Advances in Applied Mathematics 应用数学进展, **2023, 12(7), 3250-3260**
Published Online July 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.127324

Hans汉斯

一类深度神经网络的稳定性研究

倪炜达

浙江师范大学数学科学学院，浙江 金华

收稿日期：2023年6月18日；录用日期：2023年7月13日；发布日期：2023年7月20日

摘 要

深度神经网络近年来逐渐成为一个研究热点，它在建模复杂的数据集上面有着突出的表现。深度神经网络和动力系统有着潜在的联系，如何借助动力系统理论方法深入研究深度神经网络具有重要的理论和实际意义。本文首先介绍了**Haber**等人给出的三类可逆神经网络以及稳定性定理，然后介绍了林洁对于连续模型的稳定性做出的贡献，紧接着给出一个反例以此来说明**Haber**等人给出的关于离散模型的稳定性定理不严谨，然后对定理进行优化改进，得到新的判断欧拉格式稳定性的定理，最后将稳定性定理运用到一类哈密顿网络中。

关键词

深度神经网络，动力系统，稳定性

Research on Stability for a Class of Deep Neural Networks

Weida Ni

School of Mathematical Science, Zhejiang Normal University, Jinhua Zhejiang

Received: Jun. 18^{th}, 2023; accepted: Jul. 13^{th}, 2023; published: Jul. 20^{th}, 2023

文章引用: 倪炜达. 一类深度神经网络的稳定性研究[J]. 应用数学进展, 2023, 12(7): 3250-3260.
DOI: 10.12677/aam.2023.127324

Abstract

In recent years, deep neural networks have gradually become a research hotspot, and they have relatively good performance in modeling complex data sets. There is a potential connection between deep neural networks and dynamical systems. It is of great theoretical and practical significance to deeply study deep neural networks using dynamic system theory and methods. In this paper, firstly three types of reversible neural networks and stability theorems given by Haber *et al.* are introduced, and then Lin Jie's contribution to the stability of continuous models is presented. Next, we give a counter example to illustrate that the stability theorem for discrete models given by Haber *et al.* is not rigorous, and then optimize and improve the theorem to obtain a new theorem for judging the stability of Euler schemes. Finally, we apply the stability theorem to a class of Hamiltonian networks.

Keywords

Deep Neural Network, Dynamical System, Stability

1. 引言

机器学习的首要任务是创建一个足够丰富的函数类，可以以所需的精度表示数据。最直接的方法就是逼近论：从线性函数开始，然后使用样条曲线、小波或其他基函数来构建非线性函数 [1, 2]。这种方法的明显障碍就是维数，为了处理这个问题，人们通常必须假设函数的特殊形式，例如一个加法形式或乘法形式 [3]。近年来，一类新的技术取得了显著的成功，即深度神经网络模型 [4]，深度神经网络可以看成是由感知机衍生而来，感知机是由若干个输入以及一个输出组成，它的输入是某个对象的特征，输出是它所属类别。以此为基础，在其中加入多层隐藏层，增加输出的个数，并选择合适的激活函数，演化出来的系统就称为深度神经网络。最近的经验表明，具有多层的深度网络似乎在建模复杂的数据集方面做得出奇的好，并影响了社会的各个方面，从计算机视觉 [5]、自然语言处理 [6]到生物学 [7]和电子商务。神经网络与传统逼近论的区别在于，神经网络利用简单函数的组成来逼近复杂的函数。研究深度神经网络的过程就叫做深度学习，他是

机器学习领域中的一个新的研究方向。虽然我们仍然缺乏一个理解深度神经网络的理论框架，但它的实际成功却非常令人鼓舞。

为了更好的研究深度神经网络，Weinan E提出将深度神经网络看作是连续动力系统的离散化，那么从理论上讲，我们更容易理解和分析连续动力系统. 连续动力系统提供了更多的灵活性（例如添加约束，使动力系统适应问题，对动力系统附加一定的结构），并且比离散系统更容易分析。同时，从离散动力系统的角度来看，也有许多理论和方法可以借鉴。

