

積分量を出力するニューラルネットワークを用いた ファンデルポール方程式の解の予測

メタデータ	言語: Japanese 出版者: 武蔵野大学数理工学センター 公開日: 2023-03-14 キーワード (Ja): キーワード (En): 作成者: 丸野, 恵蔵, 高石, 武史, 谷口, 隆晴 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/1981

積分量を出力するニューラルネットワークを用いたファンデル ポール方程式の解の予測

Predicting The Trajectory of The Solution of van der Pol equation Using a Neural Network that Outputs The Integral Values

丸野 恵 蔵¹

Keizo Maruno

高石 武 史²

Takeshi Takaishi

谷 口 隆 晴³

Takaharu Yaguchi

概要

ファンデルポール方程式に現れる非線形振動をニューラルネットワークを用いて予測するために、ハミルトニアンニューラルネットワークに用いられた手法を応用し、入力変数の時間変化の積分にあたるスカラー量を出力するニューラルネットワークから軌道の予測を行った。提案したニューラルネットワークはいずれも解軌道を予測できることがわかったが、ファンデルポール方程式の特性を反映させた予測モデルは、学習データのノイズがある程度大きい場合においても位相がずれずに解軌道を予測できることがわかった。

1 はじめに

古典力学で記述できる現象をニューラルネットワークで予測するための方法がいくつか提案されている。時系列データに対して、そのダイナミクスを考慮せずに、単純に過去のデータを入力として次の時間ステップでの値を求める方法は、学習時と等しい時間間隔でのデータの予測に限定される上、例えばエネルギーが保存するような場合に、その拘束条件を満たすように運動を再現することは難しい。ニューラルネットワークを用いた時間発展問題の予測に対して、これらを改善する手法としては、ニューラル常微分方程式 (Neural Ordinary Differential Equation, NODE) [2] や、それを離散化した ResNet[1] が挙げられる。これらはダイナミクスの予測とは異なる目的で提案されたものであるが、ダイナミクスの予測にも利用可能である。NODE は常微分方程式 $\frac{dx}{dt} = f(x)$ の右辺項をニューラルネットワークに学習させることで $x(t)$ を連続量として扱えるうえ、ode solver と組み合わせることで高次の予測を可能としている。また、ResNet は、NODE を離散化したものと考えることができ、次の時間ステップ

¹ 武蔵野大学大学院工学研究科数理工学専攻 / アビームシステムズ株式会社

² 武蔵野大学数理工学センター員 / 武蔵野大学工学部数理工学科教授

³ 神戸大学 大学院システム情報学研究科

の値ではなく、時間ステップ間での差分 $\Delta x = x^{n+1} - x^n$ を学習させることで予測精度を上げる方法と解釈することができる。この2つの方法で予測精度の向上や連続的な時間変化へ対応できるようになったものの、エネルギー保存が必要な問題においてこの条件を精度良く満たすことは難しい。

ハミルトニアンニューラルネットワーク (HNN) では、運動における状態変数の時間変化を直接予測するのではなく、エネルギー関数であるハミルトニアンを予測することで、エネルギー保存則を満たした運動を正確に再現できることが知られている [3]。特に、この研究では、ハミルトン方程式という物理的な知識と、ニューラルネットワークによるエネルギー関数の予測を組み合わせることで精度を向上させており、大枠としてのモデル方程式 (ハミルトン方程式) をニューラルネットワークと組み合わせることで、現象の正確な予測・解析が可能となることを示したといえる。

筆者らは、入力データの時間変化の積分であるスカラー量を出力するという HNN の手法を一般の力学系に応用することで、非線形振動を精度良く予測ができるか調べた。具体的には、ファンデルポール方程式を題材に、積分量を出力とするニューラルネットワークを構築することで十分な精度で予測が可能であり、学習データにある程度の大きさのノイズが入った場合にはスカラー量に関する拘束条件があるモデルの方が非線形振動が正確に再現されているらしいことを見出した [4, 5]。本研究では、これらの成果を踏まえ、そこから得られる解軌道 (リミットサイクル) の再現性を、振動の位相のずれにも注目することでどの程度運動の予測が可能か調べた。いくつかの適用方法を検討しその予測結果を比較したところ、それぞれの方法の有効性が確認でき、また、学習用入力データにノイズを加えたところ、解軌道の再現性について適用方法による差異が生じることがわかった。2 ではハミルトニアンニューラルネットワークについて説明し、この手法のファンデルポール方程式への適用方法について述べる。3 では数値実験の方法と結果について述べ、4 では結果について考察する。

