

Deep Generative Models

Lecture 13

Roman Isachenko

Moscow Institute of Physics and Technology
Yandex School of Data Analysis

2025, Autumn

Recap of Previous Lecture

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

Variance Exploding SDE (NCSN)

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

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

Variance Preserving SDE (DDPM)

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

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

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

Recap of Previous Lecture

Discrete-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-Time Objective

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

NCSN

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

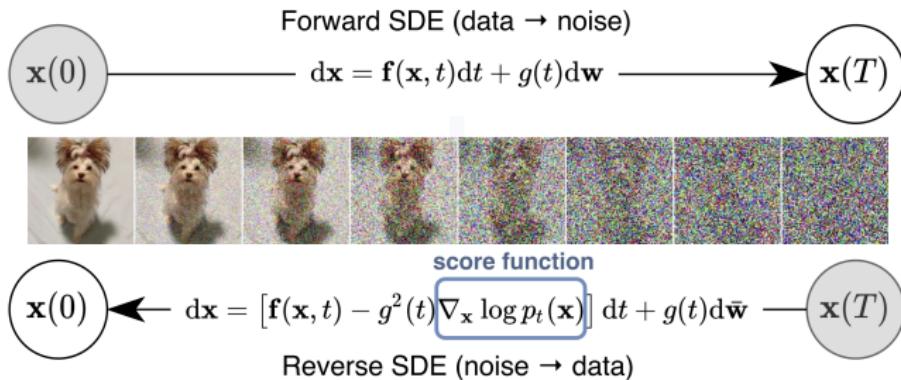
DDPM

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

Recap of Previous Lecture

Sampling

To sample, solve the reverse SDE using numerical solvers (SDESolve).



- ▶ Discretizing the reverse SDE gives us ancestral sampling.
- ▶ If we use the probability flow ODE instead, then the reverse ODE yields DDIM sampling.

Recap of Previous Lecture

Consider ODE dynamics $\mathbf{x}(t)$ in the interval $t \in [0, 1]$ with $\mathbf{x}_0 \sim p_0(\mathbf{x}) = p(\mathbf{x})$, $\mathbf{x}_1 \sim p_1(\mathbf{x}) = \pi(\mathbf{x})$.

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

KFP Theorem (Continuity Equation)

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

Solving the continuity equation using the adjoint method is complicated and unstable.

Flow Matching

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

Recap of Previous Lecture

Introduce the latent variable \mathbf{z} :

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

- ▶ $p_t(\mathbf{x}|\mathbf{z})$ is a **conditional probability path**
- ▶ $\mathbf{f}(\mathbf{x}, \mathbf{z}, t)$ is a **conditional vector field**

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

Theorem

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

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

Recap of Previous Lecture

Flow Matching (FM)

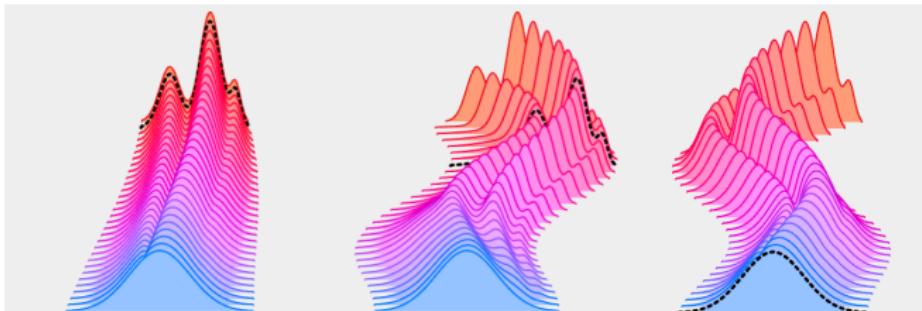
$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x})} \|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

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 the FM objective equals the optimum for CFM.



Outline

1. Conditional Flow Matching
2. Conical Gaussian Paths
3. Linear Interpolation
4. Link with Score-Based Models

Outline

1. Conditional Flow Matching
2. Conical Gaussian Paths
3. Linear Interpolation
4. Link with Score-Based Models

Conditional Flow Matching

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

$$p(\mathbf{x}) = \mathcal{N}(0, \mathbf{I}) = \mathbb{E}_{p(\mathbf{z})} p_0(\mathbf{x}|\mathbf{z}); \quad \pi(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_1(\mathbf{x}|\mathbf{z}).$$

Training

1. Sample a time $t \sim U[0, 1]$ and $\mathbf{z} \sim p(\mathbf{z})$.
2. Draw $\mathbf{x}_t \sim p_t(\mathbf{x}|\mathbf{z})$.
3. Compute the loss $\mathcal{L} = \|\mathbf{f}(\mathbf{x}_t, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}_t, t)\|^2$.