2. 深度神经网络的稳定性介绍

深度学习中的一个困难来自于潜在的正向模型的不稳定性，这对特征的传播是很不利的。一些网络对于原始特征中的小扰动可能是不稳定的，即原有的特征在经过网络传播之后，会无限的放大，或者消失，我们把他称为梯度爆炸或梯度消失。这些结果令人不安，因为具有不稳定正向传播的网络所做的预测对输入特征的小扰动非常敏感，这可能会导致网络在实践中是无用的。因此，如何训练出一个稳定且适定的网络就成为了一个关键的问题。

于是Haber等人在 [8]中提出了一个关于深度神经网络的稳定性与可逆性的理论框架，推导出三种可逆神经网络架构

1) The two-layer Hamiltonian network

连续模型：

$$\begin{cases} \dot{Y}(t) = \sigma(K(t)Z(t) + b(t)), \\ \dot{Z}(t) = -\sigma(K(t)^T Y(t) + b(t)), \end{cases} \tag{a1}$$

离散模型：

$$\begin{cases} Y_{j+1} = Y_j + hK_{j1}^T \sigma(K_{j1}Z_j + b_{j1}), \\ Z_{j+1} = Z_j - hK_{j2}^T \sigma(K_{j2}Y_{j+1} + b_{j2}), \end{cases} \tag{a2}$$

2) The midpoint network

连续模型：

$$\dot{y} = \sigma((K - K^T)y + b), \tag{b1}$$

离散模型：

$$Y_{j+1} = \begin{cases} 2h\sigma((K_j - K_j^T)Y_j + b_j), j = 0, \\ Y_{j-1} + 2h\sigma((K_j - K_j^T)Y_j + b_j), j > 0, \end{cases} \tag{b2}$$

3) The leapfrog network

连续模型：

$$\ddot{Y}(t) = -K(t)^T \sigma(K(t)Y(t) + b(t)), \tag{c1}$$

离散模型：

$$Y_{j+1}=\begin{cases}2Y_j-h^2K_j^T\sigma(K_jY_j+b_j),j=0,\\2Y_j-Y_{j-1}-h^2K_j^T\sigma(K_jY_j+b_j),j>0,\end{cases}\tag{c2}$$

在这个基础上，Haber等人在 [9]中，提出了一种特殊形式的递归网络，称为反对称递归神经网络，首先讨论系统

$$\mathbf{y}'(t)=\tanh((\mathbf{W}_y-\mathbf{W}_y^T)\mathbf{y}(t)+\mathbf{V}_y\mathbf{x}(t)+\mathbf{b}(y)),\tag{1}$$

其中，$\mathbf{y}(t)\in\mathbb{R}^n$，$\mathbf{x}(t)\in\mathbb{R}^m$，$\mathbf{W}_y\in\mathbb{R}^{n\times n}$，$\mathbf{V}_y\in\mathbb{R}^{n\times m}$，$\mathbf{b}_y\in\mathbb{R}^n$，这里$\mathbf{W}_y-\mathbf{W}_y^T$是一个反对称矩阵，于是系统的Jacobi矩阵为

$$\mathbf{J}(t)=diag[\tanh'((\mathbf{W}_y-\mathbf{W}_y^T)\mathbf{y}(t)+\mathbf{V}_y\mathbf{x}(t)+\mathbf{b}(y))](\mathbf{W}_y-\mathbf{W}_y^T),\tag{2}$$

然后将系统(1)进行正向欧拉离散格式就得到了反对称递归神经网络

$$\mathbf{y}_t=\mathbf{y}_{t-1}+h\tanh((\mathbf{W}_y-\mathbf{W}_y^T)\mathbf{y}_{t-1}+\mathbf{V}_y\mathbf{x}_t+\mathbf{b}(y)),\tag{3}$$

其中$\mathbf{y}_t\in\mathbb{R}^n$是t时刻的隐藏状态，$\mathbf{x}_t\in\mathbb{R}^m$是t时刻的输入，$\mathbf{W}_y\in\mathbb{R}^{n\times n}$,$\mathbf{V}_y\in\mathbb{R}^{n\times m}$，$\mathbf{b}_y\in\mathbb{R}^n$是网络的参数，$h$表示步长。