2 ニューラルネットワークによる力学系の解の予測

ハミルトン力学で記述される保存系の運動においては、ハミルトニアン $H(q, p)$ を用いて系の運動が次のように記述される。

$$\begin{cases} \frac{dq}{dt} = \frac{\partial H}{\partial p} \\ \frac{dp}{dt} = -\frac{\partial H}{\partial q} \end{cases} \quad (1)$$

このとき、

$$\frac{dH}{dt} = \frac{\partial H}{\partial q} \frac{dq}{dt} + \frac{\partial H}{\partial p} \frac{dp}{dt} = \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} = 0$$

となることから、(1) に従う運動においては系のエネルギーにあたるハミルトニアン H は時間に関して一定となる (保存する)。ハミルトニアンニューラルネットワークでは、ニューラルネットワークによる出力を入力データの時間変化 $\frac{dq}{dt}, \frac{dp}{dt}$ の積分量である H とし、ネットワー

積分量を出力する NN を用いたファンデルポール方程式の解の予測 (丸野, 高石, 谷口)

クを用いた自動微分で $\frac{\partial H}{\partial p}, \frac{\partial H}{\partial q}$ を計算することで, H を保存したまま運動を再現できるようにしている. また, 学習においては損失関数として次の \mathcal{L}_{HNN} を最小化するようにする.

$$\mathcal{L}_{HNN} = \left\| \frac{dq}{dt} - \frac{\partial H}{\partial p} \right\|^2 + \left\| \frac{dp}{dt} + \frac{\partial H}{\partial q} \right\|^2 \quad (2)$$

ここで $\frac{dq}{dt}, \frac{dp}{dt}$ のデータが必要となるが, そのようなデータが得られない場合には, 各々時系列データ $(q^n, p^n), n = 1, 2, 3, \dots$ の差分 $\frac{q^n - q^{n-1}}{\Delta t}, \frac{p^n - p^{n-1}}{\Delta t}$, で与えることで, 時系列データのみからニューラルネットワークの重みを学習することもできる. そのため, 実験データなどを入力に用いることも可能で, 広い応用が期待される.

この方法は次のような散逸系の場合に応用することができる. ポテンシャル関数 $E(q, p)$ を用いて, 系の運動が次のような勾配流の方程式で記述されるとする.

$$\begin{cases} \frac{dq}{dt} = -\frac{\partial E}{\partial q} \\ \frac{dp}{dt} = -\frac{\partial E}{\partial p} \end{cases} \quad (3)$$

このとき,

$$\frac{dE}{dt} = \frac{\partial E}{\partial q} \frac{dq}{dt} + \frac{\partial E}{\partial p} \frac{dp}{dt} = -\left(\frac{\partial E}{\partial q}\right)^2 - \left(\frac{\partial E}{\partial p}\right)^2 \leq 0$$

となり, ポテンシャル関数 E は時間とともに減少することがわかる. ここで, ハミルトニアニューラルネットワークの手法を応用し, 出力をポテンシャル関数 E とするようなニューラルネットワークを構築し, ネットワークを用いた自動微分で $\frac{\partial E}{\partial q}, \frac{\partial E}{\partial p}$ を計算することで, E が時間とともに減少する性質を保持したまま運動を再現することが期待できる. このとき, 学習においては損失関数として次の \mathcal{L}_{grd} を最小化するようにする.

$$\mathcal{L}_{grd} = \left\| \frac{dq}{dt} + \frac{\partial H}{\partial q} \right\|^2 + \left\| \frac{dp}{dt} + \frac{\partial H}{\partial p} \right\|^2 \quad (4)$$

一般の力学系について, この手法を適用する. ここでは次の 2 変数の力学系について考える.

$$\begin{cases} \frac{dq}{dt} = f(q, p) \\ \frac{dp}{dt} = g(q, p) \end{cases} \quad (5)$$

において, $\frac{\partial F}{\partial q} = f(q, p), \frac{\partial G}{\partial p} = g(q, p)$ となるスカラー関数を用いて,

$$\begin{cases} \frac{dq}{dt} = \frac{\partial F}{\partial q} \\ \frac{dp}{dt} = \frac{\partial G}{\partial p} \end{cases} \quad (6)$$

の方程式を解くことにする. このスカラー関数 $F(q, p), G(q, p)$ をニューラルネットワークの出力とし, ネットワークを用いた自動微分で $\frac{\partial F}{\partial q}, \frac{\partial G}{\partial p}$ を計算することで, 運動の再現が可能かどうか調べてみる. この場合はスカラー関数が変数の数だけ必要となり, また, エネルギーの

