

一种新型单层递归神经网络解决非光滑伪凸优化问题

喻 昕^{①③} 卢惠霞^① 伍灵贞^{*②} 徐柳明^①

^①(广西大学计算机与电子信息学院 南宁 530004)

^②(桂林航天工业学院计算机科学与工程学院 桂林 541004)

^③(广西多媒体通信与网络技术重点实验室 南宁 530004)

摘 要: 非光滑伪凸优化问题是一类比较特殊的非凸优化问题, 常出现在各类科学与工程应用中, 因此具有很大的研究价值。针对现有神经网络模型解决非光滑伪凸优化问题存在的不足, 该文基于微分包含理论, 提出一种新型单层递归神经网络模型。通过理论分析, 证明了神经网络状态解在有限时间内收敛到可行域, 且永驻其中, 最终神经网络状态解收敛于原优化问题的最优解。最后, 通过数值实验, 验证了所提理论的有效性。与现有的神经网络相比, 该文所提神经网络模型结构简单仅为单层, 不需要提前计算罚参数, 且对初始点选取没有任何特殊的要求。

关键词: 神经网络; 非光滑伪凸优化; 收敛; 最优解

中图分类号: TP183

文献标识码: A

文章编号: 1009-5896(2021)08-2421-09

DOI: 10.11999/JEIT200558

A New One-layer Recurrent Neural Network for Solving Nonsmooth Pseudoconvex Optimization Problems

YU Xin^{①③} LU Huixia^① WU Lingzhen^② XU Liuming^①

^①(Department of Computer and Electronic Information, Guangxi University, Nanning 530004, China)

^②(School of Computer Science and Engineering, Guilin University of Aerospace Technology, Guilin 541004, China)

^③(Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China)

Abstract: Pseudoconvex optimization problems are a special kind of nonconvex optimization problems, which often appear in various scientific and engineering applications, so they have great research value. Considering the shortcomings of the existing neural network model to solve the nonsmooth pseudoconvex optimization problem, a new single-layer recurrent neural network model based on differential inclusion theory is proposed. Through theoretical analysis, it is proved that the state solution of the neural network converges to the feasible region within a limited time and stays in it forever. Finally, the state solution of the neural network converges to the optimal solution of the original optimization problem. At the end of the article, the validity of the proposed theory is verified through numerical experiments. Compared with existing neural networks, the neural network model proposed in this paper is simple in structure, does not need to calculate penalty parameters in advance, and has no special requirements for the selection of initial points.

Key words: Neural network; Nonsmooth pseudoconvex optimization; Convergence; Optimal solution

1 引言

优化问题常出现在机器学习、图像识别等各类科学和工程应用中。研究学者在非线性规划、微分包含等理论的基础上, 提出了较多用于解决优化问题的神经网络模型。1986年, Tank和Hopfield^[1]将Hopfield神经网络运用到解决线性规划问题中, 研

究学者由此得到启发, 相继提出各种用于解决优化问题的神经网络模型。1988年, Kennedy和Chua^[2]提出了一种包含罚参数的动态非线性规划电路(Nonlinear Programming Circuit, NPC)用于解决非线性规划问题的神经网络模型, 该模型中含有罚参数, 且要求目标函数和约束函数都是光滑的。随着NPC的引入, 优化问题的研究得到更进一步的发展, 由此之后, 研究学者提出了大量的神经网络模型用于解决优化问题。Xue等人^[3]基于次梯度方法, 提出了一种用于解决非光滑凸优化问题的单层递归神经网络模型, 该模型需要计算精确的罚参数, 但精确罚参数的计算与选取, 有时候非常困难,

收稿日期: 2020-07-07; 改回日期: 2020-12-06; 网络出版: 2020-12-16

*通信作者: 伍灵贞 327467000@qq.com

基金项目: 国家自然科学基金(61862004, 61462006)

Foundation Items: The National Natural Science Foundation of China (61862004, 61462006)

且计算量会很大。同样是解决文献[3]中的非光滑凸优化问题，Qin等人^[4]基于Tikhonov正则化方法，提出了一种动态神经网络模型，该模型不要求可行域是有界的或者目标函数是强制的。后来，Qin等人^[5]基于微分包含理论，提出了一种用于解决这类非光滑凸优化问题的双层神经网络模型，该模型不需要计算精确的罚参数，且适当放宽假设条件，遗憾的是，网络结构为双层，相对其他单层神经网络增加了模型复杂度。

随着凸优化问题研究越来越成熟，研究学者发现不能利用上述的神经网络解决科学和工程应用中常见的非凸优化问题。又因为伪凸优化问题是一类比较特殊的非凸优化问题，所以，研究伪凸优化问题具有很大的理论指导意义。于是，Li等人^[6]基于正则化方法，提出一种用于解决伪凸优化问题的神经网络模型。但它要求可行域有界，且初始点必须在可行域中。文献[7]基于惩罚函数思想，提出了一种用于解决包含等式约束和方体约束的非光滑伪凸优化问题的神经网络模型，但该模型要求计算罚参数。为避免罚参数的计算，Hosseini^[8]基于微分包含理论，提出了一种神经网络模型，可是它只能解决包含不等式约束的非凸优化问题，因此该模型的应用范围受到限制。为更好地解决伪凸优化问题，Qin等人^[9]提出了一种单层神经网络模型，该模型结构简单，不需要罚参数，并且能解决包含等式约束和不等式约束的非光滑伪凸优化问题。Bian等人^[10]基于光滑思想理论，提出了一种用于解决非光滑伪凸优化问题的神经网络模型，但要求初始点的选取必须在等式可行域内，由此可知，该模型在实际应用的范围会受到限制。近年来，也有研究者运用神经网络解决各类具体的优化问题，如高鑫等人^[11]基于端到端的神经网络模型以提高车辆密集区域的检测精度。

根据上述已有的神经网络模型以及前人的工作^[12-16]，本文提出了一种新型单层递归神经网络模型，用于解决包含等式约束和不等式约束的非光滑伪凸优化问题。与现有的神经网络相比，本文所提出的神经网络其优势有：(1)不同于文献[5,17]，本文的神经网络结构简单仅为单层；(2)与文献[3,18,19]不同，本文的神经网络不需要计算精确罚参数；(3)与文献[5-7,10,17,18,20]不同，本文的神经网络在初始点的选取上，没有任何特殊要求；(4)能解决更一般的优化问题；(5)与其他论文不同，本文直接证明神经网络状态解进入可行域，而大多数的神经网络会先证明进入等式约束集，再证明进入不等式约束集。

本文组织结构如下：第2节介绍研究内容和相关定义；第3节介绍神经网络模型；第4节对该神经网络模型进行理论分析；第5节通过仿真实验验证理论结论的正确性；第6节总结全文。

2 研究内容和相关定义

2.1 主要研究内容

本文主要研究非光滑伪凸优化问题，具体描述为

$$\begin{cases} \min f(\mathbf{x}) \\ \text{s.t. } g(\mathbf{x}) \leq 0 \\ \mathbf{Ax} = \mathbf{b} \end{cases} \quad (1)$$