上述网络模型的提出是基于Haber在文献 [10]中给出的一个关于残差网络稳定性的“定理”：正向欧拉格式是稳定的如果满足条件

$$\max_{i=1,2,\cdots,n}|1+h\lambda_i(J(t))|\le 1,$$

其中$J(t)$是离散系统的Jacobi矩阵。这个“定理”的条件其实是不充分的，我们将在下一节进行讨论。

3. 离散模型的稳定性判定

林洁在 [11]中针对文献 [10]给出的稳定性“定理”存在的问题，证明了一个关于连续模型的稳定性判据，接下来我将讨论离散模型的稳定性。Haber在文献 [10]中给出的一个关于残差网络稳定性的“定理”：

正向欧拉公式是稳定的如果满足条件

$$\max_{i=1,2,\cdots,n}|1+h\lambda_i(J(t))|\le 1,$$

其中$J(t)$是离散系统的Jacobi矩阵。这个“定理”的条件其实是不充分的，我们现在就来讨论完善它。

为了讨论这个“定理”的合理性，我们需要借助以下定义与定理

定义1. 若一种数值方法在节点值y_n上产生大小为δ的扰动，在以后各节点值$y_m(m>n)$上产生的偏差均不超过δ，则称该方法是稳定的。

考察下列模型方程$y' = \lambda y$，它的欧拉格式为

$$y_{n+1} = (1 + h\lambda)y_n, \tag{4}$$

设在节点y_n上有一扰动ϵ_n，它的传播使节点值y_{n+1}产生大小为ϵ_{n+1}的扰动值，设$y_n^* = y_n + \epsilon_n$，于是

$$\begin{aligned} y_{n+1}^* &= (1 + hy)y_n^* \\ &= (1 + hy)(y_n + \epsilon_n) \\ &= (1 + hy)y_n + (1 + hy)\epsilon_n \\ &= y_{n+1} + (1 + hy)\epsilon_n, \end{aligned}$$

则扰动值ϵ_{n+1}满足

$$\epsilon_{n+1} = (1 + h\lambda)\epsilon_n,$$

所以根据定义3.1，上述模型的欧拉方法的稳定性条件为

$$|1 + h\lambda| \leq 1,$$

但是这个方法只适用于一维系统的情况，对于n维系统，不能简单的推出结论。

考虑一个简单的二维方程

$$\dot{y} = Jy,$$

其中，$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$，它的欧拉离散格式为

$$y_{n+1} = (I + hJ)y_n,$$

将其写成分量的形式为

$$\begin{aligned} y_{n+1}^1 &= (1 + h\lambda)y_n^1 + hy_n^2, \\ y_{n+1}^2 &= (1 + h\lambda)y_n^2, \end{aligned} \tag{5}$$

设在节点$y_n = \begin{pmatrix} y_n^1 \\ y_n^2 \end{pmatrix}$上有一扰动$\epsilon_n = \begin{pmatrix} \epsilon_n^1 \\ \epsilon_n^2 \end{pmatrix}$，它的传播使节点值$y_{n+1} = \begin{pmatrix} y_{n+1}^1 \\ y_{n+1}^2 \end{pmatrix}$产生大小为$\epsilon_{n+1} = \begin{pmatrix} \epsilon_{n+1}^1 \\ \epsilon_{n+2}^2 \end{pmatrix}$ 的扰动值，记

$$\begin{pmatrix} z_n^1 \\ z_n^2 \end{pmatrix} = \begin{pmatrix} y_n^1 \\ y_n^2 \end{pmatrix} + \begin{pmatrix} \epsilon_n^1 \\ \epsilon_n^2 \end{pmatrix},$$

于是

$$\begin{aligned} z_{n+1}^1 &= (1 + h\lambda)z_n^1 + hz_n^2 \\ &= (1 + h\lambda)(y_n^1 + \epsilon_n^1) + h(y_n^2 + \epsilon_n^2) \\ &= y_{n+1}^1 + (1 + h\lambda)\epsilon_n^1 + h\epsilon_n^2, \end{aligned}$$

同理

$$\begin{aligned} z_{n+1}^2 &= (1+h\lambda)z_n^2 \\ &= (1+h\lambda))(y_n^2+\epsilon_n^2) \\ &= y_{n+1}^2 + (1+h\lambda)\epsilon_n^2, \end{aligned}$$

