Numer. Math. Theor. Meth. Appl. Vol. **18**, No. 2, pp. 325-352
doi: 10.4208/nmtma.OA-2024-0114 May 2025

Convergence Analysis of OT-Flow for Sample Generation

Yang Jing[1,*] and Lei Li[1,2,3]

[1] *School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, P.R.China*
[2] *Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, P.R.China*
[3] *Shanghai Artificial Intelligence Laboratory, P.R.China*

Received 27 September 2024; Accepted (in revised version) 5 January 2025

Abstract. Deep generative models aim to learn the underlying distribution of data and generate new ones. Despite the diversity of generative models and their high-quality generation performance in practice, most of them lack rigorous theoretical convergence proofs. In this work, we aim to establish some convergence results for OT-Flow, one of the deep generative models. First, by reformulating the framework of OT-Flow model, we establish the Γ-convergence of the formulation of OT-Flow to the corresponding optimal transport (OT) problem as the regularization term parameter α goes to infinity. Second, since the loss function will be approximated by Monte Carlo method in training, we established the convergence between the discrete loss function and the continuous one when the sample number N goes to infinity as well. Meanwhile, the approximation capability of the neural network provides an upper bound for the discrete loss function of the minimizers. The proofs in both aspects provide convincing assurances for the stability of OT-Flow.

AMS subject classifications: 49Q22, 68T07

Key words: Generative models, continuous normalizing flows, OT-Flow, Benamou-Brenier functional, Γ-convergence.

1. Introduction

Deep generative models [15, 21, 23, 33] are increasingly being adopted as the preferred methodology across various tasks due to their impressive performance, including solving inverse problems [7], image generation [10], text-to-image [32] and video generation [25]. The widely-used frameworks include diffusion probabilistic models

*Corresponding author. *Email addresses:* `sharkjingyang@sjtu.edu.cn` (Y. Jing), `leili2010@sjtu.edu.cn` (L. Li)

(DPMs) [3, 17, 36], continuous normalizing flows (CNFs) [8, 16], variational autoencoders (VAEs) [21, 23] and generative adversarial networks (GANs) [2, 15]. Among above four popular frameworks, CNFs are characterized by continuous-time ordinary differential equations (ODEs), and DPMs utilize stochastic differential equations (SDEs) as their backbone. Through DPMs and CNFs, samples evolve from data points to Gaussian distribution in the forward process and gradually remove noise to generate samples in the backward process. In comparison with GANs and VAEs, samples of DPMs and CNFs are generated in smoother ways, not only achieving superior sample quality but also enabling exact likelihood computation. Despite the diversity of generative models and their outstanding performance in downstream tasks, the mathematical principles behind the models and rigorous convergence proofs are developed far behind the rapid iteration of the models. In this paper, our focus lies in establishing convergence results for OT-Flow, which stands as one of the practical CNFs. Such convergence analysis ensures stability during the training and aids in comprehending the underlying mechanisms of the model.

The continuous normalizing flows (CNFs) are a class of sample generative models based on particle transportation purely. The CNFs aim to build continuous and invertible mappings between an arbitrary distribution ρ_0 and standard normal distribution ρ_1 by setting the velocity field as an output of neural network. In particular, for a given time T, one is trying to obtain a mapping $z : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$, which defines a continuous evolution $x \mapsto z(x, t)$ of every $x \in \mathbb{R}^d$. Then the density $\rho(z(x, t), t)$ satisfies

$$\log \rho_0(x) = \log \rho\big(z(x,t),t\big) + \log|\det \nabla z(x,t)| \quad \text{for all} \quad x \in \mathbb{R}^d. \tag{1.1}$$

Especially at time T we have

$$\log \rho_0(x) = \log \rho_1\big(z(x,T),T\big) + \log|\det \nabla z(x,T)|.$$

Define

$$\ell(x,t) := \log|\det \nabla z(x,t)|,$$

then $z(x, t)$ and $\ell(x, t)$ satisfy the following ODE system:

$$\partial_t \begin{bmatrix} z(x,t) \\ \ell(x,t) \end{bmatrix} = \begin{bmatrix} v(z(x,t),t;\boldsymbol{\theta}) \\ \operatorname{tr}(\nabla v(z(x,t),t;\boldsymbol{\theta})) \end{bmatrix}, \quad \begin{bmatrix} z(x,0) \\ \ell(x,0) \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}. \tag{1.2}$$

To train the dynamics, CNFs minimize the expected negative log-likelihood given by the right-hand-side in (1.1), or equivalently the KL divergence between target distribution and final distribution under the constraint (1.2) [16, 31, 33]

$$J = \mathbb{KL}\big[\rho(\boldsymbol{z}(\boldsymbol{x},T))\|\rho_1(\boldsymbol{z}(\boldsymbol{x},T))\big]. \tag{1.3}$$

For convenience we solve (1.2) together to obtain the change of ρ, which will lead to a more efficient estimation of density.

From the ODE system (1.2), we can see that once the velocity field is learned, one can track the evolution of density and invert the transport map by running the ODE