Sampling

1. Sample $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})$.
2. Solve the ODE to obtain \mathbf{x}_1 :

$$\mathbf{x}_1 = \text{ODESolve}_f(\mathbf{x}_0, \theta, t_0 = 0, t_1 = 1).$$

Conditional Flow Matching

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

Open Questions

- ▶ How should we choose the conditioning latent variable \mathbf{z} ?
- ▶ How can we define $p_t(\mathbf{x}|\mathbf{z})$ to enforce the constraints?

Gaussian Conditional Probability Path (Choice 2)

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

- ▶ There are infinitely many vector fields that generate a particular probability path.
- ▶ Let's consider the following dynamics:

$$\mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{z}) + \boldsymbol{\sigma}_t(\mathbf{z}) \odot \boldsymbol{\epsilon}, \quad \text{with fixed } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$$

Flow Matching

Gaussian Conditional Probability Path

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

Theorem

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

Proof

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

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

Outline

1. Conditional Flow Matching
2. Conical Gaussian Paths
3. Linear Interpolation
4. Link with Score-Based Models

Endpoint Conditioning

Conditional Flow Matching

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

Conditioning Latent Variable (Choice 1)

Let us 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 the boundary constraints:

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

Conical Gaussian Paths

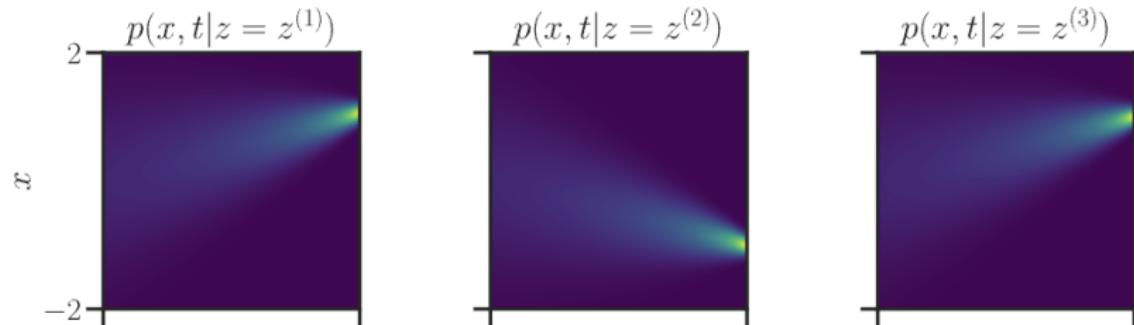
$$p_0(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(0, \mathbf{I}); \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1)$$

Gaussian Conditional Probability Path

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

Let's consider straight conditional paths:

$$\begin{cases} \boldsymbol{\mu}_t(\mathbf{x}_1) = t\mathbf{x}_1 \\ \boldsymbol{\sigma}_t(\mathbf{x}_1) = 1 - t \end{cases} \Rightarrow \begin{cases} p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2 \cdot \mathbf{I}) \\ \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0 \end{cases}$$



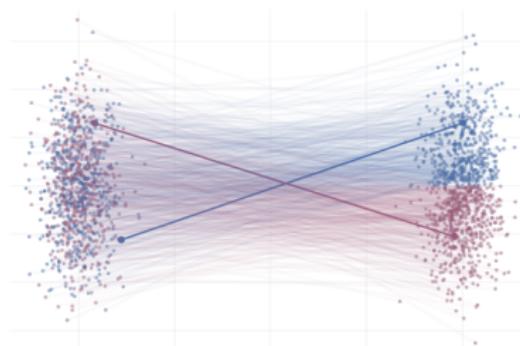
Conical Gaussian Paths

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2 \mathbf{I}); \quad \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0$$

Conditional Vector Field

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

$$\begin{aligned}\mathbf{f}(\mathbf{x}_t, \mathbf{x}_1, t) &= \mathbf{x}_1 - \frac{1}{1-t} \cdot (\mathbf{x}_t - t\mathbf{x}_1) = \frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} = \\ &= \frac{\mathbf{x}_1 - t\mathbf{x}_1 - (1-t)\mathbf{x}_0}{1-t} = \mathbf{x}_1 - \mathbf{x}_0\end{aligned}$$



The conditional vector field $\mathbf{f}(\mathbf{x}_t, \mathbf{x}_1, t)$ defines straight lines between $\pi(\mathbf{x})$ and $\mathcal{N}(0, \mathbf{I})$.

