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

喻　昕[①]　　许治健[*①]　　陈昭蓉[①]　　徐辰华[②]
[①](广西大学计算机与电子信息学院　南宁　530004)
[②](广西大学电气工程学院　南宁　530004)

摘　要：非凸非光滑优化问题涉及科学与工程应用的诸多领域，是目前国际上的研究热点。该文针对已有基于早期罚函数神经网络解决非光滑优化问题的不足，借鉴 Lagrange 乘子罚函数的思想提出一种有效解决带等式和不等式约束的非凸非光滑优化问题的递归神经网络模型。由于该网络模型的罚因子是变量，无需计算罚因子的初始值仍能保证神经网络收敛到优化问题的最优解，因此更加便于网络计算。此外，与传统 Lagrange 方法不同，该网络模型增加了一个等式约束惩罚项，可以提高网络的收敛能力。通过详细的分析证明了该网络模型的轨迹在有限时间内必进入可行域，且最终收敛于关键点集。最后通过数值实验验证了所提出理论的有效性。

关键词：拉格朗日神经网络；收敛；非凸非光滑优化
中图分类号：　TP183　　**文献标识码**：A　　**文章编号**：1009-5896(2017)08-1950-06
DOI: 10.11999/JEIT161049

Lagrange Neural Network for Nonsmooth Nonconvex Optimization Problems with Equality and Inequality Constrains

YU Xin[①]　　XU Zhijian[①]　　CHEN Zhaorong[①]　　XU Chenhua[②]
[①](*School of Computer, Electronics and Information, Guangxi University, Nanning* 530004, *China*)
[②](*School of Electrical Engineering, Guangxi University, Nanning* 530004, *China*)

Abstract: Nonconvex nonsmooth optimization problems are related to many fields of science and engineering applications, which are research hotspots. For the lack of neural network based on early penalty function for nonsmooth optimization problems, a recurrent neural network model is proposed using Lagrange multiplier penalty function to solve the nonconvex nonsmooth optimization problems with equality and inequality constrains. Since the penalty factor in this network model is variable, without calculating initial penalty factor value, the network can still guarantee convergence to the optimal solution, which is more convenient for network computing. Compared with the traditional Lagrange method, the network model adds an equality constraint penalty term, which can improve the convergence ability of the network. Through the detailed analysis, it is proved that the trajectory of the network model can reach the feasible region in finite time and finally converge to the critical point set. In the end, numerical experiments are given to verify the effectiveness of the theoretic results.

Key words: Lagrange neural network; Convergence; Nonsmooth nonconvex optimization

1　引言

作为解决优化问题的并行计算模型，递归神经网络在过去的几十年里受到了极大的关注，不少神经网络模型被提出[1–3]。然而这些网络都是为解决光滑优化问题而设计的，它们却无法解决非光滑优化问题。为此，Forti 等人[4]提出了通用非线性规划神经网络模型(G-NPC)，用于解决不等式限制的非光滑优化问题。G-NPC 是利用 Clark 次梯度和惩罚项构造的微分包含梯度系统，其动态行为和最优化能力适用于解决凸的和非凸的问题。Bian 等人在文献[5]和文献[6]分别提出了利用 Clark 次梯度神经网络来解决非光滑凸和非光滑非凸最优化问题，通过使用 Clark 次梯度和惩罚项建立微分包含递归神经网络模型。在此模型下给定一个足够大的罚因子参数，神经元状态轨迹将收敛到平衡点集。此外，Liu 等人[7]利用投影方法提出了解决线性等式和 R^n 上闭凸子集共同约束的非光滑非凸优化问题的递归神经网络模型。Bian 等人[8]利用光滑逼近技术构造光滑

收稿日期：2016-10-12；改回日期：2017-04-18；网络出版：2017-05-18
*通信作者：许治健　zhongxiawuyu@126.com
基金项目：国家自然科学基金(61462006, 51407037)，广西自然科学基金(2014GXNSFAA118391)
Foundation Items: The National Natural Science Foundation of China (61462006, 51407037), The Natural Science Foundation of Guangxi Province (2014GXNSFAA118391)

神经网络模型解决非光滑非凸的优化问题，在此模型下神经网络的平衡点就是原始问题的最优解。Qin 等人[9]提出了一种单层神经网络模型以解决目标函数为伪凸非光滑的优化问题。

目前提出的解决非光滑优化问题的递归神经网络模型大多基于早期罚函数的思想，需要罚因子足够大才能保证网络收敛到优化问题的最优解，因此必须在网络执行计算前计算出罚因子。而这个参数在某些目标函数下是难以计算的，这为网络执行计算带来困难。为此本文拟借鉴 Lagarange 乘子罚函数的思想提出一种解决非光滑非凸优化问题的递归神经网络模型，该网络模型的罚因子是变量，且无需事先计算罚因子的初始值仍能保证神经网络收敛到优化问题的最优解。此外，与传统 Lagrange 方法不同，该网络模型增加了一个等式约束惩罚项，可以提高网络的收敛能力。

本论文组织结构如下：第 2 节，介绍了本文需要解决的优化问题；第 3 节，介绍了增广拉格朗日神经网络模型；第 4 节，对该神经网络模型进行相关的理论分析；第 5 节，给出两个仿真实例验证理论的正确性和有效性。