保存や減少といった性質の保持はないが、よりなめらかなプロファイルを持つことが期待されるスカラー関数をニューラルネットワークに予測させることで、学習時間の低減や、再現した運動の精度の向上が可能となることを期待している。このとき、学習においては損失関数として次の \mathcal{L}_{dyn} を最小化するようにする。

$$\mathcal{L}_{dyn} = \left\| \frac{dq}{dt} - \frac{\partial F}{\partial q} \right\|^2 + \left\| \frac{dp}{dt} - \frac{\partial G}{\partial p} \right\|^2 \quad (7)$$

ファンデルポール方程式はコンデンサー、コイルに加えて、負性抵抗を持つ電子回路における電流の変動を表す方程式として、次のように書き表すことができる。

$$\begin{cases} \frac{dq}{dt} = -\mu \left(\frac{1}{3}q^3 - q \right) + p \\ \frac{dp}{dt} = -q \end{cases} \quad (8)$$

ここで、 $F(q, p) = -\mu \left(\frac{1}{12}q^4 - \frac{1}{2}q^2 \right) + pq$, $G(q, p) = -pq$ とすると、(8) は (6) として表すことができる。本研究では、 $\frac{dq}{dt}$, $\frac{dp}{dt}$ の積分量であるスカラー関数 F, G をニューラルネットワークの出力とし、自動微分で $\frac{\partial F}{\partial q}$, $\frac{\partial G}{\partial p}$ を計算することで、解の軌道の再現がどの程度可能なのか確かめる。一方、ファンデルポール方程式 (8) において、 $\frac{\partial F}{\partial p} = q$ となるので、

$$\begin{cases} \frac{dq}{dt} = \frac{\partial F}{\partial q} \\ \frac{dp}{dt} = -\frac{\partial F}{\partial p} \end{cases} \quad (9)$$

と書くこともできる。このとき、 $\frac{dF}{dt} = \frac{\partial F}{\partial q} \frac{dq}{dt} + \frac{\partial F}{\partial p} \frac{dp}{dt} = \left(\frac{\partial F}{\partial q} \right)^2 - \left(\frac{\partial F}{\partial p} \right)^2$ となるため、スカラー関数は時間とともに保存する性質も減少する性質も持たないが、出力するスカラー関数は1つで済み、ネットワークの重みの最適化においてより制限が課されることから、学習効率や解軌道の再現性が向上することが予想される。

3 数値計算結果

3.1 積分量を出力するニューラルネットワークでの予測

本研究では、[5] と同様に、(6) を用いて、入力を (q, p) とし出力を (F, G) としたネットワーク (図 3(a), 以下 1net) で学習させた。200 ノードを持つ層を 3 層全結合し、活性化関数として $\tanh(x)$ を用いたネットワークを使った。本研究では、リミットサイクルを持つ $\mu = 2$ の場合について、緩和振動を生じる解軌道の再現性を調べた。平均 0 分散 4 となるような正規分布から選んだ 100 組の乱数を q, p の初期値として、 $\Delta t = 8 \times 10^{-2}$ で (8) を $t \in [0, 40]$ の範囲で数値計算した 100 個の軌道によって学習用データセットを作成し、入力データとして学習させた。図 1 から、学習率を $lr = 10^{-5}$ とした場合には 2000 エポック程度でほぼ学習が終了していることがわかったため、以下は 5000 エポックまで学習させたネットワークを用い

積分量を出力する NN を用いたファンデルポール方程式の解の予測 (丸野, 高石, 谷口)

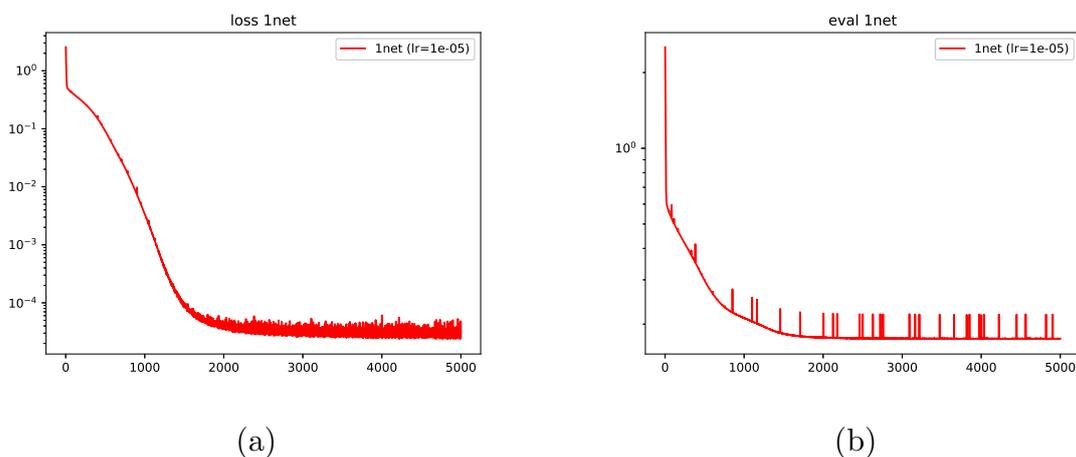


図 1 学習率 $lr = 10^{-5}$ における 1net (図 3(a)) の学習結果. (a) 学習データでの損失関数の値 (loss) と, (b) 評価データでの損失関数の値 (eval).