所以我们可以得到

$$\epsilon_{n+1} = \begin{pmatrix} 1+h\lambda & h \\ 0 & 1+h\lambda \end{pmatrix} \epsilon_n, \tag{6}$$

当$|1+h\lambda|<1$时，

$$\begin{aligned} \epsilon_{n+1}^1 &= (1+h\lambda)\epsilon_n^1 + h\epsilon_n^2 \\ &= (1+h\lambda)[(1+h\lambda)\epsilon_{n-1}^1 + h\epsilon_{n-1}^2] + h(1+h\lambda)\epsilon_{n-1}^2 \\ &= (1+h\lambda)^2\epsilon_{n-1}^1 + 2h(1+h\lambda)\epsilon_{n-1}^2 \\ &= (1+h\lambda)^{n+1}\epsilon_0^1 + (n+1)h(1+h\lambda)^n\epsilon_0^2, \\ \epsilon_{n+1}^2 &= (1+h\lambda)\epsilon_n^2 \\ &= (1+h\lambda)^{n+1}\epsilon_0^2, \end{aligned} \tag{7}$$

因为$|1+h\lambda|<1$，于是当$n\to+\infty$时，$(1+h\lambda)^{n+1}\to 0$，所以$\epsilon_{n+1}^2\to 0$，我们还需要说明$\epsilon_{n+1}^1\to 0$，所以只需证明

$$\lim_{n\to+\infty}(n+1)h(1+h\lambda)^n = 0$$

为此我们使用洛必达法则

$$\begin{aligned} &\lim_{n\to+\infty}(n+1)h(1+h\lambda)^n \\ &= \lim_{n\to+\infty}\frac{(n+1)h}{(1+h\lambda)^{-n}} \\ &= \lim_{n\to+\infty}\frac{h}{-(1+h\lambda)^{-n}\ln(1+h\lambda)} \\ &= \lim_{n\to+\infty}\frac{-h(1+h\lambda)^n}{\ln(1+h\lambda)} \\ &= 0, \end{aligned}$$

所以，当$|1+h\lambda|<1$时，系统的稳定性是可以保证的，但是，当$|1+h\lambda|=1$时

$$\begin{aligned} \epsilon_{n+1}^1 &= \epsilon_n^1 + h\epsilon_n^2, \\ \epsilon_{n+1}^2 &= \epsilon_n^2, \end{aligned}$$

于是

$$\begin{aligned}\epsilon_{n+1}^1 &= \epsilon_n^1 + h\epsilon_n^2 \\ &= \epsilon_n^1 + h\epsilon_0^2 \\ &= \epsilon_{n-1}^1 + 2h\epsilon_0^2 \\ &= \epsilon_0^1 + nh\epsilon_0^2,\end{aligned}$$

当$n \to +\infty$时，$\epsilon_{n+1}^1 \to +\infty$，显然它是不稳定的，于是我们需要Haber的定理进行修正。

考虑如下的微分方程

$$\dot{x} = f(t, x),$$

取它的一阶变分方程

$$\dot{y} = \mathbf{A}(t)y, \tag{8}$$

当$\mathbf{A}$是常数矩阵时，我们做相似变换$y = T\mathbf{w}$，于是

$$\dot{\mathbf{w}} = \Lambda\mathbf{w}, \ \ \Lambda \triangleq \mathbf{T}^{-1}\mathbf{A}\mathbf{T}. \tag{9}$$

其中$\Lambda = \begin{pmatrix} \Lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda_l \end{pmatrix}$，而$\Lambda_i$是 $r_i \times r_i$ Jordan 块 $\begin{pmatrix} \lambda_i & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{pmatrix}$，$\lambda_i$是 $\mathbf{A}$ 的特征值，$1 \le i \le l$，$m = \sum_{i=1}^{l} r_i$。它的欧拉格式为

$$\mathbf{w}_{n+1} = (\mathbf{I} + h\Lambda_n)\mathbf{w}_n,$$

由此我们可以得出以下稳定性定理：

定理1. 当$\mathbf{A}$是常数矩阵时，正向欧拉公式稳定的充要条件是$|1 + h\lambda_i| \le 1$,且$|1 + h\lambda_i| = 1$的特征值λ_i所对应的代数重数等于几何重数，即相应的*Jordan* 块为$\lambda_i\mathbf{I}$

