

基于投影算子的回归神经网络模型 及其在最优化问题中的应用*

马儒宁^{1,2}, 陈天平²

(1. 南京航空航天大学 数学系, 南京 210016;

2. 复旦大学 数学研究所, 上海 200433)

(刘曾荣推荐)

摘要: 研究了一种基于投影算子的神经网络模型. 与以前研究投影算子的值域一般是 n 维欧氏空间中的紧凸子集不同, 而是 n 维欧氏空间中未必有界的闭凸子集, 同时目标函数也是一般的连续可微函数, 未必为凸函数. 证明了所研究的神经网络模型具有整体解轨道, 以及当目标函数满足某些条件时解轨道的整体收敛性. 此外, 还将所研究的模型应用于闭凸约束极小化问题以及非线性互补问题和隐互补问题中, 并通过数值模拟说明了该神经网络方法的有效性.

关键词: 回归神经网络模型; 投影算子; 整体收敛性; 最优化; 互补问题

中图分类号: O29; TP18 **文献标识码:** A

引 言

自从 Hopfield 和 Tank 的开创性工作^[1, 2]以来, 用于最优化问题的神经网络方法得到了迅速的发展. 和其他解决最优化问题的数值方法相比, 神经网络具有易于硬件实现、并行计算等优点. 这样, 神经网络逐渐成为最优化问题实时解决的最佳选择之一.

通过回归型的神经网络解最优化问题, 特别是约束优化问题, 近年来已成为神经网络领域发展的热门课题, 出现了很多文章, 提出了许多不同的模型(例如文献[3~6]等). 由于闭凸约束的最优化问题可以转化为寻求闭凸集上非线性投影算子的零点问题^[7], 因而, 基于投影算子的神经网络模型成为解决约束最优化问题的有力工具.

在文献[5]和[8]中, 基于投影算子的神经网络模型分别被用于解决具有有界矩形约束和紧凸约束的连续可微优化问题; 在文献[9]中, 同样的模型被用于解决具有闭凸约束的连续可微凸优化问题; 在文献[6]中, 作者利用文献[5]中研究过的模型解决了一些可以化为非线性投影方程的约束优化问题.

在许多实际的应用中, 解决一般闭凸集(未必有界)上连续可微(未必凸)优化问题是很重要的, 因此, 我们需要研究这样的投影算子神经网络, 其值域未必有界而且目标函数未必凸.

在给出我们所研究的回归型神经模型之前, 首先给出一点在闭凸集上投影的定义:

定义 设 Ω 为 \mathcal{R}^n 的非空闭凸子集, 点 $x \in \mathcal{R}^n$ 在集合 Ω 上的投影, 记作 $\text{Proj}_\Omega(x)$, 定义

* 收稿日期: 2004.03.24; 修订日期: 2006.01.10

作者简介: 马儒宁(1976—), 男, 山东济宁人, 博士(联系人. Tel: + 86_25_81672925;

E-mail: mrning@nuaa.edu.cn).

为 $\min_{y \in \Omega} \|x - y\|$, 其中 $\|x\| = \sqrt{x^T x}$ 表示向量 $x \in \mathcal{R}$ 的欧氏范数. (显然上述极小化问题的解一定存在且唯一)

我们考虑如下的回归神经网络模型(RNN):

$$\tau \frac{dx}{dt} = -x + \text{Proj}_{\Omega}(x - \alpha \cdot \dot{E}(x)), \quad (\text{A})$$

其中 $\tau > 0$ 和 $\alpha > 0$ 分别为 RNN 的时间常数和步长参数.

在本文中, 首先, 我们将在 Ω 闭凸和目标函数 $E(x)$ 连续可微的前提下, 证明上述 RNN(A) 存在从任意 Ω 内的点出发的整体解轨道; 其次, 我们将证明当目标函数的梯度有界以及 Lipschitz 连续时, 解轨道收敛到神经网络的平衡点; 最后, 我们将该神经网络模型应用于连续可微最优化问题以及非线性和隐互补问题, 并通过具体的数值模拟证实了所研究的 RNN 的有效性.

