

# Deep Generative Models

## Lecture 10

Roman Isachenko



2026, Spring

# Recap of Previous Lecture

## DDPM Objective

$$\mathbb{E}_{p_{\text{data}}(\mathbf{x}_0)} \mathbb{E}_{t \sim U\{1, T\}} \mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} \left[ \frac{(1 - \alpha_t)^2}{2\tilde{\beta}_t \alpha_t} \left\| \mathbf{s}_{\theta, t}(\mathbf{x}_t) - \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t | \mathbf{x}_0) \right\|_2^2 \right]$$

In practice, this coefficient is usually omitted.

## NCSN Objective

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

**Note:** The objectives of DDPM and NCSN are almost identical; however, their sampling procedures differ:

- ▶ NCSN utilizes annealed Langevin dynamics,
- ▶ DDPM employs ancestral sampling.

# Recap of Previous Lecture

## Unconditional Generation

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} \cdot \mathbf{x}_t + \frac{\beta_t}{\sqrt{1 - \beta_t}} \cdot \nabla_{\mathbf{x}_t} \log p_{\theta}(\mathbf{x}_t) + \sigma_t \cdot \epsilon$$

## Conditional Generation

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} \cdot \mathbf{x}_t + \frac{\beta_t}{\sqrt{1 - \beta_t}} \cdot \nabla_{\mathbf{x}_t} \log p_{\theta}(\mathbf{x}_t | \mathbf{y}) + \sigma_t \cdot \epsilon$$

## Conditional Distribution

$$\begin{aligned}\nabla_{\mathbf{x}_t} \log p_{\theta}(\mathbf{x}_t | \mathbf{y}) &= \nabla_{\mathbf{x}_t} \log p_{\theta}(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t) \\ &= \mathbf{s}_{\theta, t}(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t)\end{aligned}$$

Here,  $p(\mathbf{y} | \mathbf{x}_t)$  denotes a classifier operating on noisy samples (which must be trained separately).

# Recap of Previous Lecture

## Guided Score Function

$$\mathbf{s}_{\theta,t}(\mathbf{x}_t, \mathbf{y}) = \mathbf{s}_{\theta,t}(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$$

$$\mathbf{s}_{\theta,t}^{\gamma}(\mathbf{x}_t, \mathbf{y}) = \mathbf{s}_{\theta,t}(\mathbf{x}_t) + \gamma \cdot \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$$

**Note:** Increasing  $\gamma$  sharpens  $p(\mathbf{y}|\mathbf{x}_t)$ , making it more contrast

$$\hat{p}(\mathbf{y}|\mathbf{x}_t) \propto p(\mathbf{y}|\mathbf{x}_t)^{\gamma}.$$

## Training

- ▶ Train the DDPM as before.
- ▶ Train an additional classifier  $p(\mathbf{y}|\mathbf{x}_t)$  on noisy data.

## Guided Sampling

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1-\beta_t}} \cdot \mathbf{x}_t + \frac{\beta_t}{\sqrt{1-\beta_t}} \cdot \mathbf{s}_{\theta,t}^{\gamma}(\mathbf{x}_t, \mathbf{y}) + \sigma_t \cdot \epsilon$$

## Recap of Previous Lecture

The previous method requires an additional classifier  $p(\mathbf{y}|\mathbf{x}_t)$  trained on noisy data. Let's try to avoid this requirement.

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) = \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t|\mathbf{y}) - \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t)$$

$$\begin{aligned}\nabla_{\mathbf{x}_t}^\gamma \log p_\theta(\mathbf{x}_t|\mathbf{y}) &= \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t) + \gamma \cdot \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) = \\ &= (1 - \gamma) \cdot \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t) + \gamma \cdot \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t|\mathbf{y})\end{aligned}$$

## Scaled Guided Score Function

$$\mathbf{s}_{\theta,t}^\gamma(\mathbf{x}_t, \mathbf{y}) = (1 - \gamma) \cdot \mathbf{s}_{\theta,t}(\mathbf{x}_t) + \gamma \cdot \mathbf{s}_{\theta,t}(\mathbf{x}_t, \mathbf{y})$$