证明：设$\mathbf{A}$是$n \times n$的常数矩阵，他有m个特征值$\lambda_1, \cdots, \lambda_k, \lambda_{k+1}, \cdots, \lambda_m$，其中$\lambda_1, \cdots, \lambda_k$满足条件$|1 + h\lambda_i| < 1$，$\lambda_{k+1} \cdots, \lambda_m$ 满足条件$|1 + h\lambda_i| = 1$且他们所对应的代数重数等于几何重数，于是我们将系统(8)，进行做相似变换$y = T\mathbf{w}$，得

$$\dot{\mathbf{w}} = \Lambda\mathbf{w}, \ \ \Lambda \triangleq \mathbf{T}^{-1}\mathbf{A}\mathbf{T}. \tag{10}$$

其中$\Lambda = \begin{pmatrix} \Lambda_1 & & & & & \\ & \ddots & & & & \\ & & \Lambda_k & & & \\ & & & \Lambda_{k+1} & & \\ & & & & \ddots & \\ & & & & & \Lambda_m \end{pmatrix}$，$\Lambda_p = \begin{pmatrix} \lambda_p & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_p \end{pmatrix}$，$1 \le p \le k$，$\Lambda_q = \lambda_q \mathbf{I}, k < q \le m$，于是(10)的欧拉格式为

$$\mathbf{w}_{n+1} = \begin{pmatrix} \mathbf{I}+h\Lambda_1 & & & & & \\ & \ddots & & & & \\ & & \mathbf{I}+h\Lambda_k & & & \\ & & & \mathbf{I}+h\Lambda_{k+1} & & \\ & & & & \ddots & \\ & & & & & \mathbf{I}+\Lambda_m, \end{pmatrix} \mathbf{w}_n \tag{11}$$

设在节点$w_j = \begin{pmatrix} w_j^1 \\ \vdots \\ w_j^n \end{pmatrix}$上有一扰动$\epsilon_j = \begin{pmatrix} \epsilon_j^1 \\ \vdots \\ \epsilon_j^n \end{pmatrix}$，它的传播使节点值$w_{j+1} = \begin{pmatrix} w_{j+1}^1 \\ \vdots \\ w_{j+1}^n \end{pmatrix}$产生大小为$\epsilon_{j+1} = \begin{pmatrix} \epsilon_{j+1}^1 \\ \vdots \\ \epsilon_{j+1}^n \end{pmatrix}$的扰动值。

我们按(11)中不同的分块来讨论扰动的变化，对于其中的第i个分块，他对应的扰动为$\epsilon_{j+1}^{r_i}, \cdots, \epsilon_{j+1}^{s_i}$，当$1 \le i \le k$时，有

$$\epsilon_{j+1}^{r_i} = (1+h\lambda_i)\epsilon_j^{r_i} + h\epsilon_j^{r_i+1},$$

$$\vdots$$

$$\epsilon_{j+1}^{s_i} = (1+h\lambda_i)\epsilon_j^{s_i},$$

于是根据前面的讨论，当$j \to +\infty$时，$\epsilon_{j+1}^{r_i}, \cdots, \epsilon_{j+1}^{s_i} \to 0$

当$k < i \le m$时

$$\epsilon_{j+1}^{r_i} = \epsilon_j^{r_i} = \epsilon_0^{r_i},$$

$$\vdots$$

$$\epsilon_{j+1}^{s_i} = \epsilon_j^{s_i} = \epsilon_0^{s_i},$$

他的扰动值不发生改变，于是根据上述两种讨论，稳定性得证。

当$\mathbf{A}(t)$与时间t有关时，假设$\mathbf{A}(t)$关于时间t慢变，即$\mathbf{A}(t)=\mathbf{D}+\mathbf{B}(t)$，满足条件$||\mathbf{B}(t)||<\epsilon, 0<\epsilon<1$，于是我们系统(8)就变为

$$\dot{y}=(\mathbf{D}+\mathbf{B}(t))y, \tag{12}$$

做相似变换$y=\mathbf{T}\mathbf{w}$，于是

$$\dot{\mathbf{w}}=\mathbf{T}^{-1}\mathbf{D}\mathbf{T}w+\mathbf{T}^{-1}\mathbf{B}(t)\mathbf{T}\mathbf{w}, \tag{13}$$