2　优化问题

本文考虑如下的优化问题：

$$\left.\begin{array}{ll}\min & f(x)\\ \text{s.t.} & \begin{cases} g(x)\le 0\\ h(x)=0\end{cases}\end{array}\right\} \tag{1}$$

这里，$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，目标函数 $f:R^n\to R$，约束函数 $\boldsymbol{g}:R^n\to R^m$，$\boldsymbol{h}:R^n\to R^r$，其中 $m\le n$，f 是非光滑非凸函数，$\boldsymbol{g}=(g_1,g_2,\cdots,g_m)$ 是非光滑凸函数，$\boldsymbol{h}=(h_1,h_2,\cdots,h_r)$ 是仿射约束函数。这里定义 $S_1=\{\boldsymbol{x}\in R^n:\boldsymbol{g}(\boldsymbol{x})\le 0\}$，$S_2=\{\boldsymbol{x}\in R^n:\boldsymbol{h}(\boldsymbol{x})=0\}$，满足约束函数的可行域定义为 $S=\{\boldsymbol{x}\in R^n:\boldsymbol{g}(\boldsymbol{x})\le 0,\boldsymbol{h}(\boldsymbol{x})=0\}$，显然，$S=S_1\cap S_2$。对于本文研究的优化问题，假设可行域是凸的，并且存在 $\tilde{\boldsymbol{x}}\in R^n$ 和 $\delta>0$，满足 $\tilde{\boldsymbol{x}}\in \mathrm{int}(S_1)\cap S_2$，$S_1\subseteq B(\tilde{\boldsymbol{x}},\delta)$，其中 $\mathrm{int}(S_1)$ 表示集合 S_1 的内部，$B(\tilde{\boldsymbol{x}},\delta)$ 表示以 $\tilde{\boldsymbol{x}}$ 为中心半径为 δ 的开球区域，即可行域是有界凸的，同时假定优化问题式(1)有解。

3　神经网络模型

为了便于解决问题，定义：$G(\boldsymbol{x})=\begin{cases}0, & \boldsymbol{x}\in S_1\\ \sum_{j\in J^+(x)} g_j(\boldsymbol{x}), & \boldsymbol{x}\notin S_1\end{cases}$，$H(\boldsymbol{x})=\|\boldsymbol{h}(\boldsymbol{x})\|_2$，$J^+(x)=\{j\in\{1,2,\cdots,m\}:g_j(x)>0\}$，显然，$\{\boldsymbol{x}:G(\boldsymbol{x})\le 0\}=S_1$，$\{\boldsymbol{x}:H(\boldsymbol{x})=0\}=S_2$。

定义 1　增广拉格朗日函数：$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})=f(\boldsymbol{x})+\boldsymbol{\lambda}^{\mathrm{T}}H(\boldsymbol{x})+\boldsymbol{\mu}^{\mathrm{T}}G(\boldsymbol{x})+\frac{1}{2}c\,||\,\boldsymbol{h}(\boldsymbol{x})\,||^2$，其中 $\boldsymbol{\lambda}\in R^r$，$\boldsymbol{\mu}\in R^m$ 为 Lagrange 乘子，标量 $c>0$。

基于增广拉格朗日函数，我们提出以下神经网络模型以解决优化问题：

$$\left.\begin{array}{l}\dot{\boldsymbol{x}}\in-\partial f(\boldsymbol{x})-\lambda\partial H(\boldsymbol{x})-\mu\partial G(\boldsymbol{x})-c(\partial\boldsymbol{h}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}),\\ \qquad \boldsymbol{x}(t_0)\in B(\tilde{\boldsymbol{x}},\delta)\\ \dot{\lambda}=\begin{cases}0, & \boldsymbol{x}\in S_2\\ H(\boldsymbol{x})+\varepsilon_1, & \boldsymbol{x}\notin S_2\end{cases},\ \lambda(t_0)>0\\ \dot{\mu}=\begin{cases}0, & \boldsymbol{x}\in S_1\\ G(\boldsymbol{x})+\varepsilon_2, & \boldsymbol{x}\notin S_1\end{cases},\ \mu(t_0)>0\end{array}\right\}\tag{2}$$

其中，$\partial f(\boldsymbol{x})$，$\partial G(\boldsymbol{x})$，$\partial H(\boldsymbol{x})$，$\partial \boldsymbol{h}(\boldsymbol{x})$ 分别表示 $f(\boldsymbol{x})$，$G(\boldsymbol{x})$，$H(\boldsymbol{x})$，$\boldsymbol{h}(\boldsymbol{x})$ 的 Clark 广义梯度，$\varepsilon_1>0$，$\varepsilon_2>0$ 为常量，t_0 表示神经网络模型初始状态的时间。

上述神经网络模型中参数 c 的作用：(1)凸化目标函数，当参数 c 足够大时，可以使优化问题满足局部凸性；(2)加快网络轨迹的收敛速度。

式(2)给出了神经网络的状态方程，即网络模型的学习方法，这里给出该模型对应的硬件实现及其工作过程(见图 1)。当对此硬件电路输入初始电压 $\boldsymbol{x}_0$ 后，电流会在电路中震荡最终趋于稳定，若电路设计合理，那么电压的稳定值则是原始问题的最优解。