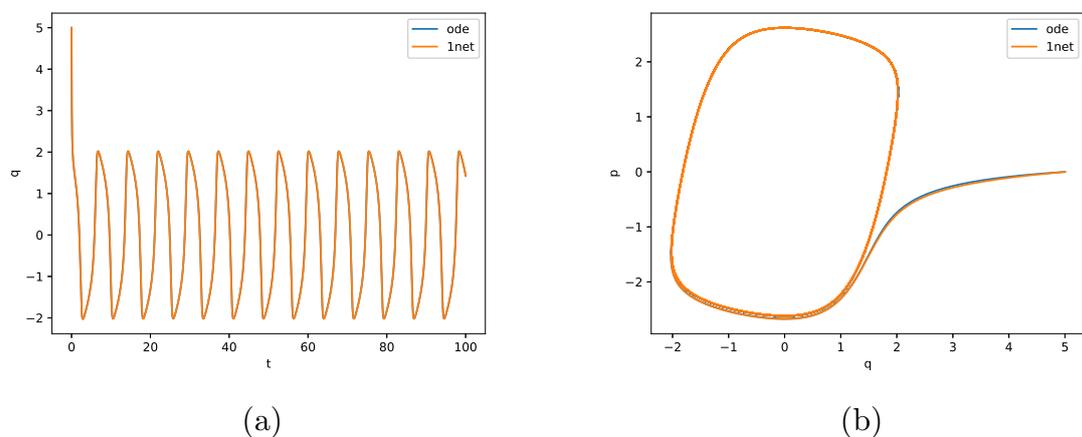


図 2 学習率 $lr = 10^{-5}$ での 1net (図 3(a)) での再現結果. $(q, p) = (5, 0)$ を初期値として予測した解軌道における (a) q の時間発展と $q - p$ の相図.

て, 解軌道の再現性を確認する. 1net (図 3(a)) で 5000 エポックまで学習済みのネットワークを用いて, 初期値 $(q, p) = (5, 0)$ から $\Delta t = 10^{-3}$ で $t \in [0, 100]$ での軌道を再現した結果を (8) の数値解 (以下 ode) 比較した (図 2). q の時間発展 (左), $q - p$ での相空間での表示とともに, ネットワークでの再現結果 (1net) と (8) の数値解の差はほとんど見られず, 解の軌道を十分再現できていることがわかる.

3.2 ネットワークの比較

以下では図 3 の 4 種類のネットワークでの軌道の再現性について比較した. 積分量を出力するネットワークとしては次の 2 種類を比較した. 図 3(a) と同様に (6) を用いて, 入力を (q, p) とし出力をスカラー量 (F) としたネットワークと, 同じ入力で出力をスカラー量 (G) とした

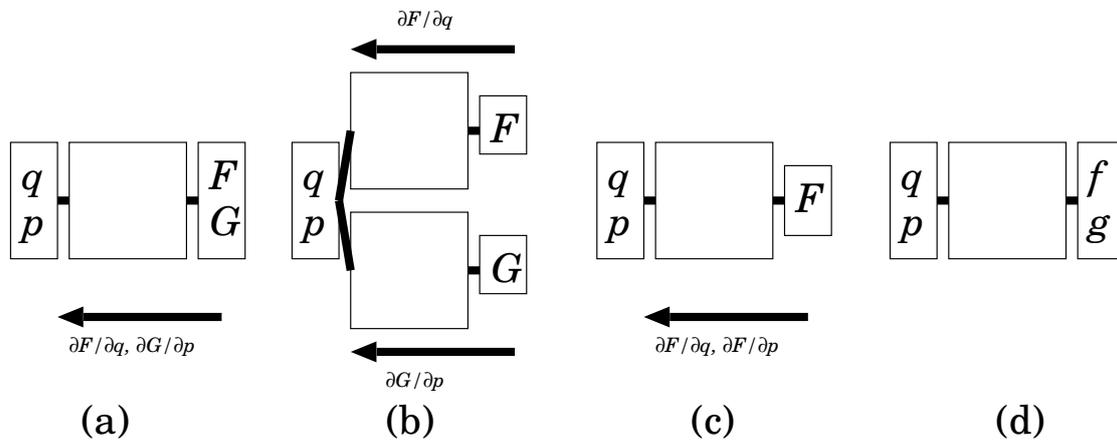


図3 ファンデルポール方程式の解軌道の学習に用いたネットワーク．積分量を出力するニューラルネットワーク ((a)1net, (b)2net, (c)F1net) と, (5) の右辺 f, g を学習させたネットワーク ((d)NN) ．

