Switched Flow Matching: Eliminating Singularities via Switching ODEs

Qunxi Zhu [1] Wei Lin [1]

[1]Fudan University

Summary

The major contributions of this study are multi-folded, including:

- **Development of SFM**: We establish Switched Flow Matching (SFM), a versatile continuous-time generative model that eliminates singularities encountered in the Flow Matching (FM) [3] via switching the candidate ODEs, and allows the intersection of probability paths from different ODEs.
- **Theoretical insights**: Through rigorous analysis, we demonstrate that FM struggles with transporting between simple distributions due to the existence and uniqueness of initial value problems of ODEs while such limitation can be effectively addressed by SFM, offering a more efficient solution.
- **Integration with advanced techniques**: SFM can seamlessly integrate with the existing advanced techniques, for example, minibatch optimal transport, to further enhance the straightness of the flow, facilitating a more efficient sampling process.
- **Empirical validation**: We validate the effectiveness of the newly proposed SFM through extensive experiments on both synthetic and real-world datasets, achieving competitive or even better performance compared to existing methods, such as FM.

Background

Continuous Normalizing Flow. Continuous Normalizing Flow is a continuous-time generative model that samples data points from the source distribution $\boldsymbol{x}_0 \sim q_0(\boldsymbol{x}_0)$ and then transforms them into diifferent ones by solving the initial value problem of the neural ODE (NODE) [2]:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{v}_t(\boldsymbol{x};\boldsymbol{\theta}), \quad t \in [0,1], \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{1}$$

where $\boldsymbol{v}_t(\boldsymbol{x};\boldsymbol{\theta})$ is a parameterized neural network with the trainable weights $\boldsymbol{\theta}$.

Static and Dynamic Optimal Transport. Consider the following squared 2-Wasserstein distance:

$$W(q_0,q_1)^2 = \inf_{\pi\in\Pi(q_0,q_1)} \int \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|^2 \mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_1), \tag{2}$$

which has the equivalent dynamic form, known as the Benamou-Brenier formula [1]:

$$W(q_0,q_1)^2 = \inf_{\boldsymbol{u}_t\in U(q_0,q_1)} \int_0^1 \int p_t(\boldsymbol{x}) \|\boldsymbol{u}_t(\boldsymbol{x})\|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t. \tag{3}$$

Motivation

Proposition 1. (Infinite number of singular points) *Suppose the source and target distributions q_0 and q_1 are defined on $\mathbb{R}^2$ with q_0 being $\mathcal{H}^1$ restricted to $\{0\}\times[-1,1]$, and q_1 being $(1/2)\mathcal{H}^1$ restricted to $\{-1,1\}\times[-1,1]$, respectively. Consider the (dynamic) optimal transport problem as defined in Eq. (2) (or Eq. (3)). If the NODE (1) exactly solves the problem, then all the points $\boldsymbol{x}(0) = (0,a), a \in [-1,1]$ are singular points as shown in Fig. 1(a).*

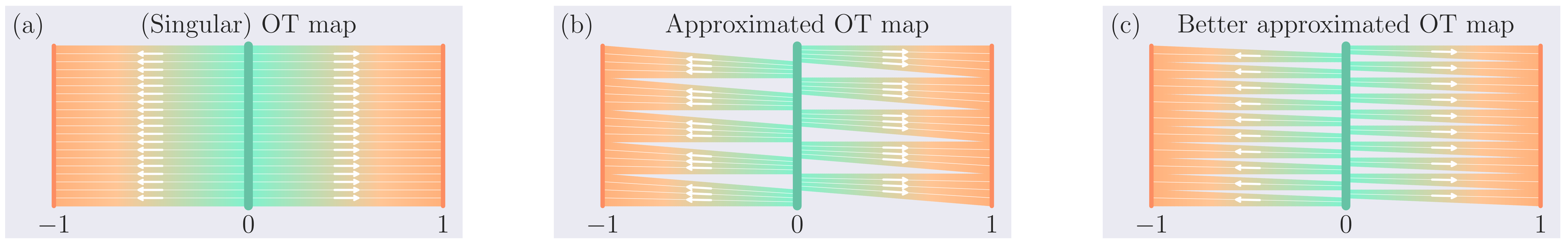

Figure 1. Illustration of the (singular) OT map (a) and the (better) approximated OT maps (b) & (c) on the example in Proposition 1.