4　主要定理及证明

引理 1[6]　$\forall\beta(\boldsymbol{x})\in\partial G(\boldsymbol{x})$，当 $\boldsymbol{x}\notin S_1$ 时，有 $<\beta(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\ \ge -G_m$，其中 $G_m=\max\limits_{1\le j\le m} g_j(\tilde{\boldsymbol{x}})<0$。

引理 2　$\forall\alpha(\boldsymbol{x})\in\partial H(\boldsymbol{x})$，当 $\boldsymbol{x}\notin S_2$ 时，有 $<\alpha(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\ \ge H_r(\boldsymbol{x})$，其中 $H_r(\boldsymbol{x})=\min\limits_{1\le i\le r}\|h_i(\boldsymbol{x})\|_2$

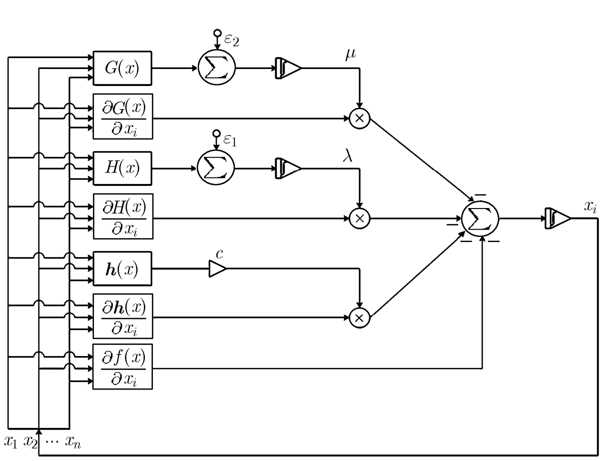

图 1 神经网络模型的硬件实现

>0，当 $\boldsymbol{x}\in S_2$，有 $<\alpha(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\geq 0$。

证明 因为 $\|\cdot\|$ 是凸函数且 $\boldsymbol{h}(\boldsymbol{x})$ 为仿射函数，所以 $H(\boldsymbol{x})$ 为凸函数。从而 $\forall\alpha(\boldsymbol{x})\in\partial H(\boldsymbol{x})$，有

$$<\alpha(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\geq\left(H(\boldsymbol{x})-H(\tilde{\boldsymbol{x}})\right) \tag{3}$$

当 $\boldsymbol{x}\notin S_2$ 时，$H(\boldsymbol{x})-H(\tilde{\boldsymbol{x}})=H(\boldsymbol{x})$，根据 $H(\boldsymbol{x})$ 及 $H_r(\boldsymbol{x})$ 的定义，必然有 $H(\boldsymbol{x})\geq H_r(\boldsymbol{x})$，因此 $<\alpha(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\geq H_r(\boldsymbol{x})$；当 $\boldsymbol{x}\in S_2$ 时，$H(\boldsymbol{x})-H(\tilde{\boldsymbol{x}})=0$，因此 $<\alpha(\boldsymbol{x}),\boldsymbol{x}-\tilde{\boldsymbol{x}}>\geq 0$。 证毕

定理 1 对于任意的初始点 $\boldsymbol{x}_0\in B(\tilde{\boldsymbol{x}},\delta)$，神经网络 $\boldsymbol{x}$ 的轨迹在有限时间内进入可行域 S 并停留在其中。

证明 这里定义能量函数：$W(\boldsymbol{x})=\lambda H(\boldsymbol{x})+\mu G(\boldsymbol{x})$，其中 $\lambda>0,\mu>0$，显然，

$$\{\boldsymbol{x}:G(\boldsymbol{x})\leq 0,H(\boldsymbol{x})=0\}=\{\boldsymbol{x}:W(\boldsymbol{x})\leq 0\}=S \tag{4}$$

对于 $B(\tilde{\boldsymbol{x}},\delta)\setminus S$ 外的任意一点 $\boldsymbol{x}(t)$，用反证法不难证明其轨迹 $\boldsymbol{x}(t)\in B(\tilde{\boldsymbol{x}},R_\delta)$，$R_\delta>\delta$。考虑 $B(\tilde{\boldsymbol{x}},\delta)\setminus S$ 外的任意一点 $\boldsymbol{x}(t)$，只需证明 $\{\boldsymbol{x}\in B(\tilde{\boldsymbol{x}},R_\delta)\setminus S_1\}$ 与 $\{\boldsymbol{x}\in S_1\setminus S_2\}$ 两个集合中的点在有限时间内能够进入可行域并停留在其中即可。

首先，考虑 $\{\boldsymbol{x}\in B(\tilde{\boldsymbol{x}},R_\delta)\setminus S_1\}$，因为 $\partial\boldsymbol{h}(\boldsymbol{x}(t))=\boldsymbol{A}^{\mathrm{T}}$，其中 $\boldsymbol{A}$ 为 $\boldsymbol{h}(\boldsymbol{x}(t))$ 的系数矩阵，$\forall\alpha(t)\in\partial H(\boldsymbol{x}(t))$，$\forall\beta(t)\in\partial G(\boldsymbol{x}(t))$，$\forall\upsilon(t)\in\partial f(\boldsymbol{x}(t))$，根据链式法则有