其中， $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ ， $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是正则伪凸函数， $\mathbf{g} = (g_1, g_2, \dots, g_p)^T: \mathbb{R}^n \rightarrow \mathbb{R}^p$ 是凸函数， $\mathbf{A} \in \mathbb{R}^{n \times m}$ 是行满秩矩阵， $\mathbf{b} = (b_1, b_2, \dots, b_m)^T \in \mathbb{R}^m$ 。定义等式约束集 $\mathbf{S}_1 = \{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax} = \mathbf{b}\}$ ，不等式约束集 $\mathbf{S}_2 = \{\mathbf{x} \in \mathbb{R}^n: \mathbf{g}(\mathbf{x}) \leq 0\}$ ，可行域 $\mathbf{S} = \mathbf{S}_1 \cap \mathbf{S}_2 = \{\mathbf{x} \in \mathbb{R}^n: \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{Ax} = \mathbf{b}\}$ 。本文中出现的 $\text{int}(\cdot)$ 表示内部， $bd(\cdot)$ 表示边界上。假设优化问题式(1)有解。

2.2 相关定义

接下来，列出与本文相关的部分必要的定义，使读者更好地理解本文研究内容。

定义1^[10] 如果对于集合 $\mathbf{E} \subset \mathbb{R}^n$ 上的任意 \mathbf{x} ，都存在一个对应的非空集合 $\mathbf{F}(\mathbf{x}) \subset \mathbb{R}^n$ ，那么称 $\mathbf{x} \rightarrow \mathbf{F}(\mathbf{x})$ 是 $\mathbf{E} \rightarrow \mathbb{R}^n$ 上的集值映射。如果对于任意的开集 $\mathbf{V} \supset \mathbf{F}(\mathbf{x}_0)$ ，对每个点 \mathbf{x}_0 ，都存在相应的邻域 \mathbf{U} ，使得 $\mathbf{F}(\mathbf{U}) \subset \mathbf{V}$ ，那么称集值映射 $\mathbf{F}: \mathbf{E} \rightarrow \mathbb{R}^n$ 在点 $\mathbf{x}_0 \in \mathbf{E}$ 处是上半连续的。

定义2^[21] 设 $\mathbf{E} \subset \mathbb{R}^n$ 是一个非空凸集合，如果对于任取的 $\mathbf{x}, \mathbf{y} \in \mathbf{E}$ ，有 $\eta \in \partial f(\mathbf{x}), \eta^T(\mathbf{y} - \mathbf{x}) \geq 0$ ，满足 $f(\mathbf{y}) \geq f(\mathbf{x})$ ，那么称 $f: \mathbf{E} \subset \mathbb{R}$ 在集合 \mathbf{E} 中是伪凸函数。

定义3^[3] 链式法则。如果 $V: \mathbb{R}^n \rightarrow \mathbb{R}$ 在任意的 \mathbf{x} 处是一个正则函数，并且在 t 附近， $\mathbf{x}: \mathbb{R} \rightarrow \mathbb{R}^n$ 是Lipschitz且可微的，那么对于几乎所有的时间 $t \in [0, +\infty)$ ，有 $(d/dt)V(\mathbf{x}(t)) = \langle \xi, \dot{\mathbf{x}}(t) \rangle$ ， $\forall \xi \in V(\mathbf{x}(t))$ 。

3 定义神经网络模型

本节提出一种单层神经网络模型解决非光滑伪凸优化问题式(1)。本文中，需要给出下面的假设。

假设1 存在 $\hat{\mathbf{x}} \in \mathbb{R}^n, R > 0$ ，满足 $\hat{\mathbf{x}} \in \text{int}(\mathbf{S}_2) \cap \mathbf{S}_1$ ，使得 $\mathbf{S}_2 \subset B(\hat{\mathbf{x}}, R), \hat{g} = \min_{1 \leq i \leq p} (-g_i(\hat{\mathbf{x}})) > 0$ ，其中 $B(\hat{\mathbf{x}}, R) = \{\hat{\mathbf{x}} \in \mathbb{R}^n: \|\mathbf{x} - \hat{\mathbf{x}}\| \leq R\}$ 。

本文基于惩罚函数的方法，为不等式约束 $g_i(\mathbf{x}), i = 0, 1, \dots, p$ 定义如式(2)的惩罚函数

$$G(\mathbf{x}) = \sum_{i=1}^p \max\{0, g_i(\mathbf{x})\} \quad (2)$$

因为 $g_i(\mathbf{x}), i = 0, 1, \dots, p$ 是凸函数，所以 $G(\mathbf{x})$ 也是凸函数。对所有的 $\mathbf{x} \in \mathbb{R}^n$ ，都存在对应的 $\partial G(\mathbf{x})$ ，且

$$\partial G(\mathbf{x}) = \begin{cases} \left\{ \sum_{I \in I^+(\mathbf{x})} \partial g_i(\mathbf{x}) + [0, 1] \sum_{I \in I^0(\mathbf{x})} \partial g_i(\mathbf{x}) \right\}, & \mathbf{x} \notin S_2 \\ \{0\}, & \mathbf{x} \in \text{int}(S_2) \\ \left\{ [0, 1] \sum_{I \in I^0(\mathbf{x})} \partial g_i(\mathbf{x}) \right\}, & \mathbf{x} \in \text{bd}(S_2) \end{cases} \quad (3)$$

其中， $I^+(\mathbf{x}) = \{i \in \{1, 2, \dots, m\} : g_i(\mathbf{x}) > 0\}$ ， $I^0(\mathbf{x}) = \{i \in \{1, 2, \dots, m\} : g_i(\mathbf{x}) = 0\}$ 。

本文中定义 $L(\mathbf{x}) = \left\| \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x} - \mathbf{b}) \right\|$ ，得 $\{\mathbf{x} : L(\mathbf{x}) \leq 0\} = \{\mathbf{x} : \mathbf{A}\mathbf{x} - \mathbf{b} = 0\} = S_1$ 。

为解决包含等式约束和不等式约束条件的伪凸优化问题式(1)，本文基于上述理论知识的储备，提出如下的神经网络模型进行求解

$$\dot{\mathbf{x}} \in - \frac{\partial f(\mathbf{x})}{\max\{\|\partial f(\mathbf{x})\|, 1\}} - \alpha \cdot \max\left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} \cdot (\partial G(\mathbf{x}) + \partial L(\mathbf{x})) \quad (4)$$

其中， $\partial f(\mathbf{x}), \partial G(\mathbf{x}), \partial L(\mathbf{x})$ 分别表示 $f(\mathbf{x}), G(\mathbf{x}), L(\mathbf{x})$ 的 Clark 广义次梯度， $\alpha > 1$ 为常量， $\hat{\mathbf{x}}, \hat{g}$ 定义在假设 1 中。式(4)是神经网络的状态方程，本文所提神经网络式(4)可以用电路实现，其具体构造及工作过程见图 1。

由表 1 可知：(1) 本文所提神经网络模型，结构为单层，而文献[5,17]的结构分别是 2 层、3 层；(2) 本文不需要提前计算精确的罚参数，而文献[3,18,19]需要；(3) 本文在选取初始点时可任意选取，而文献[5,6,10,17-19]中初始点的选取受约束。因此，本文的模型在解决伪凸优化问题方面具有一定的优越性。

4 主要定理

本节使用理论分析的方法对神经网络式(4)的有效性进行验证。在理论分析过程中，证明了一些必要的定理。首先证明了本文所提神经网络式(4)存在局部解，其次证明了全局解存在性，然后证明了神经网络式(4)的状态解可在有限时间内到达可行域 S ，且一直驻留其中。最后，在前面分析的基础上，讨论了神经网络式(4)的收敛性，且证明了神经网络式(4)收敛于伪凸优化问题式(1)的最优解。