- ▶ Introduce  $\mathbf{y} = \emptyset$  for  $\mathbf{s}_{\theta,t}(\mathbf{x}_t) = \mathbf{s}_{\theta,t}(\mathbf{x}_t, \emptyset)$ .
- ▶ Drop the labels  $\mathbf{y}$  with some fixed probability.
- ▶ Apply the model twice during inference to get  $\mathbf{s}_{\theta,t}(\mathbf{x}_t, \emptyset)$  and  $\mathbf{s}_{\theta,t}(\mathbf{x}_t, \mathbf{y})$ .

## Recap of Previous Lecture

### Continuous-Time Dynamics (Neural ODE)

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}_\theta(\mathbf{x}(t), t); \quad \text{where } \mathbf{x}(t_0) = \mathbf{x}_0.$$

$$\psi_t(\mathbf{x}_0) = \int_{t_0}^{t_1} \mathbf{f}_\theta(\mathbf{x}(t), t) dt + \mathbf{x}_0 \approx \texttt{ODESolve}_f(\mathbf{x}_0, \theta, t_0, t_1).$$

- ▶  $\mathbf{f}_\theta : \mathbb{R}^m \times [t_0, t_1] \rightarrow \mathbb{R}^m$  is a vector field.
- ▶  $\psi : \mathbb{R}^m \times [t_0, t_1] \rightarrow \mathbb{R}^m$  is a flow (the solution of ODE):

### Euler Update Step (`ODESolve`)

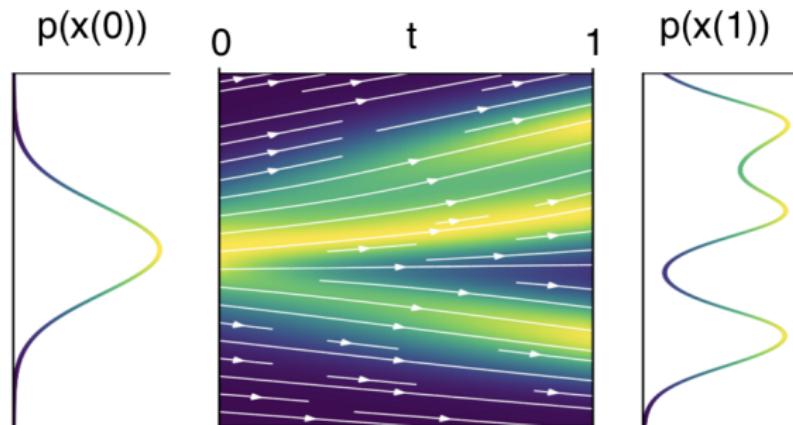
$$\mathbf{x}(t+h) = \mathbf{x}(t) + h \cdot \mathbf{f}_\theta(\mathbf{x}(t), t)$$

More advanced numerical methods (such as Runge-Kutta) are often used instead of unstable Euler update step.

## Recap of Previous Lecture

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}_\theta(\mathbf{x}(t), t); \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

- ▶ Suppose  $\mathbf{x}(0)$  is a random variable with density  $p_0(\mathbf{x})$ . Then,  $\mathbf{x}(t)$  is a random variable with density  $p_t(\mathbf{x})$ .
- ▶  $p_t(\mathbf{x}) = p(\mathbf{x}, t)$  describes the **probability path** between  $p_0(\mathbf{x})$  and  $p_1(\mathbf{x})$ .



## Recap of Previous Lecture

### Theorem (Picard)

If  $\mathbf{f}$  is uniformly Lipschitz continuous in  $\mathbf{x}$  and continuous in  $t$ , then the ODE has a **unique** solution.