Conical Gaussian Paths

$$\begin{aligned} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \| \mathbf{f}(\mathbf{x}, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}, t) \|^2 = \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_t \sim p_t(\mathbf{x}|\mathbf{x}_1)} \left\| \left(\frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} \right) - \mathbf{f}_\theta(\mathbf{x}_t, t) \right\|^2 = \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})} \| (\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_\theta(t\mathbf{x}_1 + (1-t)\mathbf{x}_0, t) \|^2 \end{aligned}$$

- ▶ We fit straight lines between the noise distribution $p(\mathbf{x})$ and the data distribution $\pi(\mathbf{x})$.
- ▶ The **marginal** path $p_t(\mathbf{x})$ does not give straight lines.



Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})} \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_{\theta}(\mathbf{x}_t, t)\|^2 \rightarrow \min_{\theta}$$

Training

1. Sample $\mathbf{x}_1 \sim \pi(\mathbf{x})$.
2. Sample time $t \sim U[0, 1]$ and $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})$.
3. Obtain the noisy image $\mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$.
4. Compute the loss $\mathcal{L} = \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_{\theta}(\mathbf{x}, t)\|^2$.

Sampling

1. Sample $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})$.
2. Solve the ODE to obtain \mathbf{x}_1 :

$$\mathbf{x}_1 = \text{ODESolve}_f(\mathbf{x}_0, \theta, t_0 = 0, t_1 = 1)$$

Outline

1. Conditional Flow Matching
2. Conical Gaussian Paths
3. Linear Interpolation
4. Link with Score-Based Models

Pair Conditioning

Conditional Flow Matching

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

Conditioning Latent Variable (Choice 1)

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

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

We must enforce boundary constraints:

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

Linear Interpolation

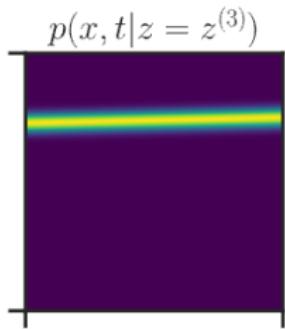
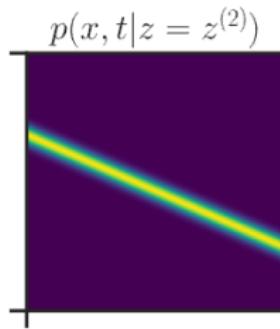
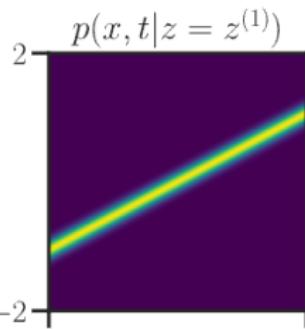
$$p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \quad p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1)$$

Gaussian Conditional Probability Path

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

Let's consider straight conditional paths:

$$\begin{cases} \boldsymbol{\mu}_t(\mathbf{x}_1) = t\mathbf{x}_1 + (1-t)\mathbf{x}_0 \\ \boldsymbol{\sigma}_t(\mathbf{x}_1) = \epsilon \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0) \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1) \end{cases}$$



Flow Matching: Conical Paths vs. Linear Interpolation

$$z = x_1$$

$$p_t(x|x_1) = \mathcal{N}(tx_1, (1-t)^2 I)$$

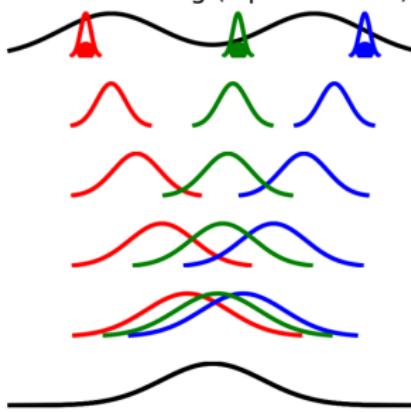
$$x_t = tx_1 + (1-t)x_0$$

$$z = (x_0, x_1)$$

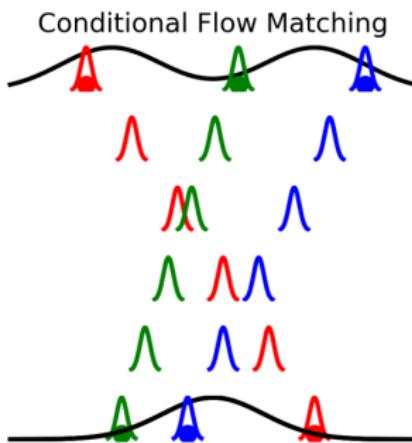
$$p_t(x|x_0, x_1) = \mathcal{N}(tx_1 + (1-t)x_0, \epsilon^2 I)$$

