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

磯部 伸（情報理工学系研究科数理情報学専攻）

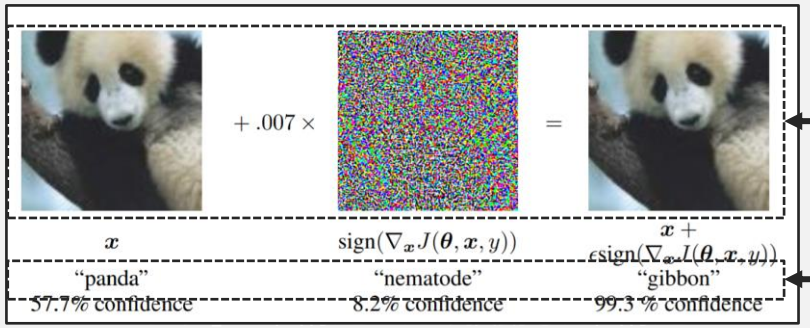
導入１ ～Deep Neural Networkの脆弱性～

以下，DNNと略記

DNN：人工知能技術の中核 ☺ しかし…

不安定な挙動を理解・制御不能 ☹

例．（敵対的事例） [Goodfellow+,14]



←入力(画像)

←出力(動物の種類)

「パンダ」の画像を ε だけ摂動した入力を
「テナガザル (gibbon)」とDNNは判定してしまった ☹

➢ **安定性**が保証されたDNNが必要

$$(\text{Output error}) \leq C(\text{Input perturbation})$$

しかし…

“Deep”を理解/制御するのは困難 ☹

（DNNを構成する）多重の関数合成

導入2 ～力学系概念の導入による安定化～

“Deep”を理解/制御するのは困難

➢ “**ODE**”として理想化する試みの勃興 [Chen+,18][E,17][Haber+,17]…

概念図

DNN (ResNet)

$$u \mapsto x_1 \mapsto x_2 \mapsto x_3 \mapsto x_4 \mapsto x_5 \mapsto y$$

$$x_{n+1} = x_n + v(x_n, \theta_n)$$

目的：

入力 u に対する出力 y の安定性

「理想化」

ODE-Net

$$u \mapsto x(t), 0 \leq t \leq 1 \mapsto y$$

$$\dot{x}(t) = v\left(x(t), \theta(t)\right)$$

中間目標：

初期値 $x(0)$に対するODEの安定性

パラメータ θ は，データの“学習”により決定

有限個の入/出力の組から定まるある関数の，最小化

既存研究：ある“機構”を学習**前**に設計し，
安定化を目指す

（“機構”の例：Lyapunov関数/Hamiltonianを導入 etc…）

設計思想：スケールが大き過ぎないように…

$$|x(t)|$$

問題意識と研究目標 ～学習“後”に着目～

既存研究：学習**前**に設計…

疑問（問題意識）

学習**後**，ODE-Netが
設計通りの働きをしているのか？

そもそも学習問題の解存在すら，
（私の知る限り）あまり研究されていない

そこで，DNN（≈離散化版ODE-Net）
の安定性保証を念頭におきながら…

本研究の目標

学習問題の解としてODE-Netを扱った上で

- 解存在，(安定性に関する) 解挙動の解明
- 安定化手法の開発・検証

この目標をクリアして得られた知見を基に
（定量的）安定性評価を目指す（今後の目標）

本研究で得られた成果の概要

ODE-Netを“平均場最適制御問題”として定式化した上で…

≒確率分布の空間上の束縛付き変分問題

主結果１

ターンパイク理論を用いた議論により，
スケールが指数的に増大する可能性を証明．

主結果２

ODE-Netが従う**保存則**に着目する観点から
学習後の挙動を考察．保存則の正当化．

主結果３

主結果２の考察を踏まえて，
新たに**運動論的正則化**を導入．そして

- “運動エネルギー保存則”
- 解の存在定理

を証明．

これらの結果の当該分野における**貢献**：

学習**後**のODE-Netの数理解析的性質を
明らかにし，ODE-Netの理論基盤を固めた

以上，一般向けの内容

以下，やや専門家向けの内容

問題設定 ～平均場最適制御問題～

理論解析の為，ODE-Netを以下で定式化

（定式化の考え方：流体力学におけるEuler的観点）

ODE-Netの学習問題 (★) [E+,19]

Given : $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d \times \mathcal{Y})$; （訓練）データ

コンパクト台を持つ確率測度のなす空間 with Wasserstein metric

Find $(\mu, \theta) \in C([0,T]; \mathcal{P}_c(\mathbb{R}^d) \times \mathcal{Y}) \times L^2(0,T;\Theta)$

such that

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

subject to

$$\begin{cases} \partial_t \mu_t + \nabla_x \left(v(\bullet, \theta_t) \mu_t \right) = 0 \text{ for } t \in (0,T), \\ \mu_t|_{t=0} = \mu_0. \end{cases}$$

（超関数の意味で）

ここで，v: $\mathbb{R}^d \times \Theta \to \mathbb{R}^d$, ℓ: $\mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}^d$,

Neural Net. パラメータの集合　損失関数　正解ラベルの集合

L: $\mathbb{R}^d \times \Theta \to \mathbb{R}, T > 0$. であり，これらは

正則化項

適当に滑らかな写像であると仮定

主結果 1 ～指数的ターンパイク評価～

直感：学習問題 (★) の解は定常問題 $\inf \widetilde{J}$ の解の近傍に留まっていそう...