This means the ODE can be **uniquely inverted**:

$$\mathbf{x}(1) = \mathbf{x}(0) + \int_0^1 \mathbf{f}_\theta(\mathbf{x}(t), t) dt$$

$$\mathbf{x}(0) = \mathbf{x}(1) + \int_1^0 \mathbf{f}_\theta(\mathbf{x}(t), t) dt$$

**Note:** Unlike discrete-time NF,  $\mathbf{f}$  need not be invertible (uniqueness ensures bijectivity).

How can we compute  $p_t(\mathbf{x})$  for any  $t$ ?

# Outline

1. Continuity Equation for NF Log-Likelihood
2. SDE Basics
3. Probability Flow ODE
4. Reverse SDE

# Outline

1. Continuity Equation for NF Log-Likelihood
2. SDE Basics
3. Probability Flow ODE
4. Reverse SDE

# Continuous-Time NF

## Theorem (Continuity Equation)

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

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

# Continuous-Time NF

## Theorem (Continuity Equation)

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

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

This result states: given  $\mathbf{x}_0 = \mathbf{x}(0)$ , the solution to the continuity equation gives the density  $p_1(\mathbf{x}(1))$ .

## Solution of the Continuity Equation

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

# Continuous-Time NF

## Theorem (Continuity Equation)

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

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

This result states: given  $\mathbf{x}_0 = \mathbf{x}(0)$ , the solution to the continuity equation gives the density  $p_1(\mathbf{x}(1))$ .

## Solution of the Continuity Equation

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

- ▶ This provides the density along the trajectory (not the total probability path).
- ▶ However, the latter term is difficult to estimate efficiently.

# Outline

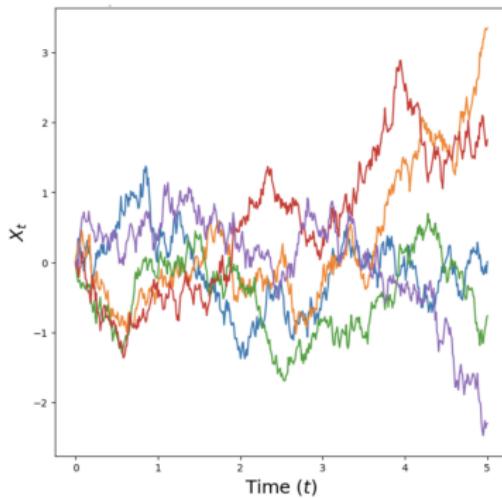
1. Continuity Equation for NF Log-Likelihood
2. SDE Basics
3. Probability Flow ODE
4. Reverse SDE

# Stochastic Differential Equation (SDE)

## Wiener Process

$w(t)$  is the standard Wiener process (Brownian motion), defined by:

1.  $w(0) = 0$  (almost surely);
2.  $w(t)$  has independent increments;
3.  $w(t)$  trajectories are continuous;
4.  $w(t) - w(s) \sim \mathcal{N}(0, (t-s)\mathbf{I})$  for  $t > s$ ;

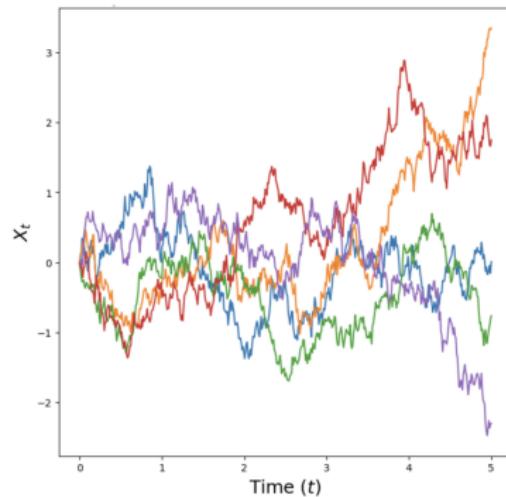


# Stochastic Differential Equation (SDE)

## Wiener Process

$\mathbf{w}(t)$  is the standard Wiener process (Brownian motion), defined by:

1.  $\mathbf{w}(0) = 0$  (almost surely);
2.  $\mathbf{w}(t)$  has independent increments;
3.  $\mathbf{w}(t)$  trajectories are continuous;
4.  $\mathbf{w}(t) - \mathbf{w}(s) \sim \mathcal{N}(0, (t-s)\mathbf{I})$  for  $t > s$ ;