$$x_t = tx_1 + (1-t)x_0$$

Flow Matching (Lipman et al.)



Conditional Flow Matching



Linear Interpolation

$$p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1 + (1-t)\mathbf{x}_0, \epsilon^2 \mathbf{I}); \quad \textcolor{teal}{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0$$

Conditional Vector Field

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{f}(\mathbf{x}_t, \mathbf{x}_0, \mathbf{x}_1, t) = \boldsymbol{\mu}'_t(\mathbf{x}_0, \mathbf{x}_1) + \frac{\boldsymbol{\sigma}'_t(\mathbf{x}_0, \mathbf{x}_1)}{\boldsymbol{\sigma}_t(\mathbf{x}_0, \mathbf{x}_1)} \odot (\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1))$$
$$\mathbf{f}(\mathbf{x}_t, \mathbf{x}_0, \mathbf{x}_1, t) = \mathbf{x}_1 - \mathbf{x}_0$$

Conditional Flow Matching

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

- ▶ This yields the same procedure as for conical paths!
- ▶ Now, we do not require that $p_0(\mathbf{x})$ is necessarily $\mathcal{N}(0, \mathbf{I})$.

Conditional Flow Matching

- ▶ This conditioning allows us to transport any distribution $p_0(\mathbf{x})$ to any distribution $p_1(\mathbf{x})$.
- ▶ It's possible to apply this approach to paired tasks, e.g., style transfer.

Training Procedure

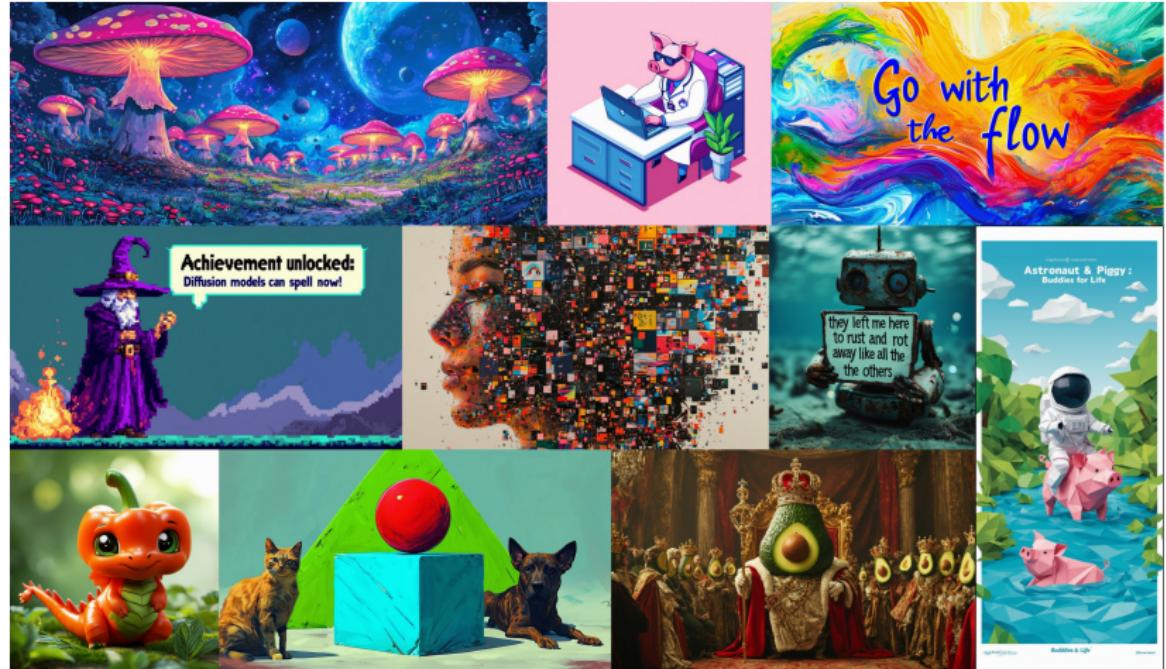
1. Sample $(\mathbf{x}_0, \mathbf{x}_1) \sim p(\mathbf{x}_0, \mathbf{x}_1)$.
2. Sample time $t \sim U[0, 1]$.
3. Compute the noisy image $\mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$.
4. Compute the loss $\mathcal{L} = \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_{\theta}(\mathbf{x}, t)\|^2$.