ネットワークを並列させて学習させた場合 (図 3(b), 以下 2net), (9) を用いて, 入力を (q, p) とし出力を単一のスカラー量 (F) としたネットワークで学習させた場合 (図 3(c), 以下 F1net) を使用した. 積分量を出力するニューラルネットワーク ((a)1net, (b)2net, (c)F1net) では (6) や (9) 右辺のスカラー量 F または G を出力し, 自動微分を用いて $\frac{dq}{dt}, \frac{dp}{dt}$ を計算する. その他に, (5) に基づいて, 入力を (q, p) に対して, 入力の時微分 $\frac{dq}{dt}, \frac{dp}{dt}$ に相当するベクトル (f, g) が出力となるように学習させたネットワーク (図 3(d), 以下 NN) を比較に用いた. このネットワークは一つ前の時系列データも用いて出力を計算するため, NODE[2] の 1 種類と考えることができる. F1net と NN は 1net と同様のネットワークを用いているが, 2net は各層を 100 ノードにした上で, 2つのネットワークの重みを同時に最適化している.

前節と同様の学習用データセットを用いて各ネットワークを学習させたところ, 2net の学習に少し多めにエポック数が必要となっているものの, 最終的には学習データでの損失関数の値はほぼ同程度まで下がっているが, 評価用データを用いた場合はネットワークによって差が生じていることがわかった (図 4). 5000 エポックで学習済みのネットワークを用いて, 初期値 $(q, p) = (5, 5)$ から $\Delta t = 10^{-3}$ で $t \in [0, 100]$ での軌道を再現した結果を比較した (図 5). どのネットワークでも (8) の数値解との差はほとんど見られず, 解の軌道を十分再現できていることがわかる.

3.3 ノイズを加えたデータでの学習結果

次に, ネットワークが学習する際の頑健さ (ロバストネス) を調べるために, 入力データとしてガウスノイズ (分散 $\sigma = 0.1$) を加えた学習用データセットを作成し, 各ネットワークで学習させた (図 7). 評価データでの損失関数の変化から, (9) を用いた F1net の学習状況が比較的良好であることがわかる.

積分量を出力する NN を用いたファンデルポール方程式の解の予測 (丸野, 高石, 谷口)

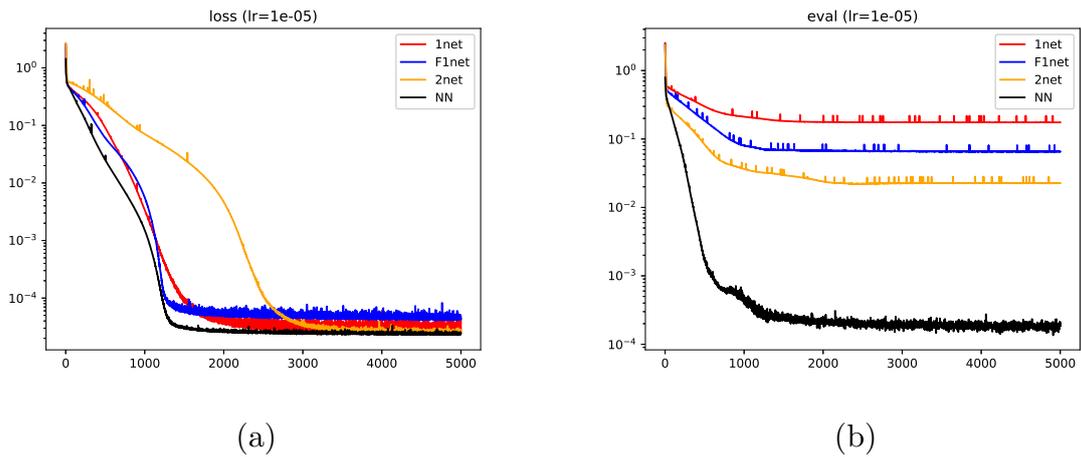


図4 各ネットワークの学習結果 ($lr = 10^{-5}$). (a) 学習データでの損失関数の値 (loss) と, (b) 評価データでの損失関数の値 (eval).