在证明之前，需要给出一些必要的引理。这里定义 $\mathbf{P} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}, \mathbf{c} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$ 。

引理 1 若假设 1 成立，则下面的结论成立。

- (1) 若 $\mathbf{x} \notin S_1$ ，则 $\|\nabla L(\mathbf{x})\| = 1$ ；
- (2) 若 $\mathbf{x} \in S_1$ ，则 $\|\nabla L(\mathbf{x})\| \leq 1$ ；
- (3) $\langle l(\mathbf{x}), \mathbf{x} - \hat{\mathbf{x}} \rangle = \begin{cases} 0, & \mathbf{x} \in S_1 \\ L(\mathbf{x}), & \mathbf{x} \notin S_1 \end{cases}, \forall l(\mathbf{x}) \in \partial L(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n$ 。

证明 (1) 根据 $L(\mathbf{x})$ 的定义，可知：当 $\mathbf{x} \notin S_1$ ， $L(\mathbf{x})$ 是严格可微函数，并且

$$\nabla L(\mathbf{x}) = \frac{\mathbf{P}^T \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} (\mathbf{A}\mathbf{x} - \mathbf{b})}{\left\| \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} (\mathbf{A}\mathbf{x} - \mathbf{b}) \right\|} \quad (5)$$

则可得

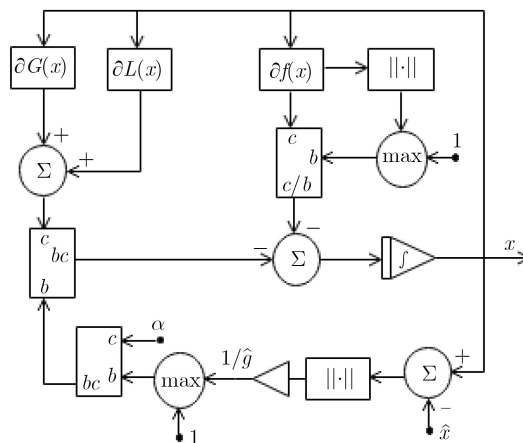


图 1 神经网络式(4)的电路实现图

表 1 与现有神经网络对比

参考文献	层次结构	神经元个数	有无精确罚因子	初始点是否为任意点
文献[3]	1	n	有	是
文献[5]	2	$n+m$	无	否
文献[6]	1	n	无	否
文献[10]	1	n	无	否
文献[17]	3	$n+m+p$	无	否
文献[18]	1	n	有	否
文献[19]	1	n	有	否
本文	1	n	无	是

$$\|\nabla L(\mathbf{x})\| = \frac{\left\| \mathbf{P}^T \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b}) \right\|}{\left\| \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b}) \right\|} = 1 \quad (6)$$

(2) 当 $\mathbf{x} \in \mathbf{S}_1$ 时, 可知 $\partial L(\mathbf{x}) = \{\mathbf{P}^T \boldsymbol{\zeta} : \|\boldsymbol{\zeta}\| \leq 1\}$, 则有 $\|\partial L(\mathbf{x})\| = \|\mathbf{P}^T \boldsymbol{\zeta}\| \leq \|\mathbf{P}\| \|\boldsymbol{\zeta}\| \leq 1$ 。(3) 当 $\mathbf{x} \in \mathbf{S}_1$ 时, 有 $\mathbf{P} \mathbf{x} - \mathbf{c} = \mathbf{P} \hat{\mathbf{x}} - \mathbf{c} = 0$ 。对于任意的 $\boldsymbol{\zeta} \in \mathbb{R}^n$ 且 $\|\boldsymbol{\zeta}\| \leq 1$, 则有 $\langle \mathbf{P}^T \boldsymbol{\zeta}, \mathbf{x} - \hat{\mathbf{x}} \rangle = \boldsymbol{\zeta}^T (\mathbf{P} \mathbf{x} - \mathbf{P} \hat{\mathbf{x}}) = 0$ 。当 $\mathbf{x} \notin \mathbf{S}_1$, 则有

$$\begin{aligned} & \langle \nabla L(\mathbf{x}), \mathbf{x} - \hat{\mathbf{x}} \rangle \\ &= (\mathbf{x} - \hat{\mathbf{x}})^T \frac{\mathbf{P}^T \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b})}{\left\| \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b}) \right\|} \\ &= \frac{\left\| \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b}) \right\|^2}{\left\| \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{x} - \mathbf{b}) \right\|} = L(\mathbf{x}) \quad (7) \end{aligned}$$

引理2^[18] 如果假设1成立, 对于任意的 $\mathbf{x} \notin \mathbf{S}_2$, $\boldsymbol{\eta} \in \partial G(\mathbf{x})$, 有 $(\mathbf{x} - \hat{\mathbf{x}})^T \boldsymbol{\eta} \geq \hat{g}$, 其中 \hat{g} 和 $\hat{\mathbf{x}}$ 同假设1。

引理3 如果假设1成立, 对于任意的 $\mathbf{x}_0 \in \mathbb{R}^n \notin \mathbf{S}_2$, 及任意满足 $\mathbf{x}(t) \notin \mathbf{S}_2, t \in [0, +\infty)$, 则存在一个正数 $\varepsilon_1 > 0$, 使得式(8)成立

$$\frac{d}{dt} \left(\frac{1}{2} \|\mathbf{x}(t) - \hat{\mathbf{x}}\|^2 \right) \leq -\varepsilon_1 \quad (8)$$

证明 对于几乎所有的时间 $t \in [0, +\infty)$, 存在可测函数 $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, $\gamma(t) \in \partial f(\mathbf{x}(t))$, 根据链式法则, 可得

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2} \|\mathbf{x}(t) - \hat{\mathbf{x}}\|^2 \right) = \langle \mathbf{x}(t) - \hat{\mathbf{x}}, \dot{\mathbf{x}}(t) \rangle \\ &= (\mathbf{x}(t) - \hat{\mathbf{x}})^T \left(-\frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \right. \\ & \quad \left. - \alpha \cdot \max \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \right) \\ &\leq \frac{\|\mathbf{x}(t) - \hat{\mathbf{x}}\| \|\gamma(t)\|}{\max\{\|\gamma(t)\|, 1\}} - \alpha \cdot \max \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} \\ & \quad \cdot (\mathbf{x}(t) - \hat{\mathbf{x}})^T \boldsymbol{\eta}(t) - \alpha \cdot \max \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} \\ & \quad \cdot (\mathbf{x}(t) - \hat{\mathbf{x}})^T \mathbf{l}(t) \quad (9) \end{aligned}$$

因为 $\mathbf{x}(t) \notin \mathbf{S}_2$, 根据引理1, 可得 $(\mathbf{x}(t) - \hat{\mathbf{x}})^T \mathbf{l}(t) \geq 0$ 。结合式(9)和引理2, 可有

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \|\mathbf{x}(t) - \hat{\mathbf{x}}\|^2 \right) &\leq \frac{\|\mathbf{x}(t) - \hat{\mathbf{x}}\| \|\gamma(t)\|}{\max\{\|\gamma(t)\|, 1\}} \\ & \quad - \alpha \cdot \max \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} \hat{g} \\ &\leq \|\mathbf{x}(t) - \hat{\mathbf{x}}\| - \alpha \\ & \quad \cdot \max \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1 \right\} \hat{g} \leq -\varepsilon_1 \quad (10) \end{aligned}$$

