Flow matching

Grigory Bartosh

April 21, 2023

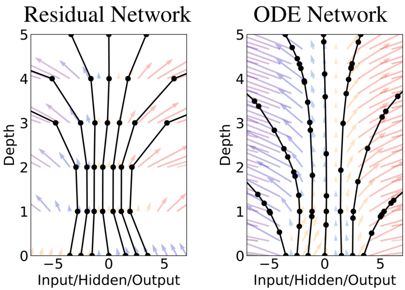
ODEs vs SDEs

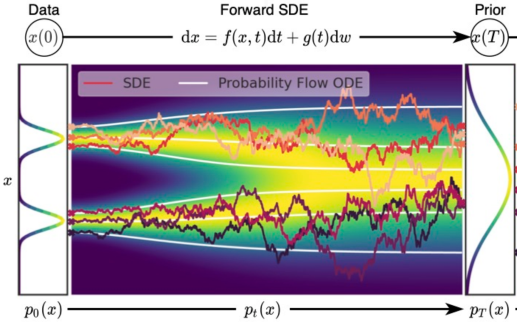
- Ordinary Differential Equation (ODE)

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

- Stochastic Differential Equation (SDE)

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





Reverse process

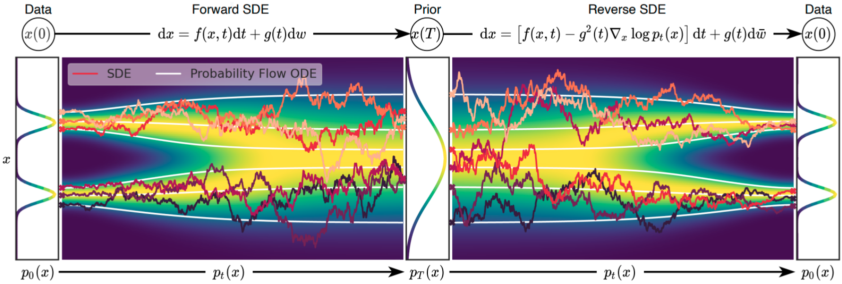
- SDE

$$\mathrm{d}x = f(x, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \tag{1}$$

- Reverse SDE (derived from the Fokker-Planck equation)

$$\mathrm{d}x = [f(x, t) - g^2(t)\nabla_x \log q_t(x)]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \tag{2}$$

- Illustration



Training, score matching

- Score matching (Fisher divergence)

$$\mathcal{L} = \mathbb{E}_{\mathcal{U}(t)}\mathbb{E}_{q(x_t)}\left[\lambda(t)\|\mathbf{s}_\theta(x_t, t) - \nabla_{x_t} \log q(x_t)\|_2^2\right] \tag{3}$$

- Conditional score matching

$$\mathcal{L} = \mathbb{E}_{\mathcal{U}(t)}\mathbb{E}_{q(x_0)}\mathbb{E}_{q(x_t|x_0)}\left[\lambda(t)\|\mathbf{s}_\theta(x_t, t) - \nabla_{x_t} \log q(x_t|x_0)\|_2^2\right] \tag{4}$$

$$+ \, const \tag{5}$$

- Fisher divergence is variational upper bound of NLL

$$\mathbb{E}_{q(x_0)}[-\log p_\theta(x_0)] \leq \mathcal{L} \tag{6}$$

Continues Normalizing Flow

- Model

$$\mathrm{d}x = f_\theta(x, t)\mathrm{d}t \tag{7}$$

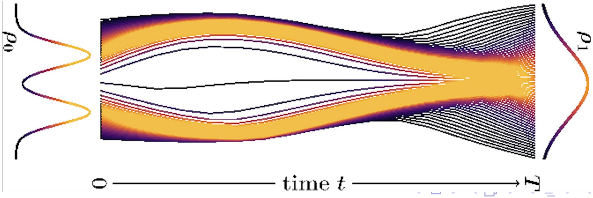
- Objective

$$\mathcal{L} = \mathbb{E}_{q(x_0)}[-\log p_\theta(x_0)] \tag{8}$$

$$= \mathbb{E}_{q(x_0)}\left[-\log p_\theta(x_1) + \int_0^1 \frac{\partial \log p(x_t)}{\partial t}\mathrm{d}t\right] \tag{9}$$

$$= \mathbb{E}_{q(x_0)}\left[-\log p_\theta(x_1) - \int_0^1 Tr\left(\frac{\partial f(x_t, t)}{\partial x_t}\right)\mathrm{d}t\right] \tag{10}$$