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

## Stochastic Differential Equation (SDE)

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

## Stochastic Differential Equation (SDE)

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

Let's define a stochastic process  $\mathbf{x}(t)$  with initial condition  $\mathbf{x}(0) \sim p_0(\mathbf{x}) = p_{\text{data}}(\mathbf{x})$ :

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

# Stochastic Differential Equation (SDE)

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

Let's define a stochastic process  $\mathbf{x}(t)$  with initial condition  $\mathbf{x}(0) \sim p_0(\mathbf{x}) = p_{\text{data}}(\mathbf{x})$ :

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

- ▶  $\mathbf{f}(\mathbf{x}, t) : \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}^m$  is the **drift** term (vector field).
- ▶  $\mathbf{g}(t) : \mathbb{R} \rightarrow \mathbb{R}$  is the **diffusion** term (if  $\mathbf{g}(t) = 0$ , we recover the standard ODE).
- ▶  $\mathbf{w}(t)$  is the standard Wiener process ( $d\mathbf{w} = \epsilon \cdot \sqrt{dt}$ ).
- ▶ We do not have the flow  $\psi_t(\mathbf{x}_0)$  notion anymore, since trajectories are stochastic.

# Stochastic Differential Equation (SDE)

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

## Theorem

If  $\mathbf{f}$  is continuously differentiable with a bounded derivative in  $\mathbf{x}$  and continuous in  $t$  and  $g(t)$  is continuous then the SDE has the solution given by unique process  $\mathbf{x}(t)$ .

- ▶ Unlike ODEs, the initial condition  $\mathbf{x}(0)$  doesn't uniquely determine the trajectory.
- ▶ There are two sources of randomness:
  - ▶ the initial distribution  $p_0(\mathbf{x})$ ;
  - ▶ the Wiener process  $\mathbf{w}(t)$ .

# Stochastic Differential Equation (SDE)

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

# Stochastic Differential Equation (SDE)

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

Discretizing the SDE (Euler Update Step) – SDESolve

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

If  $dt = 1$ , then

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{f}(\mathbf{x}_t, t) + g(t) \cdot \epsilon$$

# Stochastic Differential Equation (SDE)

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

Discretizing the SDE (Euler Update Step) – SDESolve

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

If  $dt = 1$ , then

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{f}(\mathbf{x}_t, t) + g(t) \cdot \epsilon$$

- ▶ At any time  $t$ , the process has density  $p_t(\mathbf{x}) = p(\mathbf{x}, t)$ .
- ▶  $p : \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}_+$  specifies a **probability path** from  $p_0(\mathbf{x})$  to  $p_1(\mathbf{x})$ .
- ▶ How can we obtain the probability path  $p_t(\mathbf{x})$  for  $\mathbf{x}(t)$ ?

# Stochastic Differential Equation (SDE)

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}, \quad d\mathbf{w} = \boldsymbol{\epsilon} \cdot \sqrt{dt}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

Theorem (Kolmogorov-Fokker-Planck)

The evolution of  $p_t(\mathbf{x})$  is governed by

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

# Stochastic Differential Equation (SDE)

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}, \quad d\mathbf{w} = \boldsymbol{\epsilon} \cdot \sqrt{dt}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

Theorem (Kolmogorov-Fokker-Planck)

The evolution of  $p_t(\mathbf{x})$  is governed by

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

Here,

$$\text{div}(\mathbf{v}) = \sum_{i=1}^m \frac{\partial v_i(\mathbf{x})}{\partial x_i} = \text{tr}\left(\frac{\partial \mathbf{v}(\mathbf{x})}{\partial \mathbf{x}}\right)$$

$$\Delta_{\mathbf{x}}p_t(\mathbf{x}) = \sum_{i=1}^m \frac{\partial^2 p_t(\mathbf{x})}{\partial x_i^2} = \text{tr}\left(\frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}\right)$$