本文是这样安排的: 引言给出了一些背景知识与定义, 第 1 节证明了 RNN 的解轨道的整体存在性和整体收敛性, 第 2 节我们考虑了上述 RNN 的应用, 第 3 节给出了一些连续可微优化问题和非线性互补问题的数值实验, 最后, 在第 4 节, 我们总结了全文.

1 RNN(A) 的整体解轨道以及收敛性

首先, 我们给出几个以后要用到的引理:

引理 1^[10] 投影算子 $\text{Proj}_{\Omega}(\cdot)$ 具有以下性质:

(i) 若 $y \in \Omega$, 则对 $\forall x \in \mathcal{R}$, 有 $[\text{Proj}_{\Omega}(x) - x]^T [\text{Proj}_{\Omega}(x) - y] \leq 0$.

(ii) $\text{Proj}_{\Omega}(\cdot)$ 为 \mathcal{R}^n 上的非扩展算子, 即对 $\forall x, y \in \mathcal{R}$ 有: $\|\text{Proj}_{\Omega}(x) - \text{Proj}_{\Omega}(y)\| \leq \|x - y\|$.

引理 2^[8] 关于 $y \in \mathcal{R}^n$ 的函数 $\text{dist}^2(y, \Omega)$ 可微, 并且对 $\forall y \in \mathcal{R}$ 其梯度向量为

$$\dot{\text{dist}}^2(y, \Omega) = 2(y - \text{Proj}_{\Omega}(y)),$$

其中 $\text{dist}(y, \Omega)$ 表示点 y 和闭凸集 Ω 之间的距离, 即:

$$\text{dist}(y, \Omega) = \min_{x \in \Omega} \|y - x\| = \|y - \text{Proj}_{\Omega}(y)\|.$$

引理 3 如果 $f(t) \in C[0, +\infty)$ 一致连续且在 $[0, +\infty)$ 上的广义积分收敛, 则 $\lim_{t \rightarrow +\infty} f(t) = 0$.

这是数学分析中的一个简单推论(见文献[11]).

下面, 我们研究 RNN(A) 的动力学行为, 包括其整体解轨道及收敛性. 下面的命题是关于解轨道的局部存在性:

命题 1 对任意点 $x_0 \in \mathcal{R}$, RNN(A) 存在出发自 x_0 的局部解轨道, 若 $x_0 \in \Omega$, 则出发自 x_0 的局部解轨道不会跑出 Ω .

这个命题的证明和文献[8]中定理 2 的证明类似.

今后, 我们假设目标函数 $E(x)$ 在 Ω 上存在下界, 即存在常数 A 使得任意 $x \in \Omega$ 都有 $E(x) \geq A$.

下面的定理给出了神经网络模型(A)的解轨道的整体存在性:

定理 1 若 $E(x) \in C^1(\mathcal{R})$, Ω 为 \mathcal{R} 的闭凸子集, 并且 $x_0 \in \Omega$, 则 RNN(A) 存在出发自 x_0 的整体解轨道.

证明 由命题 1, RNN(A) 存在局部解轨道, 设其极大存在区间为 $[0, t^*)$, 我们只需证 t^*

可推广为 $+\infty$. 事实上, 对任意的 $t_1, t_2, 0 < t_1 < t_2 < t^*$, 我们有

$$E(\mathbf{x}(t_1)) - E(\mathbf{x}(t_2)) = -\frac{1}{\tau} \int_{t_1}^{t_2} \left[\tau \frac{dE(\mathbf{x}(t))}{dt} \right] dt \quad (1)$$

以及

$$\begin{aligned} \tau \frac{dE(\mathbf{x}(t))}{dt} &= \dot{\mathbf{x}}^T E(\mathbf{x}(t)) + \tau \frac{d\mathbf{x}}{dt} = \\ &= \dot{\mathbf{x}}^T E(\mathbf{x}(t)) + (-\mathbf{x}(t) + \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t))))^T = \\ &= -\frac{1}{\alpha} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\|^2 + \\ &= (\text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t))) - \\ &= (\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t))))^T \cdot (\text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t))) - \mathbf{x}(t)) \leq \\ &= -\frac{1}{\alpha} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\|^2. \end{aligned} \quad (2)$$

在不等式(2)的推导中, 我们应用了引理 1(i) 和命题 1.