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))&=<\lambda\alpha(t)+\mu\beta(t),\dot{\boldsymbol{x}}(t)>\\&=<\lambda\alpha(t)+\mu\beta(t),-\upsilon(t)-\lambda\alpha(t)\\&\quad-\mu\beta(t)-c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))>\\&\leq\left(\|\lambda\alpha(t)+\mu\beta(t)\|\right)\cdot\left(\|\upsilon(t)\|\right.\\&\quad\left.+\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|-\|\lambda\alpha(t)+\mu\beta(t)\|\right)\end{aligned} \tag{5}$$

由引理 2 知 $<\alpha(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq 0$，因此 $<\lambda\alpha(t)+\mu\beta(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq<\mu\beta(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>$。

又引理 1 知 $<\beta(t),\ \boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq -G_m$，因此 $<\lambda\alpha(t)+\mu\beta(t),\ \boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq<\mu\beta(t),\ \boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq -\mu G_m$。

因为 $<\lambda\alpha(t)+\mu\beta(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\leq\|\lambda\alpha(t)+\mu\beta(t)\|\cdot\|\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}\|\leq R_\delta\|\lambda\alpha(t)+\mu\beta(t)\|$，所以，结合上述公式可知 $\|\lambda\alpha(t)+\mu\beta(t)\|\geq\dfrac{-\mu G_m}{R_\delta}$。

令 l_f 为 $f(\boldsymbol{x}(t))$ 在 $B(\tilde{\boldsymbol{x}},R_\delta)$ 上的 Lipschitz 常数，可知 $\upsilon(t)\leq l_f$。考虑到 $\boldsymbol{x}(t)\in B(\tilde{\boldsymbol{x}},R_\delta)$，可知 $\boldsymbol{h}(\boldsymbol{x}(t))$ 有上界，因此必存在一个常量 $\varphi>0$ 满足 $\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|\leq\varphi$。

假设 $\boldsymbol{x}(t)$ 的轨迹在有限时间内无法进入可行域 S 并停留其中，由 $\dfrac{\mathrm{d}\mu}{\mathrm{d}t}=G(\boldsymbol{x})+\varepsilon_1>0$，可知必存在某一时刻 t_1，当 $t\in[t_1,+\infty)$ 时，有 $\mu>\dfrac{(l_f+\varphi)R_\delta}{-G_m}$，即 $l_f+\varphi-\dfrac{-\mu G_m}{R_\delta}<0$，因此可得

$$\begin{aligned}&\|\upsilon(t)\|+\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|-\|\lambda\alpha(t)+\mu\beta(t)\|\\&\quad\leq l_f+\varphi-\frac{-\mu G_m}{R_\delta}<0\end{aligned} \tag{6}$$

又因为 $\|\lambda\alpha(t)+\mu\beta(t)\|\geq\dfrac{-\mu G_m}{R_\delta}>0$，可知 $\dfrac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))\leq-\left(\dfrac{-\mu G_m}{R_\delta}\right)\left(\dfrac{-\mu G_m}{R_\delta}-l_f-\varphi\right)$。

令 $\left(\dfrac{-\mu G_m}{R_\delta}\right)\left(\dfrac{-\mu G_m}{R_\delta}-l_f-\varphi\right)=\eta_1$，则当 $t\in[t_1,+\infty)$ 时，$\dfrac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))\leq-\eta_1<0$，可知经过有限时间后，$\boldsymbol{x}(t)$ 的轨迹必进入可行域 S 并停留其中，与假设矛盾。

其次，考虑 $\{\boldsymbol{x}\in S_1\setminus S_2\}$。同理 $\forall\alpha(t)\in\partial H(\boldsymbol{x}(t))$，$\forall\upsilon(t)\in\partial f(\boldsymbol{x}(t))$，根据链式法则有

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))&=<\lambda\alpha(t),\dot{x}(t)>\\&=<\lambda\alpha(t),-\upsilon(t)-\lambda\alpha(t)-c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))>\\&\leq\left(\|\lambda\alpha(t)\|\right)\left(\|\upsilon(t)\|+\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|-\|\lambda\alpha(t)\|\right)\end{aligned} \tag{7}$$

令 $H_m=\min\limits_{x\in S_1\setminus S_2}H_r(\boldsymbol{x})$，由引理 2 知 $\boldsymbol{x}\notin S_2$ 时，$<\alpha(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\geq H_m$。因为 $<\lambda\alpha(t),\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}>\leq\|\lambda\alpha(t)\|\cdot\|\boldsymbol{x}(t)-\tilde{\boldsymbol{x}}\|\leq R_\delta\|\lambda\alpha(t)\|$，所以 $\|\lambda\alpha(t)\|\geq\dfrac{\lambda H_m}{R_\delta}$。

同理 $\upsilon(t)\leq l_f$，并且存在一个常数 $\varphi>0$，使得 $\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|\leq\varphi$。