# Stochastic Differential Equation (SDE)

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}, \quad d\mathbf{w} = \boldsymbol{\epsilon} \cdot \sqrt{dt}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}).$$

Theorem (Kolmogorov-Fokker-Planck)

The evolution of  $p_t(\mathbf{x})$  is governed by

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

Here,

$$\text{div}(\mathbf{v}) = \sum_{i=1}^m \frac{\partial v_i(\mathbf{x})}{\partial x_i} = \text{tr}\left(\frac{\partial \mathbf{v}(\mathbf{x})}{\partial \mathbf{x}}\right)$$

$$\Delta_{\mathbf{x}}p_t(\mathbf{x}) = \sum_{i=1}^m \frac{\partial^2 p_t(\mathbf{x})}{\partial x_i^2} = \text{tr}\left(\frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}\right)$$

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

# Stochastic Differential Equation (SDE)

## Theorem (Kolmogorov-Fokker-Planck)

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

- ▶ The KFP theorem is necessary and sufficient condition (it is uniquely defines  $p_t(\mathbf{x})$ ).
- ▶ This generalizes the continuity equation for continuous-time NF:

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

# Stochastic Differential Equation (SDE)

## Theorem (Kolmogorov-Fokker-Planck)

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

- ▶ The KFP theorem is necessary and sufficient condition (it is uniquely defines  $p_t(\mathbf{x})$ ).
- ▶ This generalizes the continuity equation for continuous-time NF:

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

### Special Case: constant density $p_t(\mathbf{x})$

Let find the SDE for which  $p_t(\mathbf{x}) = \text{const}$  (i.e., if  $\mathbf{x}(0) \sim p_0(\mathbf{x})$ , then  $\mathbf{x}(t) \sim p_0(\mathbf{x})$ ).

# Stochastic Differential Equation (SDE)

## Theorem (Kolmogorov-Fokker-Planck)

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

- ▶ The KFP theorem is necessary and sufficient condition (it is uniquely defines  $p_t(\mathbf{x})$ ).
- ▶ This generalizes the continuity equation for continuous-time NF:

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

### Special Case: constant density $p_t(\mathbf{x})$

Let find the SDE for which  $p_t(\mathbf{x}) = \text{const}$  (i.e., if  $\mathbf{x}(0) \sim p_0(\mathbf{x})$ , then  $\mathbf{x}(t) \sim p_0(\mathbf{x})$ ).

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \Leftrightarrow \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

## Langevin SDE (Special Case)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \quad \Leftrightarrow \quad \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

## Langevin SDE (Special Case)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \quad \Leftrightarrow \quad \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

$$\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] = \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}$$

## Langevin SDE (Special Case)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \quad \Leftrightarrow \quad \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

$$\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] = \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}$$

$$\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x}) = \frac{1}{2} g^2(t) \frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}}$$

## Langevin SDE (Special Case)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \Leftrightarrow \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

$$\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] = \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}$$

$$\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x}) = \frac{1}{2} g^2(t) \frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}}$$

$$\mathbf{f}(\mathbf{x}, t) = \frac{1}{2} g^2(t) \frac{1}{p_t(\mathbf{x})} \frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}} = \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$$

## Langevin SDE (Special Case)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = 0 \Leftrightarrow \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] + \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = 0$$

$$\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x})] = \frac{1}{2} g^2(t) \frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2}$$

$$\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x}) = \frac{1}{2} g^2(t) \frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}}$$

$$\mathbf{f}(\mathbf{x}, t) = \frac{1}{2} g^2(t) \frac{1}{p_t(\mathbf{x})} \frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}} = \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$$

Let  $g(t) = 1$ , then  $\mathbf{f}(\mathbf{x}, t) = \frac{1}{2} \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$ .

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

## Langevin SDE (Special Case)

Let find the SDE for which  $p_t(\mathbf{x}) = \text{const}$  (i.e., if  $\mathbf{x}(0) \sim p_0(\mathbf{x})$ , then  $\mathbf{x}(t) \sim p_0(\mathbf{x})$ ).

