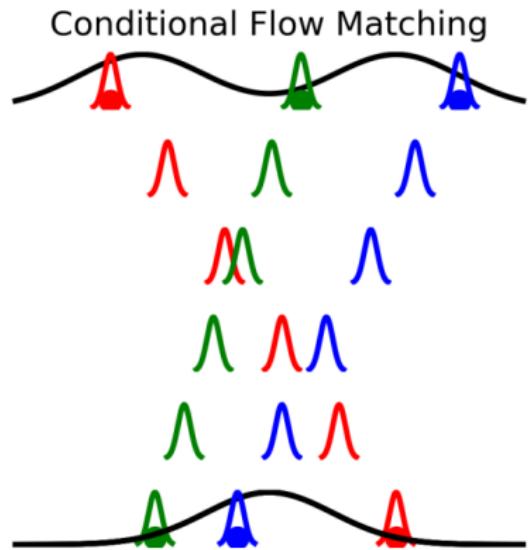
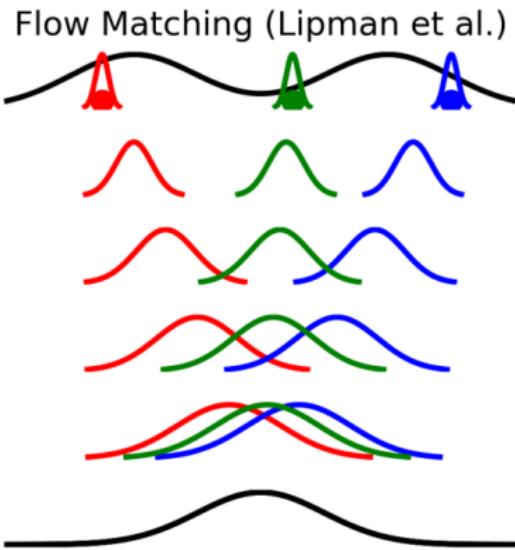
$$\begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases} \Rightarrow \begin{cases} \mu_0(\mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_0, & \sigma_0(\mathbf{x}_0, \mathbf{x}_1)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

Let consider straight conditional paths

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

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

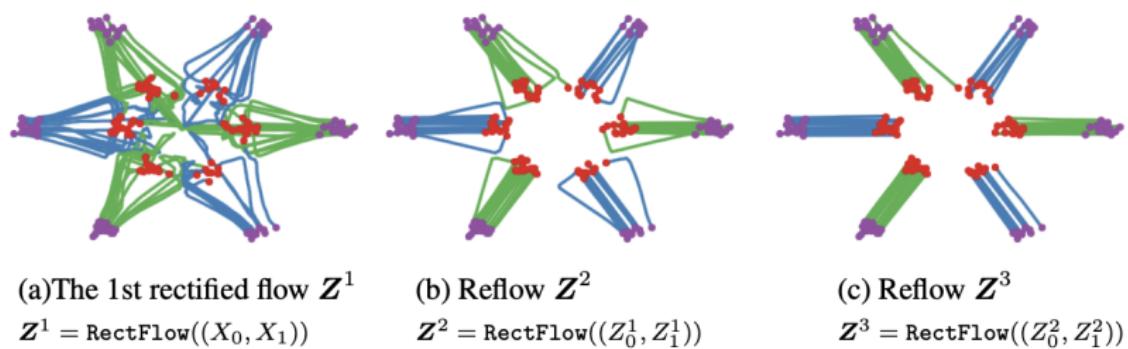
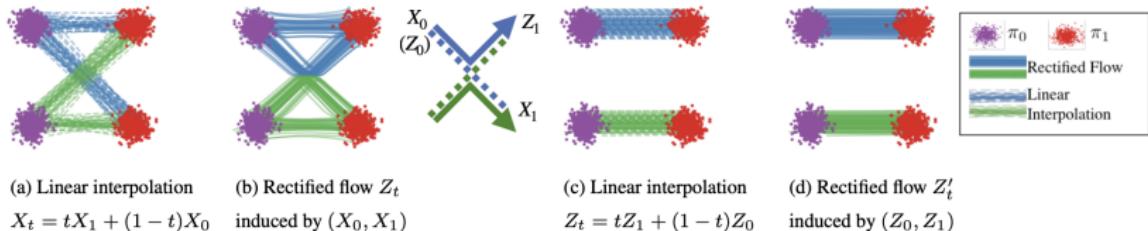
# Flow Matching