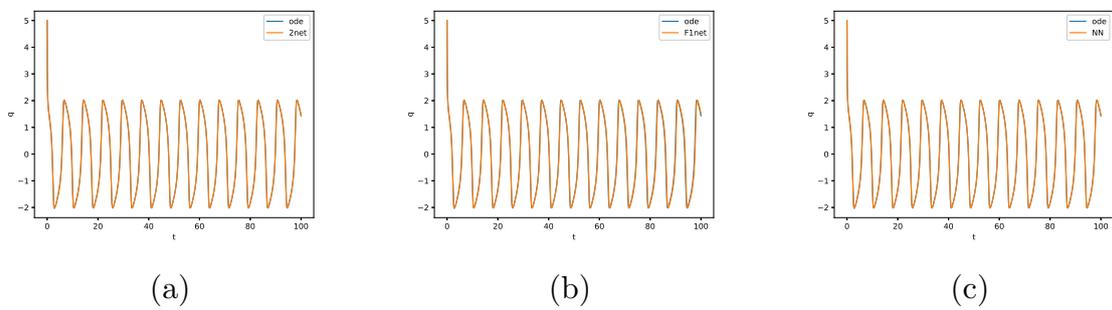


図5 図2と同様に, $lr = 10^{-5}$, 5000 エポックで学習済みの (a)2net, (b)F1net, (c)NN における, q の時間発展の再現結果.

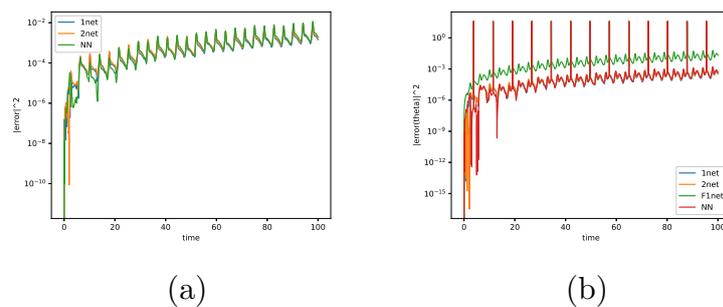


図6 $(q, p) = (1.9, 0)$ を初期値として $lr = 10^{-5}$, 5000 エポックで学習済みの (a)2net, (b)F1net, (c)NN における, q の時間発展を予測した場合の誤差.

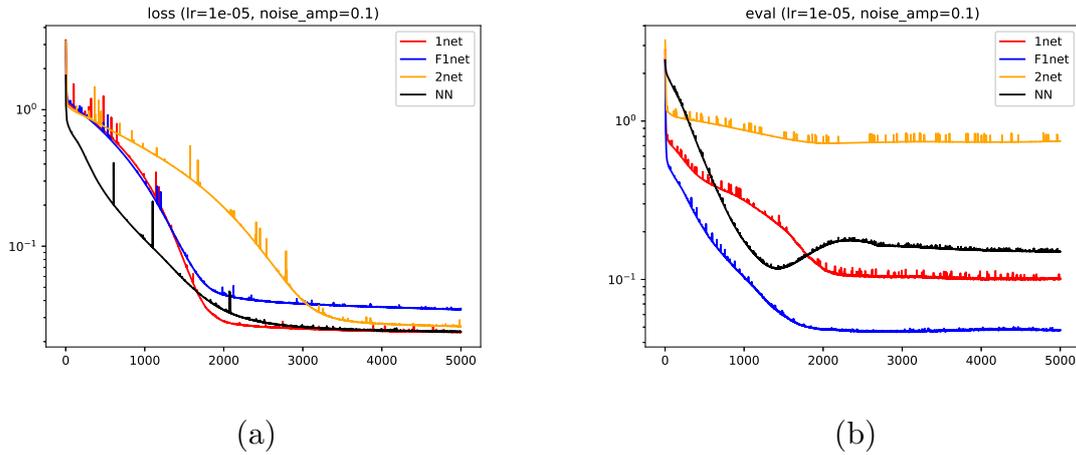


図7 $\sigma = 0.1$ のノイズを加えたデータでの各ネットワークの学習結果. (a) 学習データでの損失関数の値 (loss) と, (b) 評価データでの損失関数の値 (eval).