Switched Flow Matching

Proposition 2. (Switching ODEs) *The marginal probability path $p_t(\boldsymbol{x})$ can be effectively sampled by switching ODEs in the following three steps:*

1. ***Sampling an ODE.** Sampling a switching signal $\boldsymbol{s}$ from the distribution $q^\circ(\boldsymbol{s})$, resulting in the specified ODE $\boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{s})$;*
2. ***Sampling an initial state.** Sampling an initial state $\boldsymbol{x}_0$ (resp., backward one $\boldsymbol{x}_1$) from the conditional distribution $q_0(\boldsymbol{x}_0|\boldsymbol{s})$ (resp., $q_1(\boldsymbol{x}_1|\boldsymbol{s})$);*
3. ***Solving the IVP.** Generating the corresponding conditional probability path $p_t(\boldsymbol{x}|\boldsymbol{s})$ by the vector field $\boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{s})$ from the initial state $\boldsymbol{x}_0$ (resp., $\boldsymbol{x}_1$).*

Table 1. Properties for the ODE-based generative models, including the FM, CFM, and our proposed SFM. Particularly, the SFM can not only handle general source distributions, and optimal transport flows (OT-SFM), but also employ multiple ODEs to eliminate the singularity, allowing the intersection of trajectories from different ODEs, and owning the relatively good regularity.

ODE model	General source	OT	Multiple ODEs	Intersection	Regularity
FM	✗	✗	✗	✗	✗
I-CFM	✓	✗	✗	✗	✗
OT-CFM	✓	✓	✗	✗	✗
I-SFM	✓	✗	✓	✓	✓
OT-SFM	✓	✓	✓	✓	✓

Synthetic Datasets

- Figure 2 shows the proposed I-SFM and OT-SFM on transporting an 1-d Gaussian mixture (2-modes) to another. It is observed that an appropriate switching rule can eliminate the singularity raised from the heterogeneities of source and target distributions.

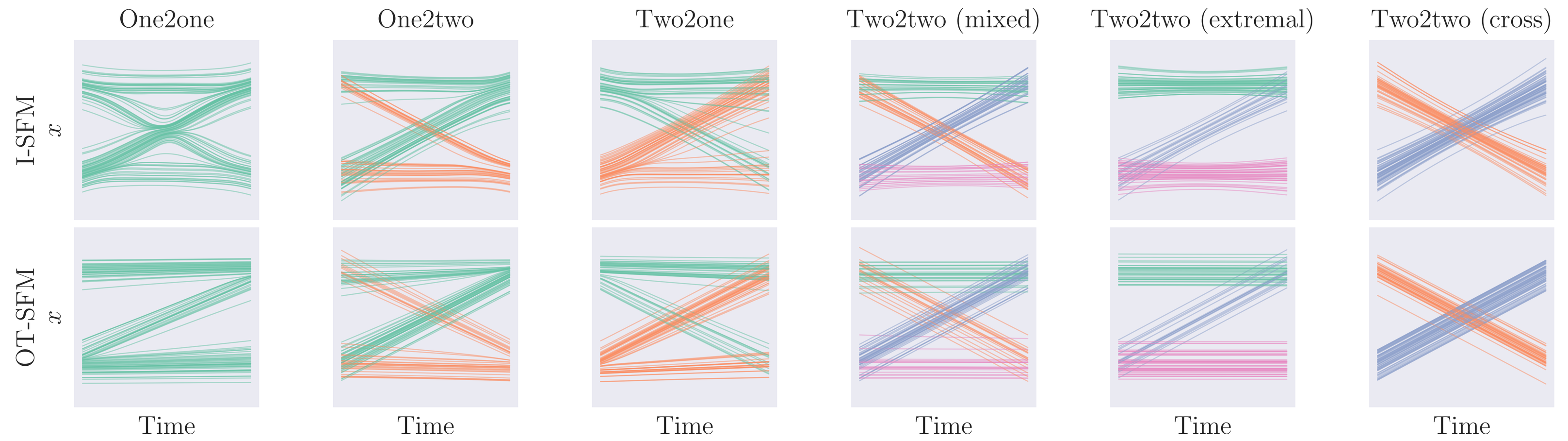

Figure 2. Trajectories of the I-SFM and the OT-SFM on 2-d Gaussian mixtures under different couplings. Particularly, in the first column ("one2one" coupling), the I-SFM and the OT-SFM are the I-CFM and OT-CFM, respectively.

- Figure 3 shows the learned flows of the I-SFM and the OT-SFM on the example of the infinite number of singular points under the optimal coupling in Proposition 1.