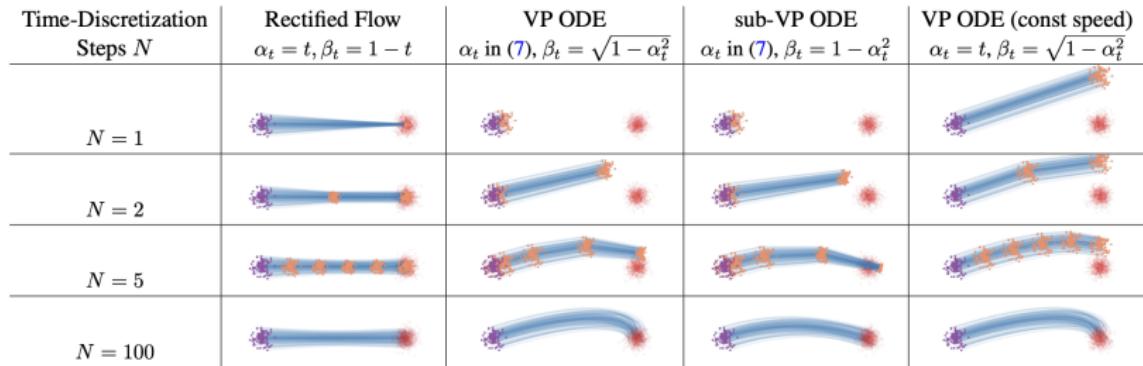
# Outline

1. Score-based generative models through SDEs
2. Flow Matching
3. Conditional Flow Matching
  - Conical gaussian paths
  - Linear interpolation
4. Rectified flows

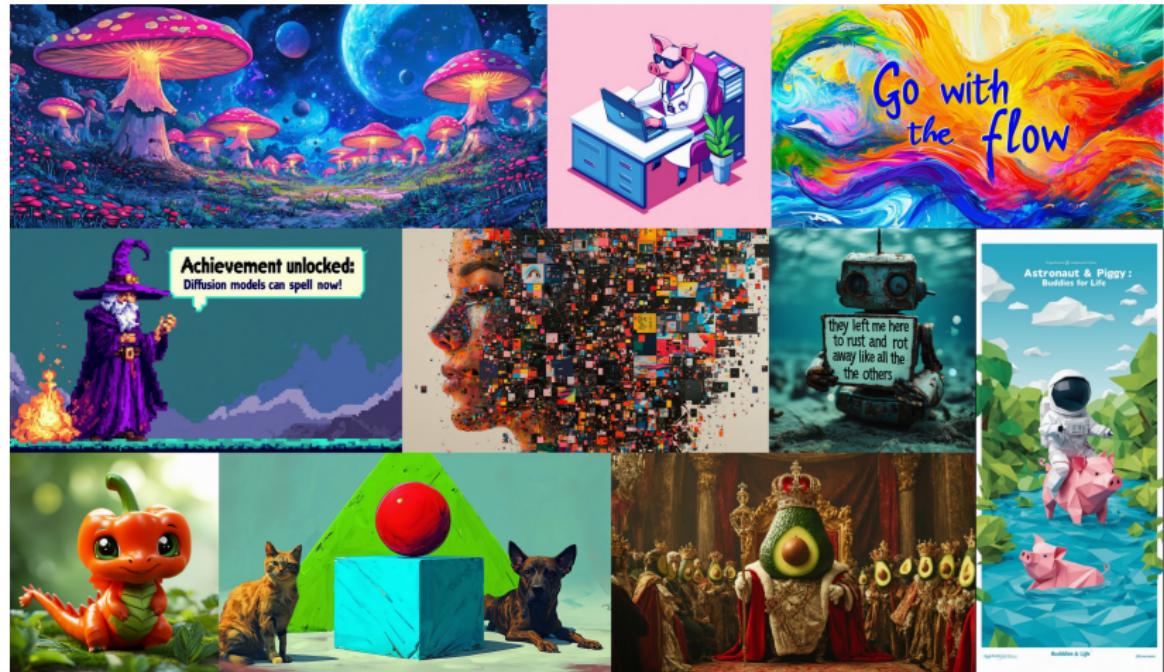
# Flow Matching



# Flow Matching



# Stable Diffusion 3



## Summary

- ▶ It is possible to train the continuous-in-time score-based generative models through forward and reverse SDEs.
- ▶