他的欧拉格式为

$$\mathbf{w}_{n+1}=\mathbf{w}_n+h\mathbf{T}^{-1}\mathbf{D}\mathbf{T}\mathbf{w}_n+\mathbf{T}^{-1}\mathbf{B}_n\mathbf{T}\mathbf{w}_n, \tag{14}$$

记$\mathbf{T}^{-1}\mathbf{D}\mathbf{T}=\tilde{\mathbf{D}}, \mathbf{T}^{-1}\mathbf{B}_n\mathbf{T}=\tilde{\mathbf{B}}_n$，于是系统就变为

$$\mathbf{w}_{n+1}=\mathbf{w}_n+h\tilde{\mathbf{D}}\mathbf{w}_n+\tilde{\mathbf{B}}_n\mathbf{w}_n, \tag{15}$$

于是

$$\epsilon_{n+1}^1=(1+h\lambda_i)\epsilon_n^1+h\epsilon_n^2+b_{11}(n)\epsilon_n^1+b_{12}(n)\epsilon_n^2+\cdots+b_{1n}(n)\epsilon_n^n,$$

假设D在复数域上可对角化，即$\tilde{\mathbf{D}}$是一个对角矩阵，于是单独讨论ϵ_{n+1}^1

$$\epsilon_{n+1}^1=(1+h\lambda_i)\epsilon_n^1+h\epsilon_n^2+b_{11}(n)\epsilon_n^1+b_{12}(n)\epsilon_n^2+\cdots+b_{1n}(n)\epsilon_n^n,$$

假设$\epsilon_n^{j_n}=\max\{\epsilon_n^1,\epsilon_n^2,\cdots,\epsilon_n^n\}$，那么

$$\begin{aligned}\epsilon_{n+1}^1&=(1+h\lambda_i)\epsilon_n^1+h\epsilon_n^2+b_{11}(n)\epsilon_n^1+b_{12}(n)\epsilon_n^2+\cdots+b_{1n}(n)\epsilon_n^n\\&<(1+h\lambda_i)\epsilon_n^{j_n}+h\epsilon_n^{j_n}+(b_{11}(n)+b_{12}(n)+\cdots+b_{1n}(n))\epsilon_n^{j_n}\\&=(1+h\lambda_i+h+b_{11}(n)+b_{12}(n)+\cdots+b_{1n}(n))\epsilon_n^{j_n},\end{aligned} \tag{16}$$

不妨设$b_{11}(n)+b_{12}(n)+\cdots+b_{1n}(n)<\eta$，其中$\eta$是一个任意小量，那么

$$\begin{aligned}\epsilon_{n+1}^1&=(1+h\lambda_i)\epsilon_n^1+h\epsilon_n^2+b_{11}(n)\epsilon_n^1+b_{12}(n)\epsilon_n^2+\cdots+b_{1n}(n)\epsilon_n^n\\&<(1+h\lambda_i+h+b_{11}(n)+b_{12}(n)+\cdots+b_{1n}(n))\epsilon_n^{j_n}\\&\approx(1+h\lambda_i+h)\epsilon_n^{j_n}\\&=(1+h\lambda_i+h)^n\epsilon_1^{j_1},\end{aligned} \tag{17}$$

由此我们可以得出以下稳定性定理：

定理2. 当$\mathbf{A}(t)$是与时间t有关时，假设$\mathbf{A}(t)$关于时间t慢变，即$\mathbf{A}(t)=\mathbf{D}+\mathbf{B}(t)$，$\mathbf{D}$在复数域上可对角化，且满足$|1+h\lambda_i(t)+h|\leq 1$, $b_{i1}(n)+b_{i2}(n)+\cdots+b_{in}(n)<\eta$，其中$\eta$是一个任意小量，那么正向欧拉公式稳定的。

由于时变矩阵的特殊性，定理(2)只能作为正向欧拉公式稳定的充分条件，更加完善的定理还需进行后续研究。

4. 稳定性的应用

根据上一节的稳定性定理我们来研究哈密顿网络

$$\begin{cases} \dot{\mathbf{y}}(t) = & \mathbf{K}^T(t)\sigma(\mathbf{z}(t)+\mathbf{b}(t)), \\ \dot{\mathbf{z}}(t) = & -\mathbf{K}(t)\sigma(\mathbf{y}(t)+\mathbf{b}(t)), \end{cases} \tag{18}$$