根据式(1)和(2), 我们有

$$\begin{aligned} E(\mathbf{x}(t_1)) - E(\mathbf{x}(t_2)) &\geq \\ &= \frac{1}{\tau \alpha} \int_{t_1}^{t_2} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\|^2 dt > 0, \end{aligned}$$

因此, 当 $0 < t < t^*$ 时, $E(\mathbf{x}(t))$ 单调下降. 既然存在常数 A , 对任意的 $\mathbf{x} \in \Omega$, $E(\mathbf{x}) \geq A$ 并且 $\mathbf{x}(t) \in \Omega, \forall t \in [0, t^*)$, 我们可以得到: 对任意的 $0 < t_1 < t_2 < t^*$,

$$E(\mathbf{x}(t_1)) - E(\mathbf{x}(t_2)) \leq E(\mathbf{x}_0) - A \stackrel{\Delta}{=} M.$$

于是有

$$\int_{t_1}^{t_2} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\|^2 dt \leq \tau \alpha M. \quad (3)$$

另一方面, 我们有:

$$\begin{aligned} \|\mathbf{x}(t_1) - \mathbf{x}(t_2)\| &= \left\| \int_{t_1}^{t_2} \tau \frac{d\mathbf{x}(t)}{dt} dt \right\| \leq \\ &= \frac{1}{\tau} \int_{t_1}^{t_2} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\| dt \leq \\ &= \frac{1}{\tau} \left\{ \int_{t_1}^{t_2} \|\mathbf{x}(t) - \text{Proj}_{\Omega}(\mathbf{x}(t) - \alpha \dot{\mathbf{x}} E(\mathbf{x}(t)))\|^2 dt \right\}^{1/2} \cdot \left\{ \int_{t_1}^{t_2} dt \right\}^{1/2}. \end{aligned} \quad (4)$$

结合式(3)和(4), 我们可得:

$$\|\mathbf{x}(t_1) - \mathbf{x}(t_2)\| \leq \sqrt{\frac{\alpha \cdot M}{\tau}} \cdot \sqrt{t_2 - t_1}. \quad (5)$$

如果 t^* 为有限数, 根据式(5)可知当 $t \rightarrow t^*$ 时 $\{\mathbf{x}(t)\}$ 为 \mathcal{R}^n 中的 Cauchy 序列. 因此, $\lim_{t \rightarrow t^*} \mathbf{x}(t)$ 有限并且 $\mathbf{x}(t)$ 可以延拓到点 t^* . 根据在点 t^* 的局部存在定理, $\mathbf{x}(t)$ 可以延拓到 $t > t^*$, 这与 $[0, t^*)$ 为解轨道的极大存在区间矛盾, 故 $t^* = +\infty$.

可以看出, 虽然 RNN(A) 具有从任意 $\mathbf{x}_0 \in \Omega$ 出发的整体解轨道(当 $E(\mathbf{x})$ 连续可微及 Ω 闭凸时), 但是式(A)的解轨道 $\mathbf{x}(t)$ 未必有界. 这样, 我们就不能利用文献[8]或[9]中的方法来证明解轨道的收敛性.

我们令 $F(\mathbf{x}) = -\mathbf{x} + \text{Proj}_{\Omega}(\mathbf{x} - \alpha \dot{\mathbf{x}} E(\mathbf{x}))$, 则 RNN(A) 被用来寻求非线性投影算子

$F(x)$ 的零点, 也就是说, 回归神经网络模型(A) 就是通过解非线性投影方程 $F(x) = \mathbf{0}$ 来寻求目标函数 (x) 的极小点.