Sampling

1. Sample $\mathbf{x}_0 \sim p_0(\mathbf{x})$.
2. Solve the ODE to obtain \mathbf{x}_1 :

$$\mathbf{x}_1 = \text{ODESolve}_f(\mathbf{x}_0, \theta, t_0 = 0, t_1 = 1)$$

Stable Diffusion 3: Scalable Flow Matching



Outline

1. Conditional Flow Matching
2. Conical Gaussian Paths
3. Linear Interpolation
4. Link with Score-Based Models

Score-Based Generative Models through SDEs

Training

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

Variance Exploding SDE (NCSN)

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

Variance Preserving SDE (DDPM)

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

Flow matching uses reverse time direction:

$$\textbf{NCSN: } p(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1, \sigma_{1-t}^2 \cdot \mathbf{I})$$

$$\textbf{DDPM: } p(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\alpha_{1-t}\mathbf{x}_1, (1 - \alpha_{1-t}^2) \cdot \mathbf{I})$$

Flow Matching vs. Score-Based SDE Models

Flow Matching Probability Path

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

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

Variance Exploding SDE Probability Path

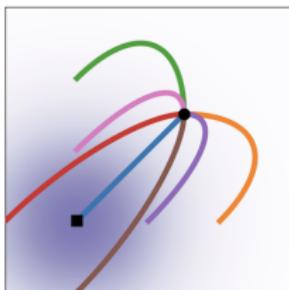
$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1, \sigma_{1-t}^2 \mathbf{I}) \quad \Rightarrow \quad \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = -\frac{\sigma'_{1-t}}{\sigma_{1-t}}(\mathbf{x} - \mathbf{x}_1)$$

Variance Preserving SDE Probability Path

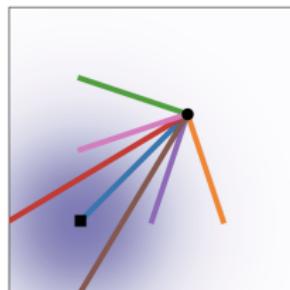
$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\alpha_{1-t}\mathbf{x}_1, (1-\alpha_{1-t}^2)\mathbf{I}) \Rightarrow \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\alpha'_{1-t}}{1-\alpha_{1-t}^2} \cdot (\alpha_{1-t}\mathbf{x} - \mathbf{x}_1)$$

Flow Matching vs. Score-Based SDE Models

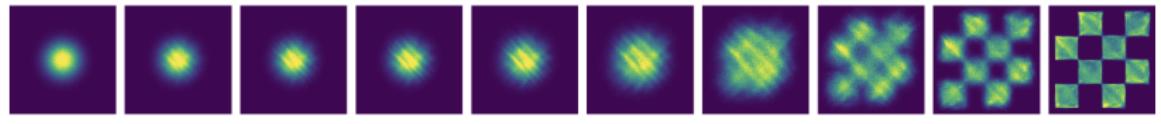
Trajectories



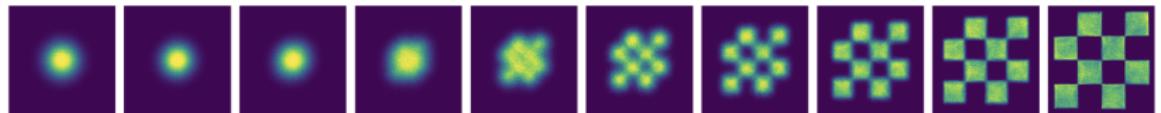
Diffusion



OT



Score matching ^{w/} Diffusion



Flow Matching ^{w/} OT

Summary

- ▶ Conditional flow matching makes the FM objective tractable.
- ▶ Conical Gaussian paths serve as an effective FM technique.
- ▶ Pair conditioning yields the same procedure, but is more general (suitable for unpaired tasks).
- ▶ Diffusion and score-based models are special cases of the flow matching approach, but use curved trajectories.