这意味着, 若 $\mathbf{x}(t) \notin \mathbf{S}_2$, 且 \mathbf{S}_2 有界, 那么必然存

在一个足够大的 R , 满足 $\|\mathbf{x}(t) - \hat{\mathbf{x}}\| \leq R, t \in [0, +\infty)$ 。也就是说, 神经网络状态解 $\mathbf{x}(t)$ 有界。证毕

4.1 神经网络存在局部解

定理1 如果假设1成立, 对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, 在区间 $t \in [0, T)$ 上, 神经网络式(4)至少存在一个局部解 $\mathbf{x}(t)$, 其中 $T > 0$ 。

证明 通过分析可知, 神经网络式(4)的右侧函数是一个非空紧凸的上半连续集值映射函数, 那么根据文献[22, Th.1, p.77], 则可得, 对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, 神经网络式(4)至少存在一个局部解 $\mathbf{x}(t)$, 其中 $t \in [0, T)$, 且 T 是这个解的最大存在区间。证毕

4.2 全局解存在性

由于局部解存在局限性, 为了更好地说明本文提出的神经网络具有良好的性质, 下面将求证神经网络式(4)全局解的存在性。

定理2 如果假设1成立, 则对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, 神经网络式(4)存在全局解 $\mathbf{x}(t)$, 其中 $t \in [0, +\infty)$ 。

证明 如果假设1成立, 可知对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, \mathbf{S}_2 有界。那么根据类似于引理3的证明, 可证明 $\mathbf{x}(t)$ 有界。则根据定理1和解的可扩展性理论, 可知神经网络式(4)至少存在一个全局解 $\mathbf{x}(t)$, 其中 $t \in [0, +\infty)$ 。证毕

引理4^[22] 设 $\mathbf{x}(t)$ 是神经网络式(4)的一个全局解。如果存在一个函数 $V(\mathbf{x}(t)) : \mathbb{R}^n \rightarrow \mathbb{R}$, 且在时间 $t \in [0, +\infty)$ 上是绝对连续的, 那么对于几乎所有的时间 $t \in [0, +\infty)$, 如果能满足 $\mathbf{x}(t) \in \{\mathbf{x} : V(\mathbf{x}) > 0\}$, 那么必定会存在常量 $\varepsilon > 0$, 使得 $(d/dt)V(\mathbf{x}(t)) \leq -\varepsilon$, 则神经网络(4)的状态解 $\mathbf{x}(t)$, 可在有限时间进到 $\{\mathbf{x} : V(\mathbf{x}) \leq 0\}$ 的范围里, 且一直驻留其中。

4.3 有限时间内收敛到S

这里求证神经网络式(4)的状态解 $\mathbf{x}(t)$ 有限时间内收敛到可行域 \mathbf{S} 。本文是直接求证神经网络式(4)的状态解 $\mathbf{x}(t)$ 在有限时间内进入可行域 \mathbf{S} , 且一直驻留其中, 而不是按照以往的思维, 先求证神经网络式(4)的状态解 $\mathbf{x}(t)$ 在有限时间内进入等式约束集 \mathbf{S}_1 , 且一直驻留其中, 然后, 求证神经网络式(4)的状态解 $\mathbf{x}(t)$ 在有限时间内进入不等式约束集 \mathbf{S}_2 , 且一直驻留其中。

定理3 如果假设1成立, 则对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, 神经网络式(4)的状态解 $\mathbf{x}(t)$ 可在有限时间内进到可行域 \mathbf{S} , 且一直驻留其中。

证明 在研究动态系统的稳定性过程中, Lyapunov能量函数的作用很关键。所以定义如下的能量函数 $D(\mathbf{x}) = G(\mathbf{x}) + L(\mathbf{x})$ 。显然, $\{\mathbf{x} \in \mathbb{R}^n : G(\mathbf{x}) \leq 0, L(\mathbf{x}) = 0\} = \{\mathbf{x} \in \mathbb{R}^n : D(\mathbf{x}) \leq 0\} = \mathbf{S}$ 。因为 $\{\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}\} = \{\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}_2\} \cup \{\mathbf{x}(t) \in \mathbf{S}_2 \setminus \mathbf{S}_1\}$,

所以考虑 $\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}$ 外的任一点 $\mathbf{x}(t)$, 仅需要 $\{\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}_2\}$ 与 $\{\mathbf{x}(t) \in \mathbf{S}_2 \setminus \mathbf{S}_1\}$ 这两个集合中的点在有限时间内进到可行域 \mathbf{S} , 且一直驻留其中。因此, 定理3可分两种情形讨论。

情形1 $\{\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}_2\}$ 对于几乎所有的时间 $t \in [0, +\infty)$, 存在可测函数 $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, 根据引理1和引理2, 则有 $\langle \boldsymbol{\eta}(t) + \mathbf{l}(t), \mathbf{x}(t) - \hat{\mathbf{x}} \rangle \geq \hat{g}$ 。由此可知 $\langle \boldsymbol{\eta}(t) + \mathbf{l}(t), \mathbf{x}(t) - \hat{\mathbf{x}} \rangle \geq \hat{g}$, 即 $\|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| \|\mathbf{x}(t) - \hat{\mathbf{x}}\| \geq \hat{g}$, 再结合引理3中的 $\|\mathbf{x}(t) - \hat{\mathbf{x}}\| \leq R$, 可得 $\|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| \geq \hat{g} / \|\mathbf{x}(t) - \hat{\mathbf{x}}\| \geq \hat{g} / R$, 由前文知 $\hat{g} / R > 0$, $(1 - \alpha) < 0$, 所以 $\|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| (1 - \alpha) < 0$ 。对于几乎所有的时间 $t \in [0, +\infty)$, 存在可测函数 $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, $\gamma(t) \in \partial f(\mathbf{x}(t))$, 根据链式法则可知

$$\begin{aligned} \frac{d}{dt} D(\mathbf{x}(t)) &= \langle \boldsymbol{\eta}(t) + \mathbf{l}(t), \dot{\mathbf{x}}(t) \rangle \\ &= \langle \boldsymbol{\eta}(t) + \mathbf{l}(t), -\frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \rangle \\ &\leq \|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| \frac{\|\gamma(t)\|}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} \|\boldsymbol{\eta}(t) + \mathbf{l}(t)\|^2 \\ &\leq \|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| \\ &\quad \cdot \left(1 - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} \frac{\hat{g}}{\|\mathbf{x}(t) - \hat{\mathbf{x}}\|}\right) \\ &\leq \|\boldsymbol{\eta}(t) + \mathbf{l}(t)\| (1 - \alpha) \leq -\delta_1 \quad (11) \end{aligned}$$

其中, $\delta_1 > 0$ 是一个正数。

情形2 $\{\mathbf{x}(t) \in \mathbf{S}_2 \setminus \mathbf{S}_1\}$ 类似情形1的证法, 由前文知 $\alpha > 1$, $\max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} > 1$, 则 $1 - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} < 0$ 。对于几乎所有的时间 $t \in [0, +\infty)$, 存在可测函数 $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, $\gamma(t) \in \partial f(\mathbf{x}(t))$, 根据链式法则, 知

$$\begin{aligned} \frac{d}{dt} D(\mathbf{x}(t)) &= \langle \boldsymbol{\eta}(t) + \mathbf{l}(t), \dot{\mathbf{x}}(t) \rangle \\ &= \langle \boldsymbol{\eta}(t) + \mathbf{l}(t), -\frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \rangle \\ &\leq \|\mathbf{l}(t)\| \frac{\|\gamma(t)\|}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} \|\mathbf{l}(t)\|^2 \\ &\leq 1 - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} \leq -\delta_2 \quad (12) \end{aligned}$$