如果 $\dot{\cdot}E(x)$ 有界且 Lipschitz 连续, 我们可以证明下述与解的收敛性密切相关的结论:

定理 2 若 $E(x) \in C^1(\mathcal{R}^p)$, Ω 为 \mathcal{R}^p 的闭凸子集, $x_0 \in \Omega$ 并且 $\dot{\cdot}E(x)$ 在 Ω 内有界且 Lipschitz 连续, 则算子 $F(x)$ 沿着 RNN(A) 的解轨道 $x(t, x_0)$ 收敛于 0 (x_0 为出发点), 即:

$$\lim_{t \rightarrow +\infty} F(x(t, x_0)) = 0. \quad (6)$$

证明 既然 $\dot{\cdot}E(x)$ 有界且 Lipschitz 连续, 则存在正常数 M' 、 M'' 使得

$$\|\dot{\cdot}E(x)\| \leq M', \quad (7)$$

以及对任意的 $x, x_1, x_2 \in \Omega$,

$$\|\dot{\cdot}E(x_1) - \dot{\cdot}E(x_2)\| \leq M'' \|x_1 - x_2\|. \quad (8)$$

令 $f(t) = \|F(x(t))\|^2 = \|x(t) - \text{Proj}_{\Omega}(x(t) - \alpha \dot{\cdot}E(x(t)))\|^2, t \in [0, +\infty)$. 则我们只需证:

$$\lim_{t \rightarrow +\infty} f(t) = 0. \quad (9)$$

由于式(3)以及 $t^* = +\infty$, 我们有(令 $t_1 \rightarrow 0$ 和 $t_2 \rightarrow +\infty$):

$$\int_0^{+\infty} f(t) dt \leq \tau \alpha M' = M. \quad (10)$$

根据式(10)和引理 3, 我们只需证 $f(t)$ 一致连续. 下面, 我们证明 $f(t)$ 在 $[0, +\infty)$ 上 Lipschitz 连续.

根据引理 1(ii), 命题 1 以及(7), 对任意的 $t > 0$ 有

$$\begin{aligned} \|x(t) - \text{Proj}_{\Omega}(x(t) - \alpha \dot{\cdot}E(x(t)))\| &\leq \\ \|x(t) - (x(t) - \alpha \dot{\cdot}E(x(t)))\| &\leq \alpha M', \end{aligned} \quad (11)$$

因此 $\|x'(t)\| \leq (\alpha/\tau)M'$, 并且对任意的 $t_1, t_2 > 0$ 有

$$\|x(t_1) - x(t_2)\| \leq (\alpha/\tau)M' |t_1 - t_2|. \quad (12)$$

对任意的 $0 < t_1 < t_2 < +\infty$, 我们有:

$$\begin{aligned} |f(t_1) - f(t_2)| &= |[x(t_1) - \text{Proj}_{\Omega}(x(t_1) - \alpha \dot{\cdot}E(x(t_1)))] + \\ & [x(t_2) - \text{Proj}_{\Omega}(x(t_2) - \alpha \dot{\cdot}E(x(t_2)))]^T [x(t_1) - \\ & \text{Proj}_{\Omega}(x(t_1) - \alpha \dot{\cdot}E(x(t_1))) - \\ & x(t_2) + \text{Proj}_{\Omega}(x(t_2) - \alpha \dot{\cdot}E(x(t_2)))]| \leq \\ & [\|x(t_1) - \text{Proj}_{\Omega}(x(t_1) - \alpha \dot{\cdot}E(x(t_1)))\| + \\ & \|x(t_2) - \text{Proj}_{\Omega}(x(t_2) - \alpha \dot{\cdot}E(x(t_2)))\|] [\|x(t_1) - x(t_2)\| + \\ & \| \text{Proj}_{\Omega}(x(t_1) - \alpha \dot{\cdot}E(x(t_1))) - \text{Proj}_{\Omega}(x(t_2) - \alpha \dot{\cdot}E(x(t_2))) \|] \leq \\ & 2\alpha M' [2 \|x(t_1) - x(t_2)\| + \alpha \| \dot{\cdot}E(x(t_2)) - \dot{\cdot}E(x(t_1)) \|] \leq \\ & 2\alpha M' (2 + \alpha M'') (\alpha/\tau) M' |t_1 - t_2|. \end{aligned} \quad (13)$$

在不等式(13)的证明中利用了式(11)和引理 1(ii) 以及式(8)、(12).

因此 $f(t)$ Lipschitz 连续, 定理得证.

考虑下面的非线性极小化问题:

$$\min E(x), \text{ s. t. } x \in \Omega, \quad (14)$$

其中目标函数 $E(x)$ 连续可微, 即 $E(x) \in C^1(\mathcal{R}^p)$, $x \in \mathcal{R}^p$, 约束集合 $\Omega \subseteq \mathcal{R}^p$ 为非空闭凸集. 当 $\Omega = \mathcal{R}^p$, 上述问题就成为 \mathcal{R}^p 上无约束的极小化问题.