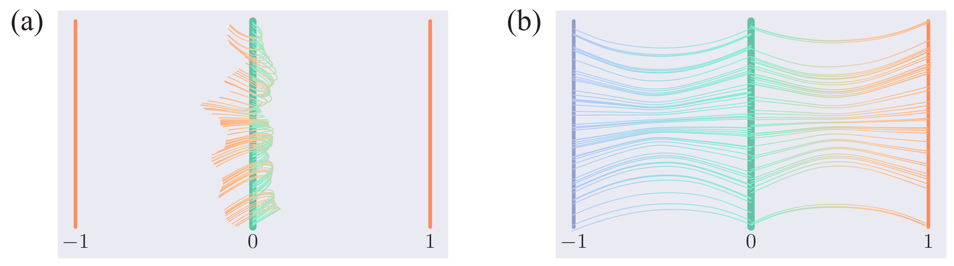

Figure 3. The learned flows of the I-CFM (a) and the I-SFM (one2two) (b) on the example in Proposition 1.

CIFAR-10 dataset

Table 2 shows the image generation results of our SFM variants on the CIFAR-10 dataset. In contrast with the existing generative models, we, here, consider a general source distribution, a Gaussian mixture with two modes (see Figure 4), instead of a standard Gaussian distribution.

Table 2. FID results of CFM and SFM on the CIFAR-10 dataset.

NFE	6	8	10	20	40	Adaptive
I-CFM (I-SFM, one2one)	144.52	130.49	122.44	106.11	99.19	94.55
OT-CFM (OT-SFM, one2one)	176.80	111.09	76.41	26.15	10.90	4.91
I-SFM (one2ten)	**109.24**	98.47	93.48	83.41	78.33	75.06
OT-SFM (one2ten)	122.74	104.19	93.04	73.47	63.94	59.72
I-SFM (two2one)	177.99	115.05	78.46	23.91	9.18	5.21
OT-SFM (two2one)	185.44	121.21	84.32	28.18	11.11	5.64
I-SFM (two2ten, mixed)	132.41	75.83	**49.53**	**15.60**	**6.98**	4.27
OT-SFM (two2ten, mixed)	133.27	76.31	49.69	15.50	7.24	4.39
I-SFM (two2ten, extremal)	128.55	**75.11**	50.12	17.14	8.39	**4.22**
OT-SFM (two2ten, extremal)	149.50	88.33	58.25	18.59	8.86	4.40

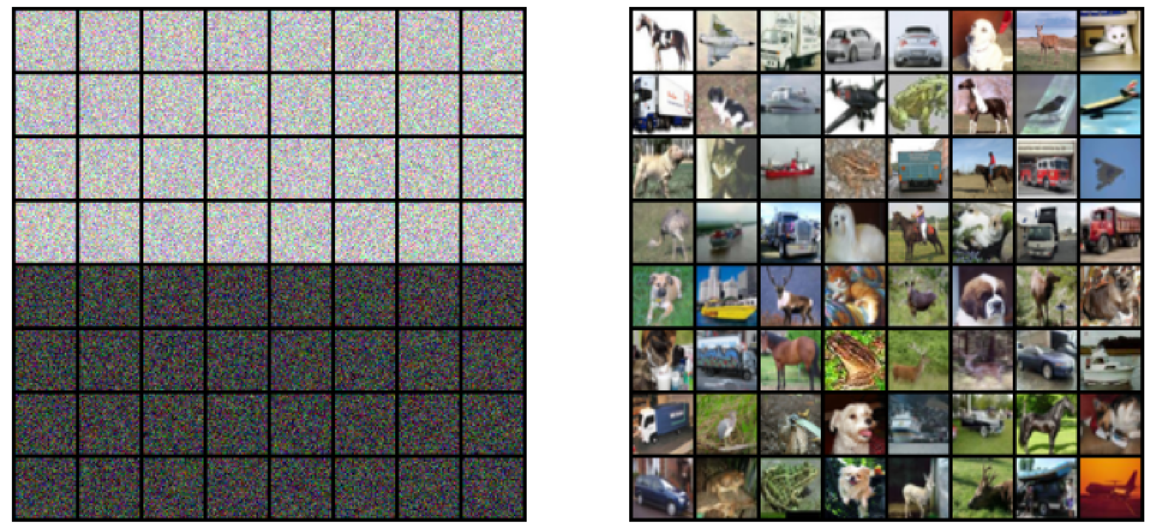

Figure 4. True samples from the source distribution (left, Gaussian mixture) and the target distribution (right, CIFAR-10 dataset).

References

[1] Jean-David Benamou and Yann Brenier.
A numerical method for the optimal time-continuous mass transport problem and related problems.
Contemporary Mathematics, 226:1–12, 1999.

[2] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud.
Neural ordinary differential equations.
Advances in Neural Information Processing Systems, 31, 2018.

[3] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le.
Flow matching for generative modeling.
Arxiv Preprint Arxiv:2210.02747, 2022.

Paper

Scan Me