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

## Langevin SDE (Special Case)

Let find the SDE for which  $p_t(\mathbf{x}) = \text{const}$  (i.e., if  $\mathbf{x}(0) \sim p_0(\mathbf{x})$ , then  $\mathbf{x}(t) \sim p_0(\mathbf{x})$ ).

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

## Discretized Langevin SDE

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \frac{\eta}{2} \cdot \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) + \sqrt{\eta} \cdot \boldsymbol{\epsilon}, \quad \eta \approx dt.$$

## Langevin Dynamic

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \frac{\eta}{2} \cdot \nabla_{\mathbf{x}} \log p_{\theta}(\mathbf{x}) + \sqrt{\eta} \cdot \boldsymbol{\epsilon}, \quad \eta \approx dt.$$

We (partially) explained, why Langevin dynamics is working.

# Outline

1. Continuity Equation for NF Log-Likelihood
2. SDE Basics
3. Probability Flow ODE
4. Reverse SDE

# Probability Flow ODE

## ODE and Continuity Equation

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

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

The only source of randomness is the initial distribution  $p_0(\mathbf{x})$ .

# Probability Flow ODE

## ODE and Continuity Equation

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

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

The only source of randomness is the initial distribution  $p_0(\mathbf{x})$ .

## SDE and KFP Equation

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

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

Now there are two sources of randomness: the initial distribution  $p_0(\mathbf{x})$  and the Wiener process  $\mathbf{w}(t)$ .

# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

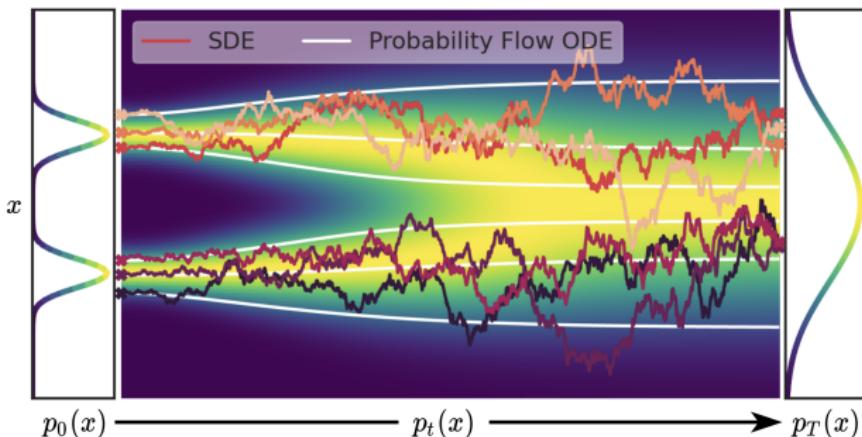
$$d\mathbf{x} = \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) dt$$

# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

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



# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

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

## Proof

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

# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

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

## Proof

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})] + \frac{1}{2}g^2(t)\frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{1}{2}g^2(t)\frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}} \right] \right)\end{aligned}$$

# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

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

## Proof

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})] + \frac{1}{2}g^2(t)\frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{1}{2}g^2(t)\frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}} \right] \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{1}{2}g^2(t)p_t(\mathbf{x})\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right] \right)\end{aligned}$$

# Probability Flow ODE

## Theorem

Suppose the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$  induces the probability path  $p_t(\mathbf{x})$ . Then, there exists an ODE with the same probability path  $p_t(\mathbf{x})$ , given by

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

## Proof

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} [\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})] + \frac{1}{2}g^2(t)\frac{\partial^2 p_t(\mathbf{x})}{\partial \mathbf{x}^2} \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{1}{2}g^2(t)\frac{\partial p_t(\mathbf{x})}{\partial \mathbf{x}} \right] \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{1}{2}g^2(t)p_t(\mathbf{x})\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right] \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) \right] \right)\end{aligned}$$