容易求得系统(18)关于(y,z)的Jacobi矩阵为

$$\mathbf{J}(t) = \begin{pmatrix} 0 & K(t)^T \\ -K(t) & 0 \end{pmatrix} \begin{pmatrix} diag(\sigma'(y+b)) & 0 \\ 0 & diag(\sigma'(z+b)) \end{pmatrix}, \tag{19}$$

假设$\varphi:(y_0(t),z_0(t))$是系统(18)的一个解，那么系统(18)在解φ处的线性化方程为

$$\begin{pmatrix} \dot{\epsilon} \\ \dot{\eta} \end{pmatrix} = \begin{pmatrix} 0 & K(t)^T diag(\sigma'(z_0(t)+b(t))) \\ -K(t)diag(\sigma'(y_0(t)+b(t))) & 0 \end{pmatrix} \begin{pmatrix} \epsilon \\ \eta \end{pmatrix}, \tag{20}$$

其中，$diag(\cdot)$表示的是由括号内元素构成的对角矩阵，对于激活函数σ，这里选取tanh函数，$\tanh' \in (0,1]$，因此我们要讨论$\mathbf{J}(t)$的特征值，只需讨论K(t)的特征值，于是我们有如下定理

定理3. 对于系统*(20)*,当$\tanh' \in (0,1]$，假设$\mathbf{K}(t)$关于时间t慢变，即$\mathbf{K}(t)=\mathbf{D}+\mathbf{B}(t)$，$\mathbf{D}$在复数域上可对角化，且满足$|1+h\lambda_i(t)+h| \le 1$, $b_{i1}(n)+b_{i2}(n)+\cdots+b_{in}(n)<\eta$，其中$\eta$是一个任意小量，那么*(20)*的正向欧拉格式是稳定的。

5. 总结与展望

本文给出一个反例以此来说明Haber等人关于离散模型的稳定性定理不严谨，给出了针对判断欧拉格式深度网络稳定性的新判据，并运用于一类哈密顿网络。在此基础上我们得到了一类哈密顿网络稳定性的判据，有了稳定性的基础，就可以展开后续对于这类哈密顿网络的优化控制问题的研究。同时，对于已经得出的稳定性定理，也还有值得讨论的地方，比如我们现在得出的只是网络的正向欧拉公式稳定的充分条件，能否给出更加完善的条件，再者能否将得到的稳定性定理运用到另外两类神经网络模型上等。

参考文献

[1] Weinan, E. (2017) A Proposal on Machine Learning via Dynamical Systems. *Communications in Mathematics and Statistics*, **5**, 1-11. https://doi.org/10.1007/s40304-017-0103-z

[2] Fan, J. and Gijbels, R. (1996) Local Polynomial Modelling and Its Applications. Chapman Hall, London.

[3] Hastie, T. (2004) The Elements of Statistical Learning: Data Mining. *Journal of the Royal Statistical Society*, **167**, 192. https://doi.org/10.1111/j.1467-985X.2004.298_11.x

[4] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[5] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vegas, NV, 27-30 June 2016, 770-778.

[6] Cho, K., Merrienboer, B.V., Gulcehre, C., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, October 2014, 1724-1734.

[7] Esteva, A., *et al.* (2017) Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*, **542**, 115-118. https://doi.org/10.1038/nature21056

[8] Chang, B., Meng, L., Haber, E., *et al.* (2017) Reversible Architectures for Arbitrarily Deep Residual Neural Networks. arXiv: 1709.03698

[9] Chang, B., Chen, M. and Haber, E. (2019) Antisymmetricrnn: A Dynamical System View on Recurrent Neural Networks. arXiv: 1902.09689

[10] Haber, E. and Ruthotto, L. (2018) Stable Architectures for Deep Neural Networks. *Inverse Problems*, **34**, Article 014004. https://doi.org/10.1088/1361-6420/aa9a90

[11] 林洁. 深度神经网络中几类动力系统的研究[D]: [硕士学位论文]. 金华: 浙江师范大学, 2021.