假设，在有限时间内 $\boldsymbol{x}(t)$ 的轨迹无法进入可行域 S 并停留其中，则由 $\dfrac{\mathrm{d}\lambda}{\mathrm{d}t}=H(\boldsymbol{x})+\varepsilon_2>0$，可知必存在某一时刻 t_2，当 $t\in[t_2,+\infty)$ 时，有 $\lambda>\dfrac{(l_f+\varphi)R_\delta}{H_m}$，即 $l_f+\varphi-\dfrac{\lambda H_m}{R_\delta}<0$，因此可得

$$\|\upsilon(t)\|+\left\|c\boldsymbol{A}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}(t))\right\|-\|\lambda\alpha(t)\|\leq l_f+\varphi-\frac{\lambda H_m}{R_\delta}<0 \tag{8}$$

又因为 $\|\lambda\alpha(t)\|\geq\dfrac{\lambda H_m}{R_\delta}>0$，可知 $\dfrac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))\leq-\left(\dfrac{\lambda H_m}{R_\delta}\right)\left(\dfrac{\lambda H_m}{R_\delta}-l_f-\varphi\right)$。

令 $\left(\dfrac{\lambda H_m}{R_\delta}\right)\left(\dfrac{\lambda H_m}{R_\delta}-l_f-\varphi\right)=\eta_2$，则当 $t\in[t_2,+\infty)$，$\dfrac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))\leq-\eta_2<0$，可知经过有限时间后，$\boldsymbol{x}(t)$ 的

轨迹必进入可行域 S 并停留其中，与假设矛盾。

证毕

定义 2　神经网络的平衡点集合 ε_σ 为

$$\varepsilon_\sigma=\Big\{\boldsymbol{x}^*,\lambda^*,\mu^*\Big|0\in\partial f\big(\boldsymbol{x}^*\big)+\lambda^*\partial H\big(\boldsymbol{x}^*\big)+\mu^*\partial G\big(\boldsymbol{x}^*\big)+c\boldsymbol{h}\big(\boldsymbol{x}^*\big)\partial\boldsymbol{h}\big(\boldsymbol{x}^*\big),\ H\big(\boldsymbol{x}^*\big)=0,G\big(\boldsymbol{x}^*\big)=0\Big\}$$

定理 2　给定任意的初始点 $\boldsymbol{x}_0\in B(\tilde{x},\delta)$，神经网络轨迹 $\boldsymbol{x}(t)$ 满足 $\lim_{t\to+\infty}\mathrm{dist}(\boldsymbol{x}(t),\varepsilon_\sigma)=0$。

证明　使用能量函数：$W(\boldsymbol{x})=f(\boldsymbol{x})+\lambda^{\mathrm{T}}H(\boldsymbol{x})+\mu^{\mathrm{T}}G(\boldsymbol{x})+\frac{1}{2}c\|\boldsymbol{h}(\boldsymbol{x})\|^2$，那么

$$\frac{\mathrm{d}}{\mathrm{d}t}W(\boldsymbol{x}(t))=\frac{\partial W(\boldsymbol{x})}{\partial\boldsymbol{x}}\cdot\frac{\partial\boldsymbol{x}}{\partial t}+\frac{\partial W(\boldsymbol{x})}{\partial\lambda}\cdot\frac{\partial\lambda}{\partial t}+\frac{\partial W(\boldsymbol{x})}{\partial\mu}\cdot\frac{\partial\mu}{\partial t}\tag{9}$$

采用反证法来证明，假设 $\lim_{t\to+\infty}\sup\mathrm{dist}(\boldsymbol{x}(t),\varepsilon_\sigma)=d>0$。

令 $S_t=S\setminus(\varepsilon_\sigma+B(0,(1/4)d))$，则存在序列 $\{t_n\}$ 满足 $n\to+\infty$ 时，$t_n\to+\infty$ 且 $\mathrm{dist}(\boldsymbol{x}(t_n),\varepsilon_\sigma)>d/2$，$\lim_{n\to+\infty}\mathrm{dist}(\boldsymbol{x}(t_n),\varepsilon_\sigma)=d$。

假定当 $t\in\big[t_0',+\infty\big]$，神经网络轨迹停留在可行域中，则对任意 $t_2'>t_1'>t_0'$，有 $\big\|\boldsymbol{x}\big(t_2'\big)-\boldsymbol{x}\big(t_1'\big)\big\|\le l_W\big|t_2'-t_1'\big|$，其中 l_W 为 $W(\boldsymbol{x})$ 在 $B(\tilde{\boldsymbol{x}},R_\delta)$ 上的 Lipschitz 常数。

令 $t\in\big[t_n,t_n+d/(4l_W)\big]$，有 $\mathrm{dist}(\boldsymbol{x}(t),\varepsilon_\sigma)\ge\mathrm{dist}\big(\boldsymbol{x}(t_n),\varepsilon_\sigma\big)-\|\boldsymbol{x}(t)-\boldsymbol{x}(t_n)\|\ge\frac{d}{2}-l_W|t-t_n|\ge\frac{d}{4}$，所以 $\boldsymbol{x}(t)\in S_t$。