# Probability Flow ODE

## Proof (Continued)

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

# Probability Flow ODE

## Proof (Continued)

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) \right] \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \tilde{\mathbf{f}}(\mathbf{x}, t) p_t(\mathbf{x}) \right] \right) = -\text{div} \left( \tilde{\mathbf{f}}(\mathbf{x}, t) p_t(\mathbf{x}) \right)\end{aligned}$$

# Probability Flow ODE

## Proof (Continued)

$$\begin{aligned}\frac{\partial p_t(\mathbf{x})}{\partial t} &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) \right] \right) = \\ &= \text{tr} \left( -\frac{\partial}{\partial \mathbf{x}} \left[ \tilde{\mathbf{f}}(\mathbf{x}, t) p_t(\mathbf{x}) \right] \right) = -\text{div} \left( \tilde{\mathbf{f}}(\mathbf{x}, t) p_t(\mathbf{x}) \right)\end{aligned}$$

$$\tilde{\mathbf{f}}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}); \quad \tilde{g}(t) = 0$$

$$d\mathbf{x} = \tilde{\mathbf{f}}(\mathbf{x}, t) dt + 0 \cdot d\mathbf{w} = \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g^2(t) \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) dt$$

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

## Probability Flow ODE

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

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

## Probability Flow ODE

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

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

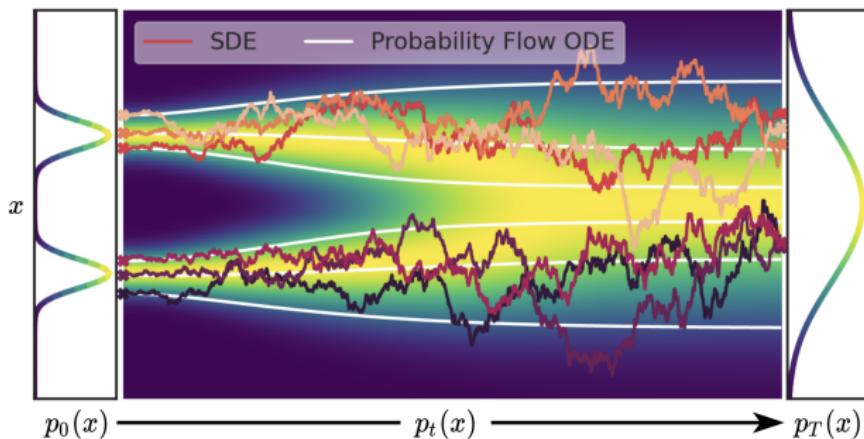
- ▶ The term  $\mathbf{s}(\mathbf{x}, t) = \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$  is the score function in continuous time.

# Probability Flow ODE

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

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

- ▶ The term  $\mathbf{s}(\mathbf{x}, t) = \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$  is the score function in continuous time.
- ▶ The ODE produces more stable trajectories.



# Outline

1. Continuity Equation for NF Log-Likelihood
2. SDE Basics
3. Probability Flow ODE
4. Reverse SDE

## Reverse SDE

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

Here  $dt$  can be  $> 0$  or  $< 0$ .

## Reverse SDE

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

Here  $dt$  can be  $> 0$  or  $< 0$ .

## Reverse ODE

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

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

## Reverse SDE

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

Here  $dt$  can be  $> 0$  or  $< 0$ .

## Reverse ODE

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

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

- ▶ How do we reverse the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ ?
- ▶ The Wiener process introduces randomness that must be reversed.

## Reverse SDE

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

Here  $dt$  can be  $> 0$  or  $< 0$ .

## Reverse ODE

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

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

- ▶ How do we reverse the SDE  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ ?
- ▶ The Wiener process introduces randomness that must be reversed.

## Theorem

There exists a reverse SDE for  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ , given by:

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

where  $dt < 0$ .

# Reverse SDE

## Theorem

There exists a reverse SDE for  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ , given by:

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

where  $dt < 0$ .

# Reverse SDE

## Theorem

There exists a reverse SDE for  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ , given by:

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

where  $dt < 0$ .

**Note:** Again, the score function appears:  $\mathbf{s}(\mathbf{x}, t) = \frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x})$ .

