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

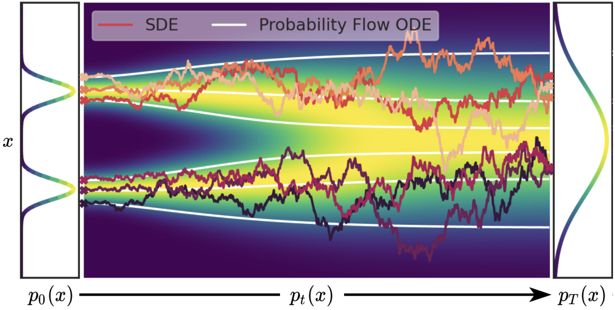
Recap of previous lecture

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

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



Song Y., et al. Score-Based Generative Modeling through Stochastic Differential Equations, 2020

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.

Song Y., et al. Score-Based Generative Modeling through Stochastic Differential Equations, 2020

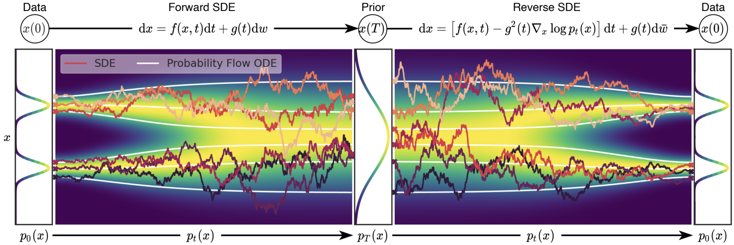
Recap of previous lecture

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

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

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

- 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.



Song Y., et al. Score-Based Generative Modeling through Stochastic Differential Equations, 2020

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.

Song Y., et al. Score-Based Generative Modeling through Stochastic Differential Equations, 2020

Outline

Flow Matching

Let consider ODE dynamic $\mathbf{x}_t = \mathbf{x}(t)$ in time interval $t \in [0, 1]$ with boundaries $\mathbf{x}_0 \sim p_0(\mathbf{x}) = p(\mathbf{x})$, $\mathbf{x}_1 \sim p_1(\mathbf{x}) = \pi(\mathbf{x})$. Here $p(\mathbf{x})$ is a base distribution ($\mathcal{N}(0, \mathbf{I})$) and $\pi(\mathbf{x})$ is a true data distribution.

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

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

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

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

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

Lipman Y., et al. Flow Matching for Generative Modeling, 2022

Flow Matching

There exists infinite number of possible $\mathbf{u}(\mathbf{x}, t)$ between $\pi(\mathbf{x})$ and $p(\mathbf{x})$.

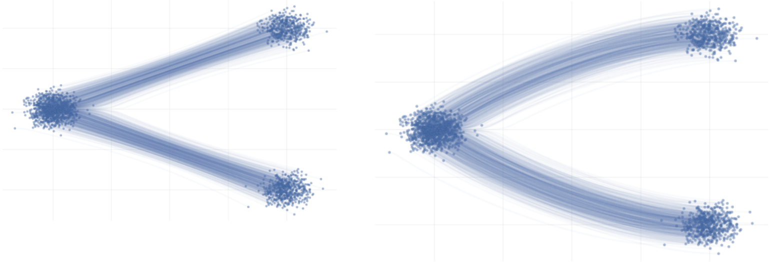


image credit: https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html

Flow Matching

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

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

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

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

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

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

Flow Matching

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

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

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = \frac{\partial}{\partial t} \int p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = \int \left(\frac{\partial p_t(\mathbf{x}|\mathbf{z})}{\partial t}\right) p(\mathbf{z}) d\mathbf{z} =$$

$$= \int \left(-\mathrm{div}\left(\mathbf{u}(\mathbf{x}, \mathbf{z}, t) p_t(\mathbf{x}|\mathbf{z})\right)\right) p(\mathbf{z}) d\mathbf{z} =$$

$$= -\mathrm{div}\left(\int \mathbf{u}(\mathbf{x}, \mathbf{z}, t) p_t(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z}\right) = -\mathrm{div}\left(\mathbf{u}(\mathbf{x}, t) p_t(\mathbf{x})\right)$$

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

Flow Matching

Flow Matching

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

Conditional Flow Matching

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

Theorem

If $\mathrm{supp}(p_t(\mathbf{x})) = \mathbb{R}^m$, 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.

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

Outline

Flow Matching

Conditional Flow Matching

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

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:

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

Gaussian conditional probability path

$$p_t(\mathbf{x}|\mathbf{x}_1)=\mathcal{N}\left(\boldsymbol{\mu}_t(\mathbf{x}_1),\boldsymbol{\sigma}_t^2(\mathbf{x}_1)\right)$$

$$\boldsymbol{\mu}_0(\mathbf{x}_1)=0, \quad \boldsymbol{\sigma}_0(\mathbf{x}_1)=1, \quad \boldsymbol{\mu}_1(\mathbf{x}_1)=\mathbf{x}_1, \quad \boldsymbol{\sigma}_1(\mathbf{x}_1)=0$$