由于解轨道 $x(t)$ 未必有界, 定理 2 还不能直接得到 $x(t)$ 的收敛性. 下面, 我们将研究 $\text{RNN}(A)$ 的拟收敛性和整体收敛性.

令 $\Omega^* = \{x \in \Omega \mid E(x) \geq E(x^*), \forall x \in \Omega\}$, $\Omega^e = \{x^e \in \mathcal{R}^n \mid F(x^e) = 0\}$, 则 Ω^* 为最优化问题(14)的极小点集, Ω^e 为投影算子 $F(x)$ 的零点集(即 $\text{RNN}(A)$ 的均衡点集). 由文献[8]中的定理 1, 我们知道 Ω^* 和 Ω^e 都是 Ω 的闭子集, 并且 $\Omega^* \subset \Omega^e$. 值得注意的是, 如果系统(A)收敛, 其解轨道 $x(t)$ 一定有界. 下面的定义确保了解轨道 $x(t)$ 的有界性.

定义 1 \mathcal{R}^n 上的算子 $F(x)$ 称为在无穷远处没有零点的(no zero_at_infinity), 如果存在 $B > 0$ 以及 $\varepsilon_0 > 0$, 使得当 $\|x\| \geq B$ 时有 $\|F(x)\| \geq \varepsilon_0$.

根据式(6), 显然若 $F(x)$ 在无穷远处没有零点, 则方程(A)的解轨道 $x(t)$ 一定是有界的. 下面的定理给出了神经网络(A)的整体收敛性.

定理 3 如果 $F(x)$ 在无穷远处没有零点, 则在定理 2 的假设下, 我们有:

(i) 神经网络模型(A)是拟收敛的(quasi-convergence), 即

$$\lim_{t \rightarrow +\infty} \text{dist}(x(t), \Omega^e) = 0, \quad (15)$$

其中 $\text{dist}(\cdot, \cdot)$ 的定义同引理 2.

(ii) 如果 Ω^e 为孤立点集, 则 $\text{RNN}(A)$ 的解轨道是整体收敛的, 即存在 $x \in \Omega^e$ 使得 $\lim_{t \rightarrow +\infty} x(t) = x$.

(iii) 如果 Ω^e 只包含一个点, 则 $\text{RNN}(A)$ 是整体渐进收敛的.

证明 (i) 若不然, 则存在 $t_n \rightarrow +\infty$ 以及 $\delta > 0$, 对任意的 n , 使得

$$\text{dist}(x(t_n), \Omega^e) \geq \delta. \quad (16)$$

既然 $F(x)$ 在无穷远处没有零点, 则解轨道 $x(t)$ 有界, 故 $\{x(t_n)\}_{n=1}^{\infty}$ 有一个收敛子列, 记作 $\lim_{k \rightarrow \infty} x(t_{n_k}) = x$. 显然当 $k \rightarrow \infty$ 时 $t_{n_k} \rightarrow +\infty$. 根据 $F(x)$ 的连续性以及式(6), 有 $F(x) = \lim_{k \rightarrow \infty} F(x(t_{n_k})) = 0$, 即 $x \in \Omega^e$. 因此 $\text{dist}(x(t_{n_k}), \Omega^e) \rightarrow 0, k \rightarrow \infty$, 这与式(16)矛盾.

(ii) 根据式(15), 可知存在 $x \in \Omega$ 以及序列 $t_n \rightarrow +\infty$, 使得 $\lim_{n \rightarrow \infty} x(t_n) = x$. 如果 $\lim_{t \rightarrow \infty} x(t) \neq x$, 由于 Ω^e 为孤立点集, 则存在 $r > 0, N > 0$ 以及 $t_n \rightarrow +\infty$, 使得: (记 $B(x, r)$ 为以 x 为中心, r 为半径的球)

$$(a) B(x, 3r/2) \cap \Omega^e = \{x\};$$

$$(b) x(t_n) \notin B(x, r);$$

$$(c) x(t_n) \in B(x, r/2).$$

既然 $t_n, t_n \rightarrow +\infty$, 根据 $x(t)$ 的连续性, 存在序列 $t_n \rightarrow +\infty$ 使得

$$x(t_n) \in B(x, r) - B(x, r/2),$$

因此, 对任意的 n 我们有 $\text{dist}(x(t_n), \Omega^e) \geq \text{dist}(x(t_n), x) \geq r/2$, 这与式(15)式矛盾.