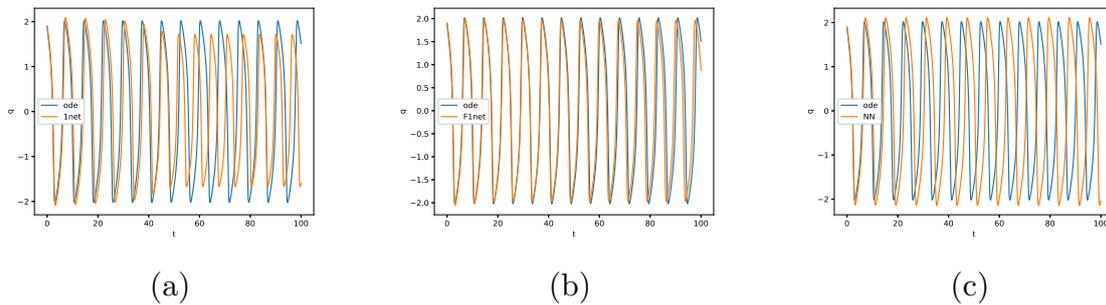


図8 $\sigma = 0.1$ のノイズを加えたデータで学習させた (a)1net, (b)F1net, (c)NN による, $(q, p) = (1.9, 0)$ を初期値として q の時間発展の予測を行った結果.

図2と同様に, 5000 エポックまで学習済みのネットワークを用いて, 初期値 $(q, p) = (1.9, 0)$ からの解軌道を比較すると, 1net は解軌道の振幅と位相がずれていき, NN は振幅はずれないが位相がずれていくのに対して, F1net は振幅と位相がともにずれることなく解軌道が再現できていることがわかる.

ここでは, 解軌道 $\vec{x}_i = (q_i, p_i), i = 1, \dots, N$ の再現性の評価として, ODE solver から得られた解に対して各ニューラルネットワークの出力をもとに計算された解軌道との平均2乗誤差 (MSE) と, q, p -平面での位相 θ との位相 $\tilde{\theta}$ との平均2乗誤差 ($e_\theta^2 = \frac{1}{N} \sum_{i=1}^N |\theta_i - \tilde{\theta}_i|^2$) を求めた. 学習データのノイズ量を $\sigma = 0, 0.01, 0.1$ とした場合に, ノイズのそれぞれのネットワークを用いて, 初期値 (q, p) を周期解のすぐ内側 $(1.7, 0)$, すぐ外側 $(1.9, 0)$ およびやや離れた外側 $(5, 0)$ とした場合の軌道の平均2乗誤差 (MSE) と位相の平均2乗誤差 (e_θ^2) を調べると, $\sigma = 0.01$ ではあまり差が出ないがノイズが大きめである $\sigma = 0.1$ でも F1net は他のネットワークに比べて解軌道を正確に再現できていることがわかる (図9). また, 位相の2乗誤差の時間発展からも, F1net での予測は安定してノイズに強いことがわかる (図10).

積分量を生じさせる NN を用いたファンデルポール方程式の解の予測 (丸野, 高石, 谷口)

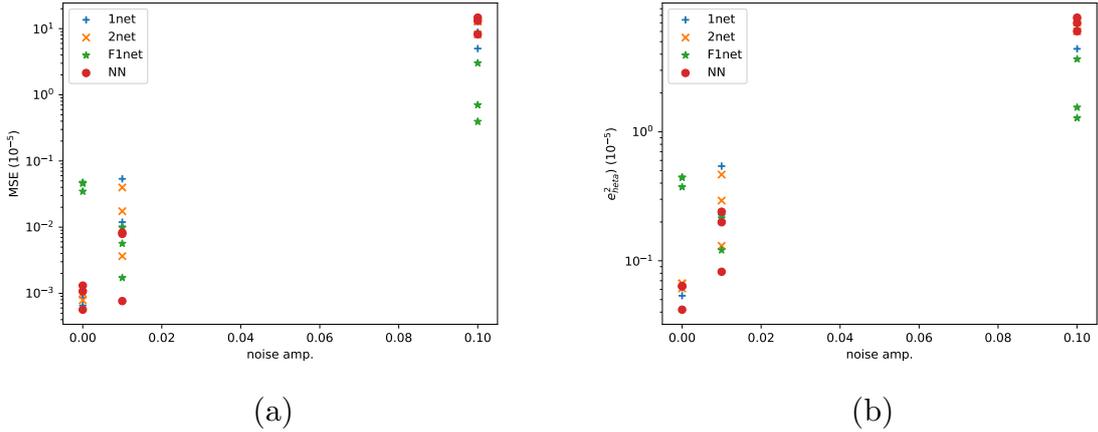


図9 各ネットワークにおける学習データのノイズの量に対する, (a) 予測軌道 (p, q) の誤差と, (b) 予測軌道の位相の誤差.