## Proof Sketch

- ▶ Convert the initial SDE to a probability flow ODE.
- ▶ Reverse the probability flow ODE.
- ▶ Convert the reversed probability flow ODE back to an SDE.

# Reverse SDE

## Proof

- ▶ Convert the initial SDE to a probability flow ODE:

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

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

# Reverse SDE

## Proof

- ▶ Convert the initial SDE to a probability flow ODE:

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

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

- ▶ Reverse the probability flow ODE:

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

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + \frac{1}{2}g^2(1 - \tau)\frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau$$

# Reverse SDE

## Proof

- ▶ Convert the initial SDE to a probability flow ODE:

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

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

- ▶ Reverse the probability flow ODE:

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

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + \frac{1}{2}g^2(1 - \tau)\frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau$$

- ▶ Convert the reversed probability flow ODE back to an SDE:

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + \frac{1}{2}g^2(1 - \tau)\frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau$$

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + g^2(1 - \tau)\frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau + g(1 - \tau)d\mathbf{w}$$

# Reverse SDE

## Theorem

There exists a reverse SDE for  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ , given by:

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

where  $dt < 0$ .

## Proof (Continued)

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + g^2(1 - \tau) \frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau + g(1 - \tau)d\mathbf{w}$$

# Reverse SDE

## Theorem

There exists a reverse SDE for  $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ , given by:

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

where  $dt < 0$ .

## Proof (Continued)

$$d\mathbf{x} = \left( -\mathbf{f}(\mathbf{x}, 1 - \tau) + g^2(1 - \tau) \frac{\partial}{\partial \mathbf{x}} \log p_{1-\tau}(\mathbf{x}) \right) d\tau + g(1 - \tau)d\mathbf{w}$$

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

Here  $d\tau > 0$  and  $dt < 0$ .

## Reverse SDE

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

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

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

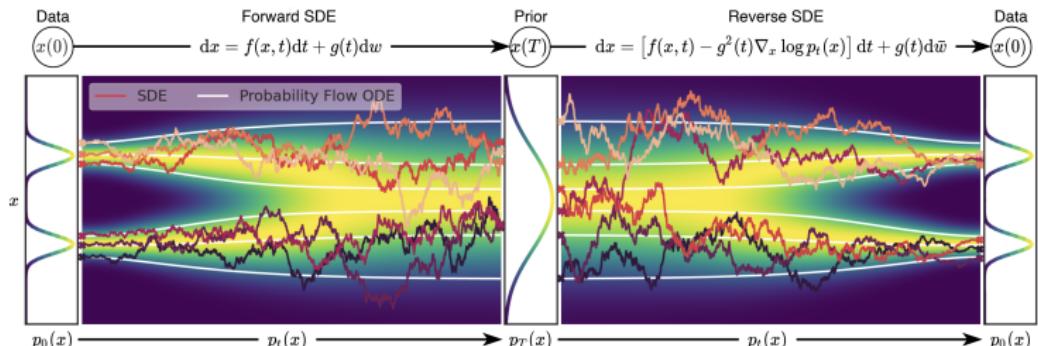
# Reverse SDE

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

$$dx = \left( \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g^2(t)\frac{\partial}{\partial \mathbf{x}} \log p_t(\mathbf{x}) \right) dt \quad - \text{Probability Flow ODE}$$

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

- ▶ This framework allows us to transform one distribution into another via an SDE with a prescribed probability path  $p_t(\mathbf{x})$ .
- ▶ We can invert this process using the score function.



## Summary

- ▶ The continuity equation allows us to compute  $\log p(\mathbf{x}, t)$  at any time  $t$ .
- ▶ An SDE defines a stochastic process with drift and diffusion terms; ODEs are a special case of SDEs.
- ▶ The KFP equation describes the probability dynamics of an SDE.
- ▶ The Langevin SDE preserves a constant probability path.
- ▶ Every SDE admits a corresponding probability flow ODE following the same probability path.
- ▶ SDEs can be reversed using the score function.