Lipman Y., et al. Flow Matching for Generative Modeling, 2022

Flow Matching

Theorem

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

Proof

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}\left(\boldsymbol{\mu}_t(\mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_1)\right)$$

$$\mathbf{x} = \boldsymbol{\mu}_t(\mathbf{x}_1) + \boldsymbol{\sigma}_t(\mathbf{x}_1) \odot \boldsymbol{\epsilon}, \quad \Rightarrow \quad \boldsymbol{\epsilon} = \frac{1}{\boldsymbol{\sigma}_t(\mathbf{x}_1)} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_1))$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t)$$

$$\frac{d\mathbf{x}}{dt} = \boldsymbol{\mu}'_t(\mathbf{x}_1) + \boldsymbol{\sigma}'_t(\mathbf{x}_1) \odot \boldsymbol{\epsilon} = \boldsymbol{\mu}'_t(\mathbf{x}_1) + \frac{\boldsymbol{\sigma}'_t(\mathbf{x}_1)}{\boldsymbol{\sigma}_t(\mathbf{x}_1)} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_1))$$

Lipman Y., et al. Flow Matching for Generative Modeling, 2022

Flow Matching

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

$$\boldsymbol{\mu}_0(\mathbf{x}_1) = 0, \quad \boldsymbol{\sigma}_0(\mathbf{x}_1) = 1, \quad \boldsymbol{\mu}_1(\mathbf{x}_1) = \mathbf{x}_1, \quad \boldsymbol{\sigma}_1(\mathbf{x}_1) = 0$$

$$\boldsymbol{\mu}_t(\mathbf{x}_1) = t\mathbf{x}_1; \quad \boldsymbol{\sigma}_t(\mathbf{x}_1) = (1 - t); \quad \mathbf{u}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\mathbf{x}_1 - \mathbf{x}}{1 - t}$$

Lipman Y., et al. Flow Matching for Generative Modeling, 2022

Flow Matching

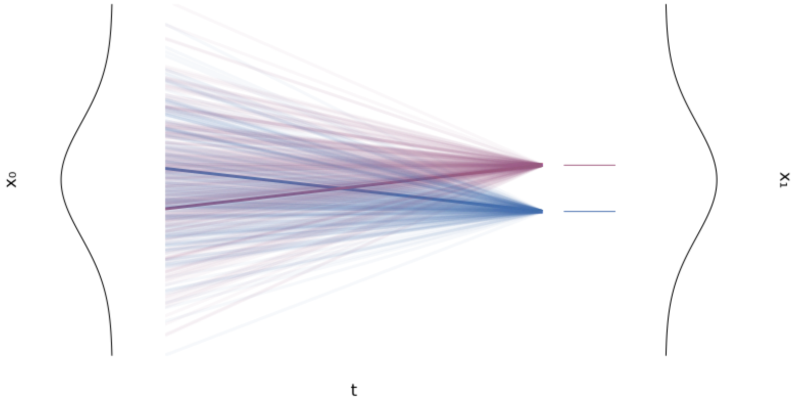
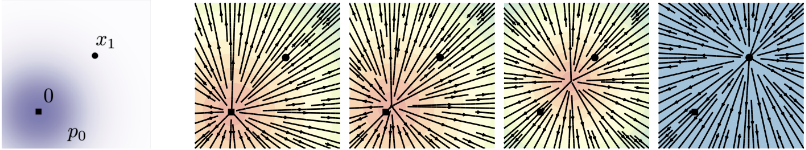
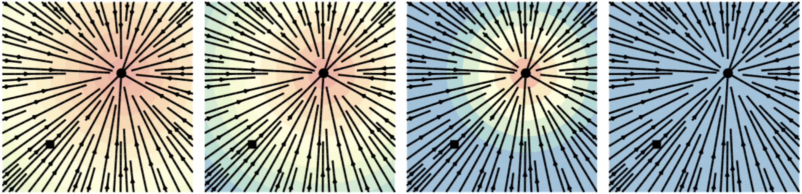


image credit: https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html

Flow Matching



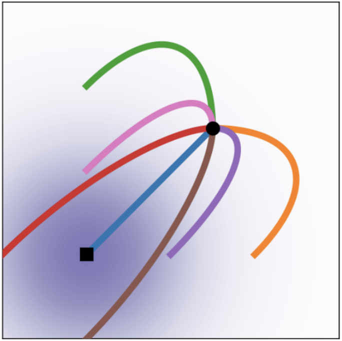
$t = 0.0$ $t = 1/3$ $t = 2/3$ $t = 1.0$



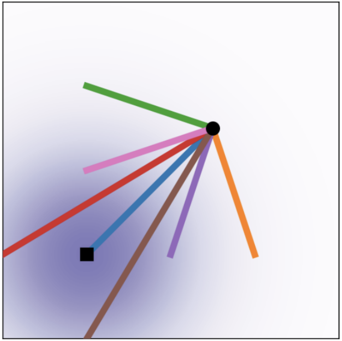
$t = 0.0$ $t = 1/3$ $t = 2/3$ $t = 1.0$

Lipman Y., et al. Flow Matching for Generative Modeling, 2022

Flow Matching



Diffusion



OT

Lipman Y., et al. *Flow Matching for Generative Modeling*, 2022

Outline

1. Flow Matching
   - Endpoint conditioning
   - Pair conditioning

Summary

-