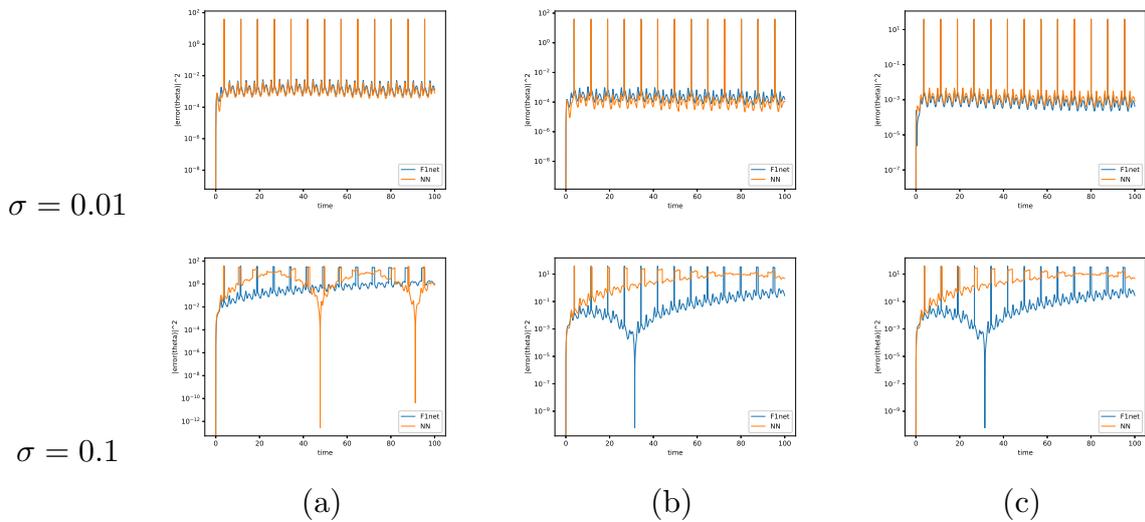


図10 分散 $\sigma = 0.01, 0.1$ のノイズを加えたデータで学習させた F1net, NN による, (a) $(q, p) = (1.7, 0)$, (b) $(1.9, 0)$, (c) $(5, 0)$ からスタートした軌道の予測における位相誤差 e_θ^2 の時間発展.

4 まとめ

ハミルトニアンニューラルネットワークの枠組みを用いて, 入力データ q, p に対して $\frac{dq}{dt}, \frac{dp}{dt}$ の積分量を生じさせるニューラルネットワークによって非線形振動を生じるファンデルポール方程式の解軌道が学習できないか試みた. ファンデルポール方程式の数値シミュレーションデータを学習したネットワークはリミットサイクルを含む解軌道を再現できることがわかり, $\frac{dq}{dt}, \frac{dp}{dt}$ を直接学習させたニューラルネットワークとほぼ同等の再現性であることがわかった. また, 入力データにある程度のノイズが混入する場合には, 方程式の性質 (9) を反映させた

ニューラルネットワーク (図 3(c)) がより正確に再現できる可能性があることがわかった。

HNN は, 等間隔な時系列データ (または, 時刻データと各時刻におけるデータ) を入力として用意するのみで教師データを必要とせず, 入力された特徴量と出力するハミルトニアンに相当するスカラー量との関係式であるハミルトン方程式をネットワーク学習の拘束条件として与えることで, エネルギーであるスカラー量 (ハミルトニアン) が時間に対して一定である運動を予測するシステムである. 本論文で取り上げたファンデルポール方程式は保存系でも純粋な散逸系でもないが, 数値計算の結果は, 入力した特徴量と出力したスカラーとを結びつける微分方程式を適切な形で仮定することでより良い予測ができる可能性を示している. この結果から, 特徴量とスカラー量の適切な関係式 (大枠としての数理モデル) を与えることで非線形な現象に対しても適切な予測ができるシステムを構築できることが期待される.

今回は安定な極限軌道を持つファンデルポール方程式を題材に選んだため, ノイズの少ない学習データでの解軌道の再現性において, 各ネットワークによる差異があまり見いだせなかった可能性が考えられる. より複雑な運動を行う方程式において, どの程度までこの方法で予測できるか, 今後の研究でさらに明らかにしたい.

参考文献

- [1] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [2] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, Neural Ordinary Differential Equations, *Advances in Neural Information Processing Systems*, 31 (NeurIPS 2018).
- [3] S. Greydanus, M. Dzamba, J. Yosinski, Hamiltonian Neural Networks, *Advances in Neural Information Processing Systems*, 32 (NeurIPS 2019).
- [4] 丸野 恵蔵, 積分を組み込んだニューラルネットワークを用いた, 力学系の運動予測, 武蔵野大学工学研究科数理工学専攻修士論文 2022.
- [5] 丸野 恵蔵, 高石 武史, 谷口 隆晴, 積分を組み込んだニューラルネットワークを用いた, 力学系の運動予測, 第 27 回計算工学講演会 2022.

(原稿提出: 2022 年 12 月 29 日; 修正稿提出: 2023 年 1 月 21 日)