其中, $\delta_2 > 0$ 是一个正数。这里令 $\delta = \min\{\delta_1, \delta_2\}$, 显然对于所有的 $\mathbf{x}(t) \in \mathbb{R}^n \setminus \mathbf{S}$, 有 $(d/dt)D(\mathbf{x}(t)) \leq -\delta$ 。结合引理4, 可知, 神经网络式(4)的状态解 $\mathbf{x}(t)$ 可在有限时间内进到可行域 \mathbf{S} , 且一直驻留其中。证毕

4.4 神经网络收敛到最优解

定理4 如果假设1成立, 则对于任取的初始点 $\mathbf{x}_0 \in \mathbb{R}^n$, 神经网络式(4)的状态解 $\mathbf{x}(t)$ 可收敛到伪凸优化问题式(4)的最优解集 \mathbf{M} 。

证明 根据定理3可知, 神经网络式(4)的状态解 $\mathbf{x}(t)$ 能够在有限时间内进到可行域 \mathbf{S} , 且一直驻留其中。假设任取的 $\mathbf{x}_0 \in \mathbf{S}$ 且 $\mathbf{x}(t) \in \mathbf{S}$, 记 $P_{\mathbf{M}}(\mathbf{x})$ 为 $\mathbf{x}(t)$ 由 \mathbb{R}^n 到 \mathbf{M} 的投影算子。对于几乎所有的时间 $t \in [0, +\infty)$, 存在 $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, $\gamma(t) \in \partial f(\mathbf{x}(t))$, 根据链式法则, 可知

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \text{dist}^2(\mathbf{x}(t), \mathbf{M}) &= \langle \mathbf{x}(t) - P_{\mathbf{M}}(\mathbf{x}(t)), \dot{\mathbf{x}}(t) \rangle \\ &= \langle \mathbf{x}(t) - P_{\mathbf{M}}(\mathbf{x}(t)), -\frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \rangle \quad (13) \end{aligned}$$

因为 $G(\mathbf{x})$ 和 $L(\mathbf{x})$ 都具有凸性, $\boldsymbol{\eta}(t) \in \partial G(\mathbf{x}(t))$, $\mathbf{l}(t) \in \partial L(\mathbf{x}(t))$, $\gamma(t) \in \partial f(\mathbf{x}(t))$ 则可得

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \text{dist}^2(\mathbf{x}(t), \mathbf{M}) &\leq -\frac{1}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad \langle \mathbf{x}(t) - P_{\mathbf{M}}(\mathbf{x}(t)), \gamma(t) \rangle \quad (14) \end{aligned}$$

接下来, 本文将分3种情形讨论分析。为方便下文的理论分析, 这里给出两个集合的定义

$$\begin{aligned} \mathbf{I} &= \{t \in [0, +\infty) : \langle \mathbf{x}(t) - P_{\mathbf{M}}(\mathbf{x}(t)), \gamma(t) \rangle \leq 0\}, \\ \mathbf{J} &= \{t \in [0, +\infty) : \langle \mathbf{x}(t) - P_{\mathbf{M}}(\mathbf{x}(t)), \gamma(t) \rangle > 0\} \quad (15) \end{aligned}$$

情形1 存在 $T \geq 0$, 满足 $t \in \mathbf{I}, \forall t \in [T, +\infty)$ 。这里令 $\bar{\mathbf{x}}$ 是 $\mathbf{x}(t)$ 中的一个簇点, 存在 $\gamma(t) \in \partial f(\bar{\mathbf{x}}(t))$, 可满足 $\langle \bar{\mathbf{x}}(t) - P_{\mathbf{M}}(\bar{\mathbf{x}}(t)), \gamma(t) \rangle \leq 0$ 。由 $f(\mathbf{x})$ 的伪凸性, 可知 $f(\bar{\mathbf{x}}) \leq f(P_{\mathbf{M}}(\bar{\mathbf{x}}))$, 则有 $\bar{\mathbf{x}} \in \mathbf{M}$, 即神经网络式(4)任意 $\mathbf{x}(t)$ 的簇点都是伪凸优化问题式(1)的最优解。然后, 通过反证法证明 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0$ 。如果 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) = \varpi > 0$, 那么必定存在递增序列 $\{t_k\}$, 满足 $\lim_{k \rightarrow +\infty} t_k = +\infty$, 且 $\text{dist}(\mathbf{x}(t_k), \mathbf{M}) \geq \varpi/2$ 。又由定理2知 $\mathbf{x}(t)$ 有界, 则必定存在一个簇点 $\bar{\mathbf{x}}'$, 可以满足 $\text{dist}(\bar{\mathbf{x}}', \mathbf{M}) \geq \varpi/2$, 即 $\bar{\mathbf{x}}' \notin \mathbf{M}$ 。这与上文的证明, 任意 $\mathbf{x}(t)$ 的簇点都是伪凸优化问题式(1)的最优解, 相互矛盾。因此, 该假设不成立, 可推得 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0$ 。

情形2 存在 $T \geq 0$, 满足 $t \in \mathbf{J}, \forall t \in [T, +\infty)$, 即 $\langle \bar{\mathbf{x}}(t) - P_{\mathbf{M}}(\bar{\mathbf{x}}(t)), \gamma(t) \rangle > 0$ 。结合式(14), 可

知 $(d/dt)(1/2)\text{dist}^2(\mathbf{x}(t), \mathbf{M}) \leq 0$ 。因此可知 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M})$ 存在。若 $\liminf_{t \rightarrow +\infty} \langle \bar{\mathbf{x}}(t) - P_{\mathbf{M}}(\bar{\mathbf{x}}(t)), \gamma(t) \rangle = h > 0$ ，则一定会存在 $T_1 \geq T$ ，满足 $\langle \bar{\mathbf{x}}(t) - P_{\mathbf{M}}(\bar{\mathbf{x}}(t)), \gamma(t) \rangle \geq h/2, \forall t \in [T_1, +\infty)$ 。结合式(14)，可知

$$\frac{d}{dt} \frac{1}{2} \text{dist}^2(\mathbf{x}(t), \mathbf{M}) \leq -\frac{h}{2 \max\{\|\gamma(t)\|, 1\}} \quad (16)$$

对式(16)从 t 到 T_1 积分，可得

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \text{dist}^2(\mathbf{x}(t), \mathbf{M}) - \frac{d}{dt} \frac{1}{2} \text{dist}^2(\mathbf{x}(T_1), \mathbf{M}) \\ \leq -\frac{h}{2 \max\{\|\gamma(t)\|, 1\}} (t - T_1) \end{aligned} \quad (17)$$