この直感から，学習問題の解の挙動について，$L(x,\theta) = \lambda_x |x|^2 + \lambda_\theta |\theta|^2 (\lambda_x, \lambda_\theta > 0)$ の場合に次の評価が得られる：

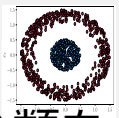
定理（気持ち：解は初期値&終端値付近で急変化し得る）

ある条件（詳しくは修論要旨を参照）を仮定する．このとき，ある定数 $C, k > 0$ が存在して，「T が充分大きいならば，学習問題(★)の任意の解 (μ, θ) について

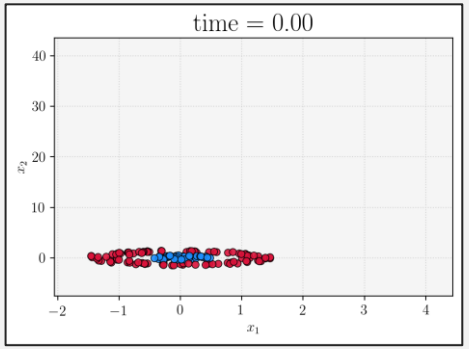
$$\int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\mu_t(x) \leq C \left(\mathrm{e}^{-kt} + \mathrm{e}^{-k(T-t)} \right)$$

が成立．」

実際に同心円状データ



の二値分類をさせてみると...☞（クリック）



主結果２，３ ～保存則に着目～

実は，**学習したODE-Netはある保存則に従う**

定理

学習問題 (★) を $L(x,\theta) = \lambda_x |x|^2 + \lambda_\theta |\theta|^2$ と $v(x,\theta) = \theta f(x)$ の下で考える．このとき，もし問題 (★) が $\theta \in L^\infty(0,T;\Theta)$ なる解を持つなら，その解は以下を満たす：

$$\frac{\lambda_\theta}{2} |\theta_t|^2 - \frac{\lambda_x}{2} \int_{\mathbb{R}^d \times \mathcal{Y}} |x|^2 \, \mathrm{d}\mu_t(x) = \text{const.}$$

この定理の知見を活かし，**運動論的正則化**

$$L_{\text{kinetic}}(x,\theta) := \frac{\lambda}{2} |v(x,\theta)|^2 + \frac{\epsilon}{2} |\theta|^2 \quad (\lambda, \epsilon > 0)$$

を導入．この設定で以下を示した：

定理（“運動エネルギー”保存則）

$$\underbrace{\frac{\lambda}{2} \int_{\mathbb{R}^d} |\theta_t f(x)|^2 \, \mathrm{d}\mu_t(x)}_{\text{“運動エネルギー”}} + \frac{\epsilon}{2} |\theta_t|^2 = \text{const.}$$

定理（解の存在定理）

学習問題 (★) には解が存在する．

既存研究 [Thorpe & Gennip,20] で必要なH^1正則化が不要 ☺

今後の展望 ～平均場ゲームの数値解析～

運動論的正則化の下での学習問題

$$\inf_{\mu,\theta} \int_{\mathbb{R}^d \times \mathcal{Y}} \ell \, \mathrm{d}\mu_T + \int_0^T \frac{\lambda}{2} \int_{\mathbb{R}^d} |v(x,\theta_t)|^2 \, \mathrm{d}\mu_t(x) \mathrm{d}t$$

は，Neural Network v が“万能”だと思うと

任意のベクトル場を $v(\bullet,\theta)$ の形で表すことができる

$$\inf_{\mu,\boldsymbol{v}} \int_{\mathbb{R}^d \times \mathcal{Y}} \ell \, \mathrm{d}\mu_T + \int_0^T \frac{\lambda}{2} \int_{\mathbb{R}^d} |\boldsymbol{v}_t(x)|^2 \, \mathrm{d}\mu_t(x) \mathrm{d}t$$

という，**平均場ゲーム** (cf. [Benamou+,17]) 問題になり，こちらでは解の一意存在や安定性評価が確立されている．

そこで今後は，（逆に）DNNの学習問題を，平均場ゲームの

- Neural Network による**関数近似**
- 確率測度の（**空間**）**離散化**
- ODEの（**時間**）**離散化**

と見做し，（時空間変分問題に対する）数値解析学の発展と連動させることで，ODE-NetやDNNの理論保証をおこなっていきたい．

■