(iii) 如果 $\Omega^e = \{x\}$, 根据(ii), 对任意的初始点 $x_0 \in \Omega$, 我们有 $\lim_{t \rightarrow +\infty} x(t) = x$, 即 $\text{RNN}(A)$ 整体渐进收敛.

2 RNN(A) 在互补问题中的应用

在本节中, 我们将所研究的 $\text{RNN}(A)$ 应用到寻求如下一般的隐互补问题(implicit complementarity problem, ICP)的解:

$$\begin{cases} \text{求 } x \in \Omega, \\ \text{s.t. } g_i(x) \geq 0, F_i(x) \geq 0 \text{ 和 } g_i(x) \cdot F_i(x) = 0, \end{cases} \quad (17)$$

其中 $\Omega \subset \mathcal{R}^n$ 为闭凸集, 并且 $g, F: \mathcal{R}^n \rightarrow \mathcal{R}^n$ 连续可微. 当 $\Omega = \mathcal{R}^n, g(x) = x$ 并且 $F(x) = Mx + b, M \in \mathcal{R}^{n \times n}, b \in \mathcal{R}^n$, 问题(17) 就成为线性互补问题, 记作 LCP(M, b); 当 $\Omega = \mathcal{R}^n, g(x) = x$ 并且 $F(x)$ 为连续可微映射时, 问题(17) 就成为非线性互补问题, 记作 NCP(F). 因此, 问题(17) 是所有互补问题的最一般的形式.

为了寻求互补问题的解, 通常采用的一种方法就是将互补问题转化为极小化问题^[12~14]. 我们先给出互补函数的定义:

定义2 令 $\phi: \mathcal{R} \times \mathcal{R} \rightarrow \mathcal{R}$, 如果

$$\phi(a, b) = 0 \Leftrightarrow a \geq 0, b \geq 0 \text{ 和 } a \cdot b = 0, \quad (18)$$

则函数 ϕ 称为互补函数.

令 $g(x) = (g_i(x)), F(x) = (F_i(x))$, 定义

$$\Phi(x) = \begin{pmatrix} \phi(g_1(x), F_1(x)) \\ \vdots \\ \phi(g_n(x), F_n(x)) \end{pmatrix}. \quad (19)$$

显然问题(17) 的解等价于在闭凸集合 Ω 上解方程 $\Phi(x) = 0$, 也等价于如下约束极小化问题:

$$\begin{cases} \min E(x) = \frac{1}{2} \|\Phi(x)\|^2, \\ \text{s.t. } x \in \Omega. \end{cases} \quad (20)$$

在本文中, 我们采用如下的互补函数:

$$\phi(a, b) = (a - b)^2 - a|a| - b|b|, \quad (21)$$

显然上面定义的 $\phi(a, b)$ 连续可微, 并且:

$$\frac{\partial \phi(a, b)}{\partial a} = 2(a - b) - 2|a|, \quad \frac{\partial \phi(a, b)}{\partial b} = 2(b - a) - 2|b|. \quad (22)$$

这样, 目标函数 $E(x)$ (由(21)式定义) 也连续可微且:

$$\dot{E}(x) = \dot{\Phi}(x) \cdot \Phi(x) = [a(x) \cdot \dot{g}(x) + b(x) \cdot \dot{F}(x)] \cdot \Phi(x), \quad (23)$$

其中 $\dot{g}(x), \dot{F}(x)$ 分别为 $g(x), F(x)$ 的梯度,

$$\begin{cases} a(x) = \text{diag}(2(g_i(x) - F_i(x)) - 2|g_i(x)|)_{i=1}^n, \\ b(x) = \text{diag}(2(F_i(x) - g_i(x)) - 2|F_i(x)|)_{i=1}^n. \end{cases} \quad (24)$$

我们有如下的结论:

