修士論文 要旨

A Study on Regularizations for Stabilizing ODE-Net Based on Mean-Field Optimal Control Formulation

（平均場最適制御問題の枠組に基づく ODE-Net 安定化のための正則化に関する研究）

数理情報学専攻 48206203　磯部 伸
指導教員　松尾 宇泰 教授

1　導入

深層ニューラルネットワーク（Deep Neural Network: DNN）が世界を席捲して数年が経つ一方で，DNN は入力の摂動，特に，敵対的攻撃に対して脆弱であることが明らかになっている [GSS15]．ゆえに，摂動に対して頑健な DNN を構成することが必要とされている．このような需要の中，深層構造を常微分方程式（Ordinary Differential Equation: ODE）の離散化で置き換えた，ODE-Net（または Neural ODE）という DNN が提案された [Che+18; E17]．ODE-Net は摂動に対して頑健であることが利点として挙げられているものの，実験的事実が報告されているに過ぎず [Car+19]，その理論的根拠は殆どない．

そこで，本研究では，ODE-Net を最適制御問題として定式化し，その定性挙動を解析する（第 2 節）．その解析を通じて得られた知見から，安定になると期待できる ODE-Net を提案し，数学的正当化を行う（第 3 節）．

2　ODE-Net の指数的増大性

2.1　問題設定，数値例

まず，教師あり学習問題としての ODE-Net を，平均場最適制御問題 [EHL18] として定式化する：$T > 0$ は“層”の深さ（終端時刻）を表し，$\Theta = \mathbb{R}^m$ を各“層”（時間）の訓練パラメータが取り得る値の集合とする．X, Y を訓練データ，つまり，それぞれ，d 次元の入力と，ある集合 $\mathcal{Y}$ の中の値をとる正解ラベルを表す確率変数とし，$\mu_0 \in \mathcal{P}_c(\mathbb{R}^d \times \mathcal{Y})$ を，確率変数の組 (X, Y) が従う，コンパクト台を持つような確率分布とする．また，写像 $v\colon \mathbb{R}^d \times \Theta \to \mathbb{R}^d, \ell\colon \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}, L\colon \mathbb{R}^d \times \Theta \to \mathbb{R}$ は，それぞれ（浅い）ニューラルネットワーク，損失関数，正則化項を表すとする．この設定の下で，「与えられたデータ μ_0 に対して，次の汎関数

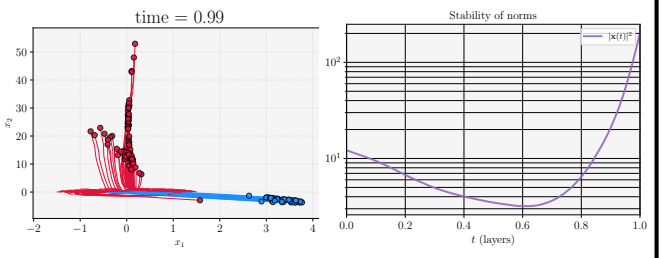
$$J(\mu, \theta) = \int_{\mathbb{R}^d \times \mathcal{Y}} \ell \, \mathrm{d}\mu_T + \int_0^T \int_{\mathbb{R}^d} L(x, \theta_t) \, \mathrm{d}\mu_t \, \mathrm{d}t \quad (1)$$

を，制約条件

$$\partial_t \mu_t + \nabla_x \cdot (v(\bullet, \theta_t)\, \mu_t) = 0, \quad \mu_t|_{t=0} = \mu_0 \quad (2)$$

のもとで最小化するパラメータ $\theta \in L^2(0, T; \Theta)$ と $\mathcal{P}_c(\mathbb{R}^d \times \mathcal{Y})$ 上の曲線 $\mu \in C([0, T], \mathcal{P}_c(\mathbb{R}^d \times \mathcal{Y}))$ を見出せ．」という問題を，「ODE-Net の学習問題」と呼ぶことにする．