令 $t \rightarrow +\infty$ ，可得 $\liminf_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) > -\infty$ ，这同 $\text{dist}(\mathbf{x}(t), \mathbf{M}) > 0$ 相矛盾，所以 $\liminf_{t \rightarrow +\infty} \langle \bar{\mathbf{x}}(t) - P_{\mathbf{M}}(\bar{\mathbf{x}}(t)), \gamma(t) \rangle = 0$ 。和情形1的分析证明相类似，由前文的证明可知 $\mathbf{x}(t)$ 有界，则存在单调递增序列 $\{t_k\}$ 和 $\bar{\mathbf{x}}$ ，使得 $\lim_{k \rightarrow +\infty} t_k = +\infty$ ，且 $\lim_{k \rightarrow +\infty} t_k = \bar{\mathbf{x}}$ 。结合式(17)，对于几乎所有时间 $t \in [0, +\infty)$ ，存在 $\gamma(t) \in \partial f(\bar{\mathbf{x}}(t))$ ，使满足 $\langle \mathbf{x}(t_k) - P_{\mathbf{M}}(\mathbf{x}(t_k)), \gamma(t_k) \rangle = 0$ 。由 $f(\mathbf{x})$ 的伪凸性，可知 $f(\bar{\mathbf{x}}) \leq f(P_{\mathbf{M}}(\bar{\mathbf{x}}))$ 。再由 $\bar{\mathbf{x}} \in \mathbf{M}$ ，可得 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t_k), \mathbf{M}) = 0$ 。因此， $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0$ 。

情形3 当集合 \mathbf{I} 和 \mathbf{J} 都为无界集合时，对于 $t \in \mathbf{I}$ ，证明方法与情形1的相似，则可证得

$$\lim_{t \rightarrow +\infty, t \in \mathbf{I}} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0 \quad (18)$$

然后，对于 $t \in \mathbf{J}$ ，这里定义 $\tau(t) = \sup_{s \leq t, s \in \mathbf{I}} s$ ，则 $\tau(t) \in \mathbf{I}$ ，且 $(\tau(t), t] \subseteq \mathbf{J}$ 。通过集合 \mathbf{I} 和 \mathbf{J} 的无界性，可得到 $\lim_{t \rightarrow +\infty, t \in \mathbf{J}} \tau(t) = +\infty$ 。结合 $(d/dt)(1/2)\text{dist}^2(\mathbf{x}(t), \mathbf{M}) \leq 0$ ，得

$$\text{dist}(\mathbf{x}(t), \mathbf{M}) \leq \text{dist}(\mathbf{x}(\tau(t)), \mathbf{M}), \forall t \in \mathbf{J} \quad (19)$$

$\tau(t) \in \mathbf{I}$ ，结合式(12)和式(13)，可得 $\lim_{t \rightarrow +\infty, t \in \mathbf{J}} \sup \text{dist}(\mathbf{x}(t), \mathbf{M}) \leq \lim_{t \rightarrow +\infty, t \in \mathbf{J}} \text{dist}(\mathbf{x}(\tau(t)), \mathbf{M}) = 0$ 。因此 $\lim_{t \rightarrow +\infty, t \in \mathbf{J}} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0$ 。结合式(18)和式(19)，可有 $\lim_{t \rightarrow +\infty} \text{dist}(\mathbf{x}(t), \mathbf{M}) = 0$ 。证毕

引理5^[9] 令 \mathbf{x}^* 是伪凸优化问题(1)的一个最优解，则对于任意的 $\mathbf{x} \in \mathbf{S}_2 \cap \mathbf{S}_1$ 且 $\gamma \in \partial f(\mathbf{x})$ ，则有 $\langle \gamma, \mathbf{x} - \mathbf{x}^* \rangle \geq 0$ 。

定理5 如果假设1成立，则对于任取的 $\mathbf{x}_0 \in \mathbb{R}^n$ ，存在一个 $\mathbf{x}^* \in \mathbf{M}$ ，使得神经网络式(4)的状态解 $\mathbf{x}(t)$ 收敛于伪凸优化问题(1)的最优解 \mathbf{x}^* 。

证明 根据定理3，存在 $T \geq 0$ ，使得神经网络式(4)的状态解 $\mathbf{x}(t) \in \mathbf{S}$ ，其中 $\forall t \in [T, +\infty)$ 。令 \mathbf{x}^* 是神经网络式(4)状态解 $\mathbf{x}(t)$ 的一个簇点。由定理

4的证明过程，可知 $\mathbf{x}^* \in \mathbf{M}$ ，进而，存在递增序列 $\{\mathbf{x}(t_k)\}$ ，使得 $\lim_{k \rightarrow +\infty} \mathbf{x}(t_k) = \mathbf{x}^*$ 。和定理3的证明相似，对于几乎所有的时间 $t > T$ ，则有

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|\mathbf{x}(t) - \mathbf{x}^*\|^2 &= \langle \mathbf{x}(t) - \mathbf{x}^*, \dot{\mathbf{x}}(t) \rangle \\ &= \langle \mathbf{x}(t) - \mathbf{x}^*, -\frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \\ &\quad - \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \rangle \\ &= -\langle \mathbf{x}(t) - \mathbf{x}^*, \frac{\gamma(t)}{\max\{\|\gamma(t)\|, 1\}} \rangle \\ &\quad - \langle \mathbf{x}(t) - \mathbf{x}^*, \alpha \cdot \max\left\{\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\hat{g}}, 1\right\} (\boldsymbol{\eta}(t) + \mathbf{l}(t)) \rangle \\ &\leq -\frac{1}{\max\{\|\gamma(t)\|, 1\}} \langle \mathbf{x}(t) - \mathbf{x}^*, \gamma(t) \rangle \end{aligned} \quad (20)$$

结合引理5，可得 $(d/dt)(1/2)\|\mathbf{x}(t) - \mathbf{x}^*\|^2 \leq 0$ 。由此可知， $\|\mathbf{x}(t) - \mathbf{x}^*\|$ 是非增函数，那么存在 $\lim_{t \rightarrow +\infty} \|\mathbf{x}(t) - \mathbf{x}^*\|$ 。又因为 $\lim_{k \rightarrow +\infty} \mathbf{x}(t_k) = \mathbf{x}^*$ ，所以有 $\lim_{t \rightarrow +\infty} \mathbf{x}(t) = \mathbf{x}^*$ 。证毕

5 仿真实验

在本节中，为了能够有效验证所提出神经网络式(4)的有效性，通过两个仿真实验模拟神经网络式(4)解决最优化问题。仿真实验在MATLAB 2012a 平台进行。

实验1 目标函数为2次函数的伪凸优化问题

$$\begin{aligned} \min f(\mathbf{x}) &= \left. \begin{aligned} &\frac{\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{a}^T \mathbf{x} + a_0 \mathbf{x}}{\mathbf{c}^T \mathbf{x} + b_0} \\ \text{s.t. } &\mathbf{A} \mathbf{x} = \mathbf{b}, -1 \leq x_i \leq 2, i = 1, 2, 3, 4 \end{aligned} \right\} \quad (21) \end{aligned}$$

$$\text{其中， } \mathbf{Q} = \begin{bmatrix} 5 & -1 & 2 & 0 \\ -1 & 5 & -1 & 3 \\ 2 & -1 & 3 & 0 \\ 0 & 3 & 0 & 5 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \left\{ \begin{aligned} &a_0 = -2 \\ &b_0 = 5 \end{aligned} \right., \mathbf{A} = \begin{bmatrix} 2 & 1 & -1 & -1 \\ 1 & 0 & 2 & -2 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \mathbf{S}_1 \cap \mathbf{S}_2 = \mathbf{S}, \mathbf{S} = \{\mathbf{x} \in \mathbb{R}^4 : -1 \leq x_i \leq 2\}.$$