- Illustration



Flow Matching

Paper[1]:

Flow Matching for Generative Modeling

Yaron Lipman[1,2] **Ricky T. Q. Chen**[1] **Heli Ben-Hamu**[2] **Maximilian Nickel**[1] **Matt Le**[1]
[1]Meta AI [2]Weizmann Institute of Science

Abstract

We introduce a new paradigm for generative modeling built on Continuous Normalizing Flows (CNFs), allowing us to train CNFs at unprecedented scale. Specifically, we present the notion of Flow Matching (FM), a simulation-free approach for training CNFs based on regressing vector fields of fixed conditional probability paths. Flow Matching is compatible with a general family of Gaussian probability paths for transforming between noise and data samples—which subsumes existing diffusion paths as specific instances. Interestingly, we find that employing FM with diffusion paths results in a more robust and stable alternative for training diffusion models. Furthermore, Flow Matching opens the door to training CNFs with other, non-diffusion probability paths. An instance of particular interest is using Optimal Transport (OT) displacement interpolation to define the conditional probability paths. These paths are more efficient than diffusion paths, provide faster training and sampling, and result in better generalization. Training CNFs using Flow Matching on ImageNet leads to state-of-the-art performance in terms of both likelihood and sample quality, and allows fast and reliable sample generation using off-the-shelf numerical ODE solvers.

---

[1] Yaron Lipman et al. "Flow matching for generative modeling". In: *arXiv preprint arXiv:2210.02747* (2022).

Flow Matching, Idea – I

- Terminal distributions of the forward process

$$q(x_0) \text{ – data distribution}, \quad q(x_1) \approx \mathcal{N}(x_1; 0, I) \tag{11}$$

- Forward process

$$\mathrm{d}x_t = u(x_t)\mathrm{d}t \tag{12}$$

- Reverse process

$$\mathrm{d}x_t = v_\theta(x_t)\mathrm{d}t \tag{13}$$

- Flow matching objective

$$\mathcal{L} = \mathbb{E}_{\mathcal{U}(t)}\mathbb{E}_{q(x_t)}\left[\|v_\theta(x_t) - u(x_t)\|_2^2\right] \tag{14}$$

Flow Matching, Idea – II

- Terminal distributions of the forward process

$$q(x_0) \text{ – data distribution}, \quad q(x_1|x_0) \approx \mathcal{N}(x_1; 0, I) \tag{15}$$

- Forward conditional process

$$\mathrm{d}x_t = u(x_t|x_0)\mathrm{d}t \tag{16}$$

- Reverse process

$$\mathrm{d}x_t = v_\theta(x_t)\mathrm{d}t \tag{17}$$

- $q(x_t)$ corresponds to $u(x_t)$, where

$$u(x_t) = \int u(x_t|x_0)q(x_0|x_t)\mathrm{d}t \tag{18}$$

- Conditional flow matching objective

$$\mathcal{L} = \mathbb{E}_{\mathcal{U}(t)}\mathbb{E}_{q(x_t)}\left[\|v_\theta(x_t) - u(x_t)\|_2^2\right] \tag{19}$$

$$= \mathbb{E}_{\mathcal{U}(t)}\mathbb{E}_{q(x_0)}\mathbb{E}_{q(x_t|x_0)}\left[\|v_\theta(x_t) - u(x_t|x_0)\|_2^2\right] + const \tag{20}$$

Flow Matching, Gaussian distribution

- Forward marginal

$$q(x_t|x_0) = \mathcal{N}(x_t; \mu_t(x_0), \sigma_t^2(x_0)I) \tag{21}$$

- Reparameterization

$$x_t(\varepsilon) = \mu_t(x_0) + \sigma_t(x_0)\varepsilon, \quad \varepsilon(x_t) = \frac{1}{\sigma_t(x_0)}(x_t - \mu_t(x_0)) \tag{22}$$

- Forward conditional process

$$u(x_t|x_0) = \frac{d}{\mathrm{d}t}(\mu_t(x_0) + \sigma_t(x_0)\varepsilon) \tag{23}$$