通常，機械学習では過学習を防止する等の理由の為，正則化項として $L_1(x, \theta) := \lambda|\theta|^2/2$ $(\lambda > 0)$ 等を採用する．また，最適制御の文脈でも，状態量のノルムを制御する為，$L_2(x, \theta) := \left(\lambda_\theta|\theta|^2 + \lambda_x|x|^2\right)/2$ $(\lambda_\theta, \lambda_x > 0)$ のような評価関数を用いる．しかし，実際に $L = L_1, L_2$ として ODE-Net の学習問題を数値的に解くと，図 1 のような，終端値付近で分散が急増大する解が得られる．



(a) μ_t の軌跡（アニメーション）　(b) $W_2(\mu_t, \delta_0)^2$ の履歴

図 1．二値分類問題を，$L = L_2$ のときの ODE-Net に学習させた結果．終端時刻付近で急激な変化が見られる．訓練データはうまく分離している．

このような指数的増大は，入力の摂動を増幅させ得る為，汎化性能低下の一因にもなっていると考えられる．

2.2　Turnpike 理論による指数的増大性の説明

図 1 のような挙動の原因は，汎関数 (1) の形にある．式 (1) の中で，μ_t $(0 \lneq t \lneq T)$ が直接現れる項は，正則化項 L に関わる第二項のみである．ゆえに，L がなるべく小さくなる定常解付近に，殆どの時間留まるような曲線 μ_t^* が，ODE-Net の学習問題の最適解となることが直感的には予想される．このような説明は Turnpike 理論 [Est+20; Zas05] でよく見られる．これらの結果を利用して，$L = L_2$ の場合に次の主張が得られる．

定理 2.1.（ODE-Net は指数的に増大し得る．） 以下の二つを仮定する．

可制御性 ≈ 普遍性　任意の二点 $\mu_1, \mu_2 \in \mathcal{P}_c(\mathbb{R}^d \times \mathcal{Y})$ を，充分小さい時間で結ぶようなパラメー

タ θ が存在する．

コスト評価　ある定数 T_0, r が存在して，任意の $\mu_0 \in B_r(\delta_0)$ について，定数 $C(T_0)$ が存在して，$\inf_{\theta \text{ s.t. } \mu(0)=\mu_0, \mu(T_0)=\delta_0} \|\theta\|_{L^2(0,T_0)} \leq C(T_0) W_2(\mu_0, \delta_0)$ を満たす．

このとき，ある定数 $T^*, C, k > 0$ が存在して，任意の $T \geq T^*$ と，ODE-Net の学習問題の，任意の最小化元 μ, θ について，

$$\int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\mu_t \leq \frac{C}{\lambda_x}\left(\mathrm{e}^{-k\lambda_x t} + \mathrm{e}^{-k\lambda_x (T-t)}\right) \quad (3)$$

が，任意の $t \in [0, T]$ で成り立つ．

式 (3) 右辺第二項をみると，ODE-Net の二次モーメントに，指数的に増大する成分が含まれ得ることが分かる．

2.3　保存則による ODE-Net の不安定性の説明

より詳細な解析の為，ニューラルネットワーク v を，“最終層のパラメータのみ学習する”ように緩和した上で，Pontryagin の最大値原理を用いる．すると，ODE-Net は学習の結果，次の保存則に従うようになる．

定理 2.2.　(ODE-Net は保存量を持つ)
ニューラルネットワークが $v(x, \theta) = \theta f(x)$ $(\theta \in \mathbb{R}^{d\times p}, f\colon \mathbb{R}^d \to \mathbb{R}^p)$ の表式を持つとき，ODE-Net の学習問題に解が存在するのであれば，その解は

$$\begin{cases} \dfrac{\lambda}{2}|\theta|^2 & (\text{when } L = L_1) \\ \dfrac{\lambda_\theta}{2}|\theta|^2 - \dfrac{\lambda_x}{2}\displaystyle\int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\mu(x) & (\text{when } L = L_2) \end{cases}$$

を保存量にもつ．

特に $L = L_2$ の場合を考えれば，$|x|$ が大きくなるほど，ベクトル場のノルム $|v(x,\theta)| = |\theta f(x)|$ も大きくなり，急変化が助長されてしまうことが分かる．