目标函数 $f(\mathbf{x})$ 是包含4个变量的2次函数，用一般的方法求解相对困难。通过计算可知， \mathbf{Q} 是对称的正定矩阵， $f(\mathbf{x})$ 在 $\mathbf{S} = \{\mathbf{x} \in \mathbb{R}^4 : \mathbf{c}^T \mathbf{x} + b_0 > 0\}$ 上是伪凸函数，且在不等式约束 $-1 \leq x_{1,2,3,4} \leq 2$ 中， $f(\mathbf{x})$ 依旧是伪凸函数。因此，上述式(21)问题能够用本文提出的神经网络式(4)解决。文献[20]提出了一种神经网络来解决式(21)这类型问题，但是文献[20]的神经网络只能解决线性等式约束的问

题。然而，本文所提神经网络式(4)可以解决包含等式约束和不等式约束的问题。

在本实验中取内点 $\hat{x} = (1, 1, 1, 1)^T$ ，步长为 10^{-4} ， $\hat{g} = 1$ ， $\alpha = 2$ 。通过前文理论知识分析可知，原函数具有最优解集。因此，本实验中随机选取10个不同的初始点(可在可行域中，也可不在)进行实验，图2展现了随机选取10个初始点的轨迹图，从图2中可以知道该神经网络的轨迹在有限的时间内收敛到原函数优化问题的最优解 $x^* = (1.4395, 1.8999, 0.7785, -0.9999)^T$ ，图3是10个随机初始点的最优解的目标函数值 $f(x^*) = 0.9699$ 。

实验2^[10] 目标函数为Guass函数的伪凸优化问题 高斯函数较特殊，其是局部Lipschitz连续且在 \mathbb{R}^n 上严格伪凸。本文的神经网络式(4)能解决这类问题

$$\left. \begin{aligned} \min \quad & f(x) = -\exp\left(-\sum_{i=1}^n \frac{x_i^2}{\delta_i^2}\right) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & Ax = b \end{aligned} \right\} \quad (22)$$

其中， $i = 1, 2, \dots, \delta_i \geq 0, i = 1, 2, \dots, n, A \in \mathbb{R}^{r \times n}, b \in \mathbb{R}^r$ 。本实验中，取 $n = 4, \delta = (1, 0.0625, 0.25, 1)^T, b = (3, 5.5, 10)^T, A = \begin{bmatrix} 2 & 0 & -1 & 0.5 \\ 0 & 3 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, g_1 = x_1^2 + x_2 - x_3 - x_4, g_2 = 3x_3^2 - x_4$ 。

文献[7,10,20,21]中，分别提出了4种神经网络来解决这类型的问题，但是文献[7,20]的神经网络

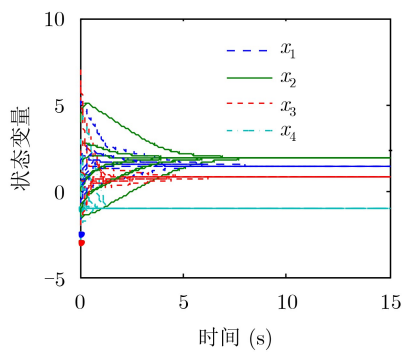


图2 实验1中10个随机初始点的收敛轨迹

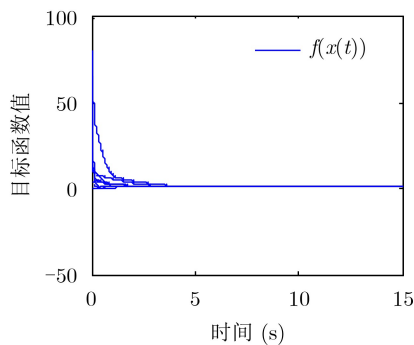


图3 实验1中10个随机初始点的目标函数值

只能解决带等式约束的问题，而文献[10,21]也能解决上述的问题，但是文献[21]需要计算罚参数，文献[10]要求初始点在等式可行域中。然而，本文所提神经网络式(4)能够克服上述文献中的不足。对于本实验中的问题，本文随机选取10个不同的初始点进行仿真实验，取内点 $\hat{x} = (1, 1, 1, 1)^T, \hat{g} = 1, \alpha = 3$ 。图4是随机选取10个初始点的神经网络式(4)的轨迹，不同初始点的神经网络其轨迹最终都能收敛到原始优化问题式(1)的最优解，最优解 $x^* = (0.3996, 3.4883, 0.5719, 5.5407)^T$ ，此时 $f(x^*) = 0$ 。为了更好地突出本文所提神经网络的优点，运用文献[10]的神经网络模型对伪凸优化问题式(22)进行同样的仿真实验，选取初始点 $(2, 3, 1, 0)^T$ 不在等式可行域内。文献[10]实验结果如图5所示，神经网络的轨迹不能得到很好的收敛。本文实验结果如图6所示，与文

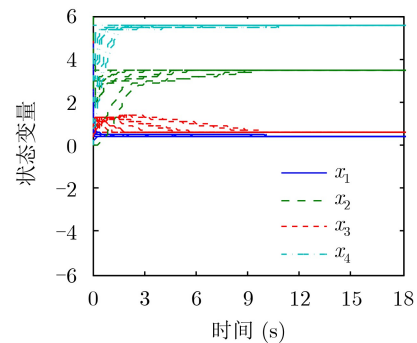


图4 实验2中10个随机初始点的收敛轨迹

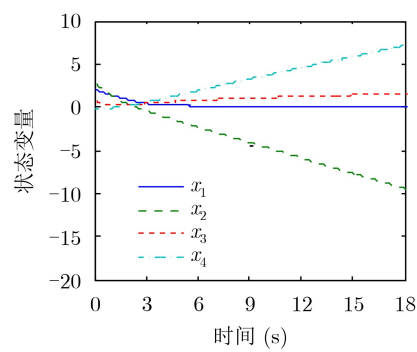


图5 文献[10]中初始点为 $(2, 3, 1, 0)^T$ 的收敛轨迹

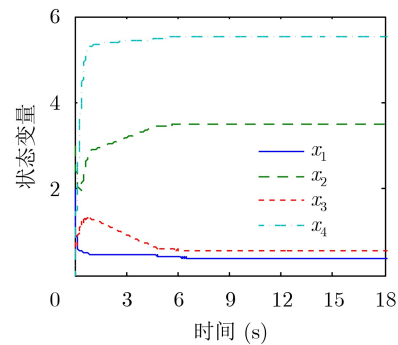


图6 实验2中初始点为 $(2, 3, 1, 0)^T$ 的收敛轨迹

献[10]的神经网络相比,神经网络式(4)对初始点的选取没有任何特殊要求。

6 结束语

本文提出一种新型单层递归神经网络模型以解决包含等式和不等式约束的非光滑伪凸优化问题。本文通过理论分析,首先证明了神经网络存在局部解。其次证明了神经网络具有全局解。在此分析的基础上,然后证明了神经网络状态解会在有限的时间内进到可行域中,且驻留其中。最后证明了神经网络最终收敛于原优化问题的最优解。此外,通过两个数值仿真实验,验证了理论结论的正确性,实验结果充分表明,本文提出的神经网络求解非光滑伪凸优化问题是有效的。与现有的神经网络模型相比,本文模型可以解决非光滑伪凸优化问题,具有结构简单、不需要计算罚参数,对初始点的选取没有任何特殊的要求等优点。