Flow Matching, non-Gaussian distribution

- Forward marginal

$$q(x_t|x_0) = ??? \tag{24}$$

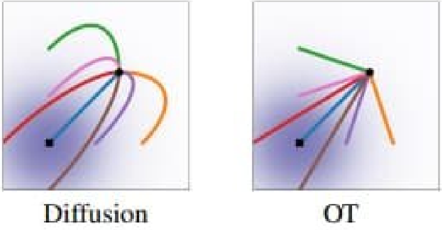
- Reparameterization

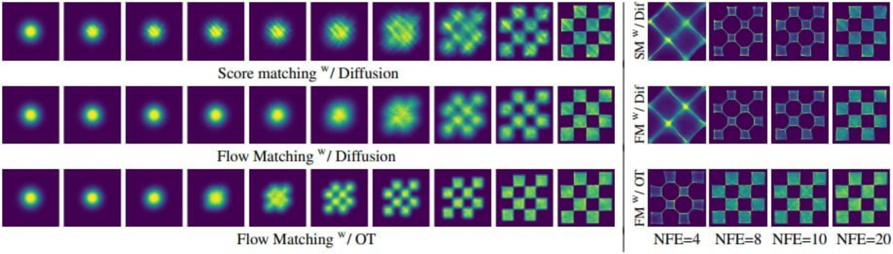
$$x_t(\varepsilon) = T_t(\varepsilon|x_0), \quad \varepsilon(x_t) = T_t^{-1}(x_t|x_0) \tag{25}$$

- Forward conditional process

$$u(x_t|x_0) = \frac{d}{\mathrm{d}t} T_t(\varepsilon|x_0) \tag{26}$$

Flow Matching, optimal transport





Flow Matching, comparison

Model	CIFAR-10		ImageNet 32×32		ImageNet 64×64	
	NLL↓	FID↓	NLL↓	FID↓	NLL↓	FID↓
Normalizing Flow						
FFJORD (Grathwohl et al., 2018)	3.40					
Glow (Kingma & Dhariwal, 2018)	3.35		4.09		3.81	
Residual Flow (Chen et al., 2019)	3.28		4.01		3.76	
Flow++ (Ho et al., 2019)	3.09		3.86		3.69	
Variational Autoencoder						
NVAE (Vahdat & Kautz, 2020)	2.91		3.92			
Very Deep VAE (Child, 2020)	2.87		3.80		3.52	
Diffusion Model						
DDPM (Ho et al., 2020)	3.75	3.17				
VDM (Kingma et al., 2021)	**2.65**	7.41	3.72		3.40	
Score SDE (Song et al., 2020b)	2.99	**2.92**				
Soft Truncation (Kim et al., 2022)	2.88	3.45	3.85	8.42		
ScoreFlow (Song et al., 2021)	2.81	5.40	3.76	10.18		
Ablation						
Score Matching w/ Diffusion path	3.16	21.96	3.57	22.38	3.40	19.61
Ours						
Flow Matching w/ Diffusion path	3.10	10.31	3.56	8.02	3.33	16.06
Flow Matching w/ OT path	3.00	6.96	**3.53**	**5.25**	**3.31**	**14.00**

Connection between SDEs and ODEs – I

- SDE

$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \tag{27}$$

- Reverse SDE

$$\mathrm{d}x = [f(x,t) - g^2(t)\nabla_x \log q_t(x)]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \tag{28}$$

- ODE

$$\mathrm{d}x = \left[f(x,t) - \frac{g^2(t)}{2}\nabla_x \log q_t(x)\right]\mathrm{d}t \tag{29}$$

Connection between SDEs and ODEs – II

- ODE

$$
\begin{aligned}
\mathrm{d}x &= \psi(x,t)\mathrm{d}t && (30)\\
&= \left[\underbrace{\psi(x,t) + \frac{g^2(t)}{2}\nabla_x \log q_t(x)}_{f(x,t)} - \frac{g^2(t)}{2}\nabla_x \log q_t(x)\right]\mathrm{d}t && (31)
\end{aligned}
$$

- SDE

$$
\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \qquad (32)
$$

- Reverse SDE

$$
\mathrm{d}x = [f(x,t) - g^2(t)\nabla_x \log q_t(x)]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \qquad (33)
$$

Further Reading

- Action Matching: Learning Stochastic Dynamics from Samples[2].
- Riemannian Flow Matching on General Geometries[3].
- Conditional Flow Matching: Simulation-Free Dynamic Optimal Transport[4].

---

[2] Kirill Neklyudov, Daniel Severo, and Alireza Makhzani. "Action Matching: A Variational Method for Learning Stochastic Dynamics from Samples". In: *arXiv preprint arXiv:2210.06662* (2022).

[3] Ricky TQ Chen and Yaron Lipman. "Riemannian flow matching on general geometries". In: *arXiv preprint arXiv:2302.03660* (2023).

[4] Alexander Tong et al. "Conditional Flow Matching: Simulation-Free Dynamic Optimal Transport". In: *arXiv preprint arXiv:2302.00482* (2023).