3　運動論的正則化による ODE-Net の安定化

定理 2.2 では，保存則の観点から ODE-Net が不安定成分を持つことをみた．この観察を活かし，安定性の意味で “よい” 保存量を保存するように，正則化項 L を設計することを考え，次の正則化項を導入する．

定義 3.1.　(運動論的正則化)　$\lambda, \epsilon > 0$ として

$$L_{\text{kinetic}}(x, \theta) := \frac{\lambda}{2}|v(x,\theta)|^2 + \frac{\epsilon}{2}|\theta|^2 \quad (4)$$

とおく．

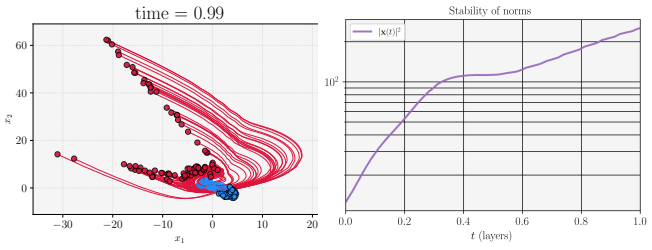
式 (4) 右辺第一項は，Otto–calculus で重要な役割を演じ，ODE-Net の生成モデルへの応用でも安定化を実現している [Fin+20]．本研究の設定でも，L_{kinetic} による正則化は，“運動エネルギー保存則” を導き，数学解析（定理 3.2）の上でも，数値実験においても（図 2），ODE-Net の安定化を達成する．

定理 3.2.　(運動エネルギー保存則)　$L = L_{\text{kinetic}}, v(x,\theta) = \theta f(x)$ としたとき，ODE-Net の学習問題に解が存在するのであれば，その解は

$$\frac{\lambda}{2}\int_{\mathbb{R}^d} |\theta f(x)|^2 \, \mathrm{d}\mu(x) + \frac{\epsilon}{2}|\theta|^2$$

を保存量にもつ．

さらに，ODE-Net の学習問題には，$\theta \in L^2(0, T; \theta)$ なる解が存在する．



(a) μ_t の軌跡（アニメーション）　(b) $W_2(\mu_t, \delta)^2$ の履歴

図 2．二値分類問題を，$L = L_{\text{kinetic}}$ のときの ODE-Net に学習させた結果．二次モーメント $W_2(\mu_t, \delta)^2$ が高々線形程度にしか増加していないことがわかる．

参考文献

[Car+19] F. Carrara et al. “On the Robustness to Adversarial Examples of Neural ODE Image Classifiers.” *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*. 2019, pp. 1–6.

[Che+18] T. Q. Chen et al. “Neural Ordinary Differential Equations.” *NeurIPS*. 2018, pp. 6572–6583.

[E17] W. E. “A Proposal on Machine Learning via Dynamical Systems.” *Communications in Mathematics and Statistics* **5**.1 2017, pp. 1–11.

[EHL18] W. E, J. Han, and Q. Li. “A Mean-Field Optimal Control Formulation of Deep Learning.” *Research in the Mathematical Sciences* **6**.1 2018, p. 10.

[Est+20] C. Esteve et al. “Turnpike in Lipschitz-Nonlinear Optimal Control.” 2020. arXiv: 2011.11091 [math.OC].

[Fin+20] C. Finlay et al. “How to Train Your Neural ODE: The World of Jacobian and Kinetic Regularization.” *Proceedings of the 37th International Conference on Machine Learning*. Ed. by H. D. III and A. Singh. **119**. Proceedings of Machine Learning Research. PMLR, 2020, pp. 3154–3164.

[GSS15] I. J. Goodfellow, J. Shlens, and C. Szegedy. “Explaining and Harnessing Adversarial Examples.” 2015. arXiv: 1412.6572 [stat.ML].

[Zas05] A. Zaslavski. “Turnpike Properties in the Calculus of Variations and Optimal Control.” **80**. Springer Science & Business Media, 2005.