定理4 如果 Ω 为有界闭凸集合, 则对任意的连续可微映射 $g(x), \Phi(x): \mathcal{R}^n \rightarrow \mathcal{R}^n$, 定理3的结论成立, 即 RNN(A) 从 $x_0 \in \Omega$ 出发的解轨道 $x(t)$ 满足:

(i) $\lim_{t \rightarrow +\infty} \text{dist}(x(t), \Omega^e) = 0$;

(ii) 若 Ω^e 为孤立点集, 则存在 $x \in \Omega^e$ 使得 $\lim_{t \rightarrow \infty} x(t) = x$;

(iii) 若 Ω^e 只包含一个点, 则 RNN(A) 整体渐进收敛.

证明 既然 Ω 有界, 则 $g(x), F(x), \dot{g}(x)$ 以及 $\dot{F}(x)$ 都在 Ω 上有界, 一致连续, 此外, 根据命题1, 神经网络(A) 发自 $x_0 \in \Omega$ 的解轨道 $x(t)$ 满足 $x(t) \subset \Omega$, 即 $x(t)$ 有界, 这样, 定理3的结论成立, 该定理证毕.

从定理4, 我们知道如果 Ω 为有界闭凸集, 神经网络(A) 是收敛的. 这样, 对任意的闭凸集

合 Ω , 我们可以利用一系列有界闭凸集合 Ω_k 来逼近 Ω , 并分别在 Ω_k 上建立类似的神经网络模型. 一方面, 我们也可以通过选定某个特定的有界闭凸集来获得互补问题 (17) 在我们所感兴趣区域的解; 另一方面, 我们也可以逐步获得互补问题在整个区域 Ω 上的解.

下面, 我们针对互补问题研究集合 Ω^* 和 Ω^e 之间的关系, 有如下的结论:

命题 2 问题 (17) 的解集 Ω^* 与神经网络 (A) 的平衡点集 Ω^e 具有如下的关系:

(i) $\Omega^* \subset \Omega^e$;

(ii) 如果 $\Omega^e \subset \text{int } \Omega$ ($\text{int } \Omega$ 表示集合 Ω 的内部) 并且对任意的 $x^e \in \Omega^e$, $\cdot \Phi(x^e)$ 非奇异, 则 $\Omega^* = \Omega^e$.

证明 (i) 若 $x^* \in \Omega^*$, 则 $x^* \in \Omega$ 并且 $\Phi(x^*) = \mathbf{0}$. 由于 $\cdot E(x^*) = \cdot \Phi(x^*)^T \cdot \Phi(x^*)$, 我们有 $\cdot E(x^*) = \mathbf{0}$ 并且 $\text{Proj}_\Omega(x^* - \alpha \cdot E(x^*)) = \text{Proj}_\Omega(x^*) = x^*$. 因此, $x^* \in \Omega^e$.

(ii) 如果 $x^e \in \Omega^e$ 并且 x^e 为集合 Ω 的内点, 根据 $x^e = \text{Proj}_\Omega(x^e - \alpha \cdot E(x^e))$, 我们有 $\cdot E(x^e) = \mathbf{0}$. 又由于 $\cdot \Phi(x^e)$ 非奇异, 可得 $\Phi(x^e) = \mathbf{0}$. 故有 $x^e \in \Omega^*$.

当 $x^* \in \Omega^*$ 并且 $\cdot \Phi(x^*)$ 非奇异时, 我们称 x^* 为互补问题 (17) 的正则解. 下面, 我们将证明正则解的 Liapunov 渐进稳定性.

定理 5 若 x^* 为集合 Ω 的内点并且 x^* 为互补问题 (17) 的正则解, 则 x^* 为 RNN(A) 的 Liapunov 渐进稳定解.