当 $\boldsymbol{x}(t)\in S_t$，$\frac{\partial W(\boldsymbol{x})}{\partial\lambda}=0$，$\frac{\partial W(\boldsymbol{x})}{\partial\mu}=0$，则 $\frac{\mathrm{d}}{\mathrm{d}t}W=-\|\dot{\boldsymbol{x}}(t)\|^2<0$。因 $\partial W(\boldsymbol{x})$ 是有界紧的，则必存在一个 $\theta>0$ 满足 $\min\limits_{\boldsymbol{x}\in S_t}\Big\{\min\limits_{\nu\in\partial W(\boldsymbol{x})}\|\nu\|\Big\}=\theta$，由此可知，$\frac{\mathrm{d}}{\mathrm{d}t}W\le-\theta^2<0$，当 $t>t_0'$，$W(x(t))$ 是非递增能量函数，由上述可知在 $d/(4l_W)$ 的时间间隔内，$W(\boldsymbol{x})$ 减小量大于 $\frac{d}{4l_W}\theta^2$，这样经过无限个 $d/(4l_W)$ 的时间间隔后，$W(\boldsymbol{x}(t))$ 必趋于负无穷，而这与 $W(\boldsymbol{x}(t))$ 有界是矛盾的，因此可以得出 $\lim_{t\to+\infty}\mathrm{dist}(\boldsymbol{x}(t),\varepsilon_\sigma)=0$。

定义 3　优化问题的关键点集合 C 为：$C=\{\boldsymbol{x}\in S:0\in\partial f(\boldsymbol{x})+N_S(\boldsymbol{x})\}$，其中 $N_S(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 在可行域 S 上的法锥。

定理 3　对 $\boldsymbol{x}\in S$，拉格朗日神经网络的平衡点 ε_σ 与优化问题的关键点集 C 的关系：$\varepsilon_\sigma\subseteq C$。

证明　假设 $\big(\boldsymbol{x}^*,\lambda^*,\mu^*\big)$ 是拉格朗日神经网络的一个平衡点，即 $\big(\boldsymbol{x}^*,\lambda^*,\mu^*\big)$ 满足以下等式：

$$\begin{cases}0\in\partial f\big(\boldsymbol{x}^*\big)+\lambda^*\partial H\big(\boldsymbol{x}^*\big)+\mu^*\partial G\big(\boldsymbol{x}^*\big)+c\partial\boldsymbol{h}\big(\boldsymbol{x}^*\big)\boldsymbol{h}\big(\boldsymbol{x}^*\big)\\0=H\big(\boldsymbol{x}^*\big)\\0=G\big(\boldsymbol{x}^*\big)\end{cases}$$

因为 $\big(\boldsymbol{x}^*,\lambda^*,\mu^*\big)$ 是平衡点，所以 $\boldsymbol{h}\big(\boldsymbol{x}^*\big)=0$，那么

$$0\in\partial f\big(\boldsymbol{x}^*\big)+\lambda^*\partial H\big(\boldsymbol{x}^*\big)+\mu^*\partial G\big(\boldsymbol{x}^*\big)\tag{10}$$

根据文献 [4] 可知 $N_S(\boldsymbol{x})=\Big\{\sum_{i=1}^{r}p_i\partial h_i(\boldsymbol{x})+\sum_{j\in J^0(x)}q_j\partial g_j(\boldsymbol{x}):p_i\in R,q_j\ge0\Big\}$，其中，$J^0(\boldsymbol{x})=\big\{j\in\{1,2,\cdots,m\}:g_j(\boldsymbol{x})=0\big\}$。

当 $\boldsymbol{x}\in S_2$，$\partial H(\boldsymbol{x})=\big\{\boldsymbol{A}^{\mathrm{T}}\xi:\xi\in R^r,\|\xi\|\le1\big\}$，因为 $\boldsymbol{A}^{\mathrm{T}}\xi=\sum_{i=1}^{r}\boldsymbol{A}_i^{\mathrm{T}}\xi_i$，其中 $\boldsymbol{A}_i$ 为 $\boldsymbol{A}$ 的第 i 行元素，所以 $\partial H(\boldsymbol{x})=\Big\{\sum_{i=1}^{r}\boldsymbol{A}_i^{\mathrm{T}}\xi_i:\xi\in R^r,\|\xi\|\le1\Big\}$，由 $\partial h_i(\boldsymbol{x})=\boldsymbol{A}_i^{\mathrm{T}}$ 知 $\Big\{\sum_{i=1}^{r}p_i\partial h_i(\boldsymbol{x}):p_i\in R\Big\}=\Big\{\sum_{i=1}^{r}p_i\boldsymbol{A}_i^{\mathrm{T}}:p_i\in R\Big\}$。

因此对于 $\lambda^*\in R$，综上可知

$$\lambda^*\partial H(\boldsymbol{x}^*)\subset\left\{\sum_{i=1}^{r}p_i\partial h_i(\boldsymbol{x}):p_i\in R\right\}\tag{11}$$

同理，当

$$\boldsymbol{x}\in S_1,\ \partial G(\boldsymbol{x})=\begin{cases}0, & \boldsymbol{x}\in\mathrm{int}(S_1)\\\sum\limits_{j\in J^0(\boldsymbol{x})}[0,1]\partial g_j(\boldsymbol{x}), & \boldsymbol{x}\in bd(S_1)\end{cases}$$