参考文献

- [1] TANK D W and HOPFIELD J. Simple ‘neural’ optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit[J]. *IEEE Transactions on Circuits and Systems*, 1986, 33(5): 533–541. doi: [10.1109/TCS.1986.1085953](https://doi.org/10.1109/TCS.1986.1085953).
- [2] KENNEDY M P and CHUA L O. Neural networks for nonlinear programming[J]. *IEEE Transactions on Circuits and Systems*, 1988, 35(5): 554–562. doi: [10.1109/31.1783](https://doi.org/10.1109/31.1783).
- [3] XUE Xiaoping and BIAN Wei. Subgradient-based neural networks for nonsmooth convex optimization problems[J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2008, 55(8): 2378–2391. doi: [10.1109/TCSI.2008.920131](https://doi.org/10.1109/TCSI.2008.920131).
- [4] QIN Sitian, FAN Dejun, WU Guangxi, et al. Neural network for constrained nonsmooth optimization using Tikhonov regularization[J]. *Neural Networks*, 2015, 63: 272–281. doi: [10.1016/j.neunet.2014.12.007](https://doi.org/10.1016/j.neunet.2014.12.007).
- [5] QIN Sitian and XUE Xiaoping. A two-layer recurrent neural network for nonsmooth convex optimization problems[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(6): 1149–1160. doi: [10.1109/TNNLS.2014.2334364](https://doi.org/10.1109/TNNLS.2014.2334364).
- [6] LI Qingfa, LIU Yaqiu, and ZHU Liangkuan. Neural network for nonsmooth pseudoconvex optimization with general constraints[J]. *Neurocomputing*, 2014, 131: 336–347. doi: [10.1016/j.neucom.2013.10.008](https://doi.org/10.1016/j.neucom.2013.10.008).
- [7] LIU Qingshan, GUO Zhishan, and WANG Jun. A one-layer recurrent neural network for constrained pseudoconvex optimization and its application for dynamic portfolio optimization[J]. *Neural Networks*, 2012, 26: 99–109. doi: [10.1016/j.neunet.2011.09.001](https://doi.org/10.1016/j.neunet.2011.09.001).
- [8] HOSSEINI A. A non-penalty recurrent neural network for solving a class of constrained optimization problems[J]. *Neural Networks*, 2016, 73: 10–25. doi: [10.1016/j.neunet.2015.09.013](https://doi.org/10.1016/j.neunet.2015.09.013).
- [9] QIN Sitian, YANG Xiudong, XUE Xiaoping, et al. A one-layer recurrent neural network for pseudoconvex optimization problems with equality and inequality constraints[J]. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3063–3074. doi: [10.1109/TCYB.2016.2567449](https://doi.org/10.1109/TCYB.2016.2567449).
- [10] BIAN Wei, MA Litao, QIN Sitian, et al. Neural network for nonsmooth pseudoconvex optimization with general convex constraints[J]. *Neural Networks*, 2018, 101: 1–14. doi: [10.1016/j.neunet.2018.01.008](https://doi.org/10.1016/j.neunet.2018.01.008).
- [11] 高鑫, 李慧, 张义, 等. 基于可变形卷积神经网络的遥感影像密集区域车辆检测方法[J]. *电子与信息学报*, 2018, 40(12): 2812–2819. doi: [10.11999/JEIT180209](https://doi.org/10.11999/JEIT180209).
GAO Xin, LI Hui, ZHANG Yi, et al. Vehicle detection in remote sensing images of dense areas based on deformable convolution neural network[J]. *Journal of Electronics & Information Technology*, 2018, 40(12): 2812–2819. doi: [10.11999/JEIT180209](https://doi.org/10.11999/JEIT180209).
- [12] LIU Na and QIN Sitian. A neurodynamic approach to nonlinear optimization problems with affine equality and convex inequality constraints[J]. *Neural Networks*, 2019, 109: 147–158. doi: [10.1016/j.neunet.2018.10.010](https://doi.org/10.1016/j.neunet.2018.10.010).
- [13] YU Xin, WU Lingzhen, XU Chenhua, et al. A novel neural network for solving nonsmooth nonconvex optimization problems[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(5): 1475–1488. doi: [10.1109/TNNLS.2019.2920408](https://doi.org/10.1109/TNNLS.2019.2920408).
- [14] LI Wenjing, BIAN Wei, and XUE Xiaoping. Projected neural network for a class of Non-Lipschitz optimization problems with linear constraints[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(9): 3361–3373. doi: [10.1109/TNNLS.2019.2944388](https://doi.org/10.1109/TNNLS.2019.2944388).
- [15] XU Chen, CHAI Yiyuan, QIN Sitian, et al. A neurodynamic approach to nonsmooth constrained pseudoconvex optimization problem[J]. *Neural Networks*, 2020, 124: 180–192. doi: [10.1016/j.neunet.2019.12.015](https://doi.org/10.1016/j.neunet.2019.12.015).
- [16] 喻昕, 许治健, 陈昭蓉, 等. 拉格朗日神经网络解决带等式和不等式约束的非光滑非凸优化问题[J]. *电子与信息学报*, 2017, 39(8): 1950–1955. doi: [10.11999/JEIT161049](https://doi.org/10.11999/JEIT161049).
YU Xin, XU Zhijian, CHEN Zhaorong, et al. Lagrange neural network for nonsmooth nonconvex optimization problems with equality and inequality constraints[J]. *Journal of Electronics & Information Technology*, 2017, 39(8): 1950–1955. doi: [10.11999/JEIT161049](https://doi.org/10.11999/JEIT161049).

- [17] CHENG Long, HOU Zengguang, LIN Yingzi, *et al.* Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks[J]. *IEEE Transactions on Neural Networks*, 2011, 22(5): 714–726. doi: [10.1109/TNN.2011.2109735](https://doi.org/10.1109/TNN.2011.2109735).
- [18] BIAN Wei and XUE Xiaoping. Subgradient-based neural networks for nonsmooth nonconvex optimization problems[J]. *IEEE Transactions on Neural Networks*, 2009, 20(6): 1024–1038. doi: [10.1109/TNN.2009.2016340](https://doi.org/10.1109/TNN.2009.2016340).
- [19] LIU Qingshan and WANG Jun. A one-layer recurrent neural network for constrained nonsmooth optimization[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2011, 41(5): 1323–1333. doi: [10.1109/TSMCB.2011.2140395](https://doi.org/10.1109/TSMCB.2011.2140395).
- [20] QIN Sitian, FAN Dejun, SU Peng, *et al.* A simplified recurrent neural network for pseudoconvex optimization subject to linear equality constraints[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2014, 19(4): 789–798. doi: [10.1016/j.cnsns.2013.08.034](https://doi.org/10.1016/j.cnsns.2013.08.034).
- [21] QIN Sitian, BIAN Wei, and XUE Xiaoping. A new one-layer recurrent neural network for nonsmooth pseudoconvex optimization[J]. *Neurocomputing*, 2013, 120: 655–662. doi: [10.1016/j.neucom.2013.01.025](https://doi.org/10.1016/j.neucom.2013.01.025).
- [22] AUBIN J P and CELLINA A. *Differential Inclusions*[M]. Berlin, Germany: Springer-Verlag, 1984: 77.
- 喻 昕: 男, 1973年生, 教授, 博士, 研究方向为人工神经网络、互联网络、优化计算.
- 卢惠霞: 女, 1993年生, 硕士生, 研究方向为神经网络、优化计算.
- 伍灵贞: 女, 1995年生, 硕士, 研究方向为神经网络、优化计算.
- 徐柳明: 男, 1994年生, 硕士生, 研究方向为神经网络、优化计算.
- 责任编辑: 余 蓉