又由 $\mu^*\ge0$，可知

$$\mu^*\partial G\big(\boldsymbol{x}^*\big)\subset\left\{\sum_{j\in J^0(\boldsymbol{x})}q_j\partial g_j(\boldsymbol{x}):q_j\ge0\right\}\tag{12}$$

结合式 (11)，式 (12) 和 $N_S(\boldsymbol{x})$ 的定义可得 $\lambda^*\partial H\big(\boldsymbol{x}^*\big)+\mu^*\partial G\big(\boldsymbol{x}^*\big)\subset N_S(\boldsymbol{x})$。

再结合式(10)可知 $0\in\partial f\big(\boldsymbol{x}^*\big)+N_S\big(\boldsymbol{x}^*\big)$，根据定义 3 可知 $\boldsymbol{x}^*\in C$，又因为 $\boldsymbol{x}^*\in\varepsilon_\sigma$，因此 $\varepsilon_\sigma\subseteq C$。

证毕

5　数值仿真及结果分析

仿真实验在 MatlabR2012 平台下进行，实验包括两个实例：

例 1 目标函数为非凸非光滑，限制条件为光滑凸函数

$$\left.\begin{aligned} \min\quad & f(\boldsymbol{x})=x_1^2-|x_2| \\ \text{s.t.}\quad & 2x_2+x_2^2-5\le 0 \\ & x_1+x_2-1=0 \end{aligned}\right\} \tag{13}$$

可知 $S_1=\left\{\boldsymbol{x}\in R^2: 2x_2+x_2^2-5\le 0\right\}$, $S_2=\left\{\boldsymbol{x}\in R^2: x_1+x_2-1=0\right\}$, $S=S_1\cap S_2$。根据是否满足 S_1 和 S_2 分情况选择开球 $B\left((1,0)^{\mathrm T},9\right)$ 内不同的初始点 $(7,0)^{\mathrm T}$, $(3,-2)^{\mathrm T}$, $(0,2)^{\mathrm T}$, $(-6,-1)^{\mathrm T}$ 。设定拉格朗日乘子 $\lambda(t_0)=0.6$， $\mu(t_0)=0.3$， 参数 $c=1$。 x_1 和 x_2 的状态轨迹分别如图 2 和图 3 所示，可以看出状态向量 $\boldsymbol{x}$ 轨迹收敛于平衡点 $(-0.4495,1.4495)^{\mathrm T}$，近似于原始优化问题的最优解 $(-0.4494,1.4494)^{\mathrm T}$ 。

例 2 目标函数为非光滑非凸，限制条件包含非光滑凸约束函数

$$\left.\begin{aligned} \min\quad & f(\boldsymbol{x})=|x_1|-x_2^2 \\ \text{s.t.}\quad & |x_1|-5\le 0 \\ & 2x_2+x_2^2-5\le 0 \\ & x_1+x_2-1=0 \end{aligned}\right\} \tag{14}$$

可知 $S_1=\left\{\boldsymbol{x}\in R^2: |x_1|-5\le 0,\ 2x_2+x_2^2-5\le 0\right\}$， $S_2=\left\{\boldsymbol{x}\in R^2: x_1+x_2-1=0\right\}$, $S=S_1\cap S_2$。根据是否满足 S_1 和 S_2 分情况选择开球 $B\left((1,0)^{\mathrm T},8\right)$ 内不同的初始点 $(-6,-1)^{\mathrm T}$, $(-4,1.5)^{\mathrm T}$, $(1,-2)^{\mathrm T}$, $(5,1)^{\mathrm T}$ 。设定拉格朗日乘子 $\lambda(t_0)=0.6$, $\mu(t_0)=0.3$，参数 $c=1$。x_1 和 x_2 的状态轨迹分别如图 4 和图 5 所示，可以看出状态向量 $\boldsymbol{x}$ 轨迹收敛于平衡点 $(4.4903,-3.4606)^{\mathrm T}$，近似于原始优化问题的最优解 $(4.4494,-3.4494)^{\mathrm T}$ 。

此外，为了比较本文所提出的增广拉格朗日网络模型和传统拉格朗日网络模型，设定初始点 $(-6,-1)^{\mathrm T}$，参数 c 分别取 $0,0.4,1$，其他参数不变。事实上当 $c=0$ 增广拉格朗日网络模型就退化为传统拉格朗日网络模型。状态向量 $\boldsymbol{x}$ 的轨迹如图 6 和图 7 所示。从图中可以发现，网络随着参数 c 的增加，收敛速度也相应增加，而当 $c=0$ 时，即在传统拉格朗日网络模型下，网络状态甚至不收敛。

6 结束语

本文旨在解决工程应用中经常出现的含有约束的非光滑非凸优化问题，提出了一种基于拉格朗日思想的神经网络模型，通过增加惩罚约束项以凸化目标函数来求解优化问题，分析了神经网络的收敛轨迹在有限时间内进入可行域，并且最终收敛于平衡点集，同时给出了平衡点集与关键点集的关系。最后实例验证了所提出的增广拉格朗日神经网络求解带等式及不等式约束非光滑非凸优化问题的有效性，同时通过与传统拉格朗日网络模型解决相同问题进行比较，表明其在收敛性方面更优。

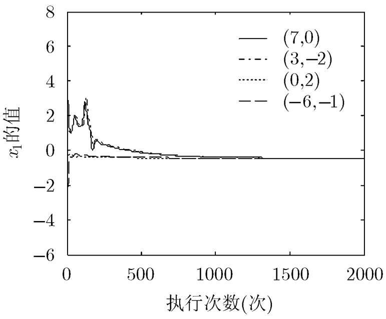

图 2 例 1 中 x_1 的轨迹

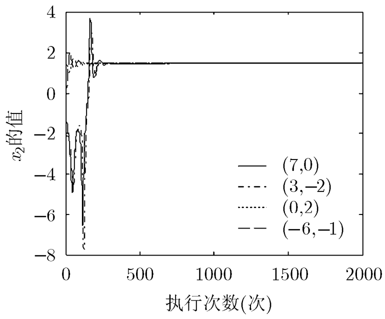

图 3 例 1 中 x_2 的轨迹

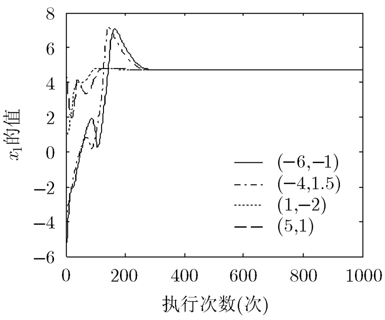

图 4 例 2 中 x_1 的轨迹

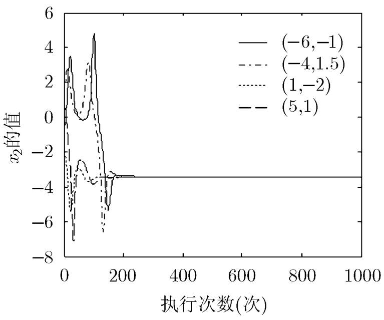

图 5 例 2 中 x_2 的轨迹

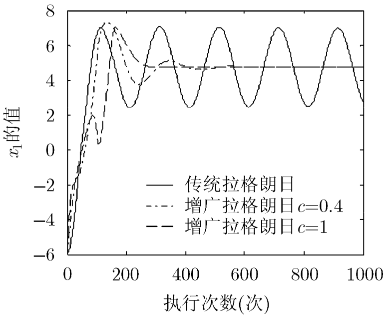

图 6　c 取不同值 x_1 的轨迹

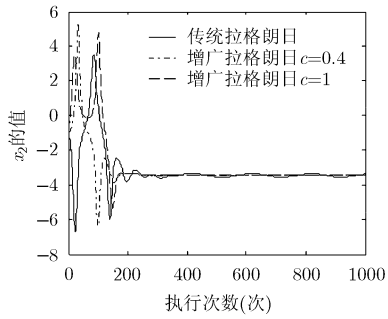

图 7　c 取不同值 x_2 的轨迹

参 考 文 献

[1] MIAO Peng, SHEN Yanjun, LI Yujiao, *et al.* Finite-time recurrent neural networks for solving nonlinear optimization problems and their application[J]. *Neurocomputing*, 2016, 177(7): 120–129. doi: 10.1016/j.neucom.2015.11.014.

[2] MESTARI M, BENZIRAR M, SABER N, *et al.* Solving nonlinear equality constrained multiobjective optimization problems using neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(10): 19–35. doi: 10.1109/TNNLS.2015.2388511.

[3] HOSSEINI A. A non-penalty recurrent neural network for solving a class of constrained optimization problems[J]. *Neural Networks*, 2016, 73(1): 10–25. doi: 10.1016/j.neunet.2015.09.013.

[4] FORTI M, NISTRI P, and QUINCAMPOIX M. Generalized neural network for nonsmooth nonlinear programming problems[J]. *IEEE Transactions on Circuits and Systems I Regular Papers*, 2004, 51(9): 1741–1754. doi: 10.1109/TCSI.2004.834493.

[5] 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.

[6] 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.

[7] LIU Qingshan and WANG Jun. A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 24(5): 812–824. doi: 10.1109/TNNLS.2013.2244908.

[8] BIAN Wei and CHEN Xiaojun. Smoothing neural network for constrained non-Lipschitz optimization with applications [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(3): 399–411. doi: 10.1109/TNNLS.2011.2181867.

[9] QIN Sitian, BIAN Wei, and XUE Xiaoping. A new one-layer recurrent neural network for nonsmooth pseudoconvex optimization[J]. *Neurocomputing*, 2013, 120(22): 655–662. doi: 10.1016/j.neucom.2013.01.025.

喻　昕：男，1973 年生，教授，博士，博士生导师，主要研究方向为人工智能、人工神经网络、优化计算.

许治健：男，1990 年生，硕士生，研究方向为人工神经网络、优化计算.

陈昭蓉：女，1991 年生，硕士生，研究方向为人工神经网络、优化计算.

徐辰华：男，1976 年生，副教授，博士，主要研究方向为过程控制、人工神经网络.