证明 首先, 我们证明 x^* 为 RNN(A) 的孤立平衡点. 由于 $\Phi(x)$ 为 \mathcal{B} 到 \mathcal{B} 的连续可微映射, $\Phi(x) = \det \cdot \Phi(x^*)$ 为 \mathcal{B} 上的连续函数, 因此集合 $A = \{x \in \mathcal{B}; |\Phi(x)| > 0\}$ 为 \mathcal{B} 中的开集. 如果 $x^* \in \text{int } \Omega \cap A$, 则存在 x^* 的邻域, 记作 A^* , 使得 $A^* \subset A \cap \Omega$. 这样, 对任意的 $x \in A^*$, $\cdot \Phi(x)$ 非奇异. 若 $x \in A^*$ 且 $x \neq x^*$, 则存在 $x \in A^*$ 使得 $\Phi(x) = \Phi(x^*) + \cdot \Phi(x)^T \cdot (x - x^*) = \cdot \Phi(x)^T \cdot (x - x^*)$. 既然 $x - x^* \neq \mathbf{0}$ 并且 $\cdot \Phi(x)$ 非奇异, 我们有 $\Phi(x) \neq \mathbf{0}$, 即 $x \notin \Omega^*$. 根据命题 2(ii), 对任意的 $x \in A^*$ 且 $x \neq x^*$ 我们有 $x \in \Omega^e$, 这样, x^* 为 RNN(A) 的孤立平衡点.

其次, 我们证明 $E(x)$ 为 RNN(A) 在 A^* 上的 Liapunov 函数. 根据集合 A^* 的性质, 我们有 $E(x^*) = 0$, 并且任意的 $x \in A^*$ 且 $x \neq x^*$, $E(x) > 0$. 我们只需证 $dE(x(t))/dt \leq 0$ 于 A^* 并且 $dE(x(t))/dt = 0$ 当且仅当 $x(t) = x^*$. 根据定理 1 的证明式 (2), 我们有

$$\frac{dE(x(t))}{dt} \leq \frac{1}{\tau\alpha} \|x(t) - \text{Proj}_\Omega(x(t) - \alpha \cdot E(x(t)))\|^2 \leq 0.$$

既然 x^* 是 RNN(A) 的孤立平衡点, 我们有

$$\frac{dE(x(t))}{dt} < 0, \quad \text{对所有 } x(t) \in A^* \text{ 和 } x(t) \neq x^*.$$

这样, x^* 是 RNN(A) 的 Liapunov 渐进稳定解, 该定理证毕.

3 数值实验与模拟

本节中, 我们给出一些闭凸约束连续可微最优化问题和非线性互补问题的例子, 并用前面讨论的神经网络模型解决它们(为简单起见, 取 $\alpha = \tau = 1$).

闭凸连续可微优化问题的例子

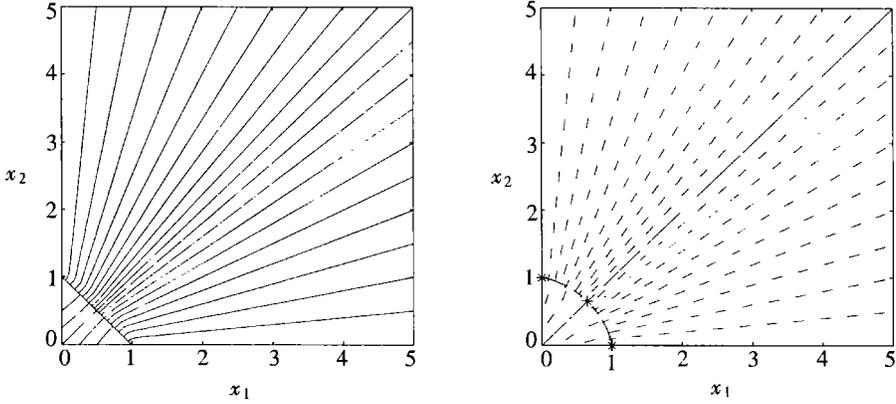
例 1 Minimize

$$E(x) = e^{x_1^2 + x_2^2 - 1} \arctan(x_1 + x_2 - 1)^2, \quad (25)$$

约束条件为

$$\mathbf{x} \in \Omega = \{ \mathbf{x} \in \mathcal{R}^2 \mid x_1 \geq 0, x_2 \geq 0 \}. \tag{26}$$

这是一个非凸连续可微优化, 其约束集合为无界的闭凸集合, 我们可得: $\Omega^* = \Omega^e = \{ \mathbf{x} \in \Omega \mid x_1 + x_2 = 1 \}$ 以及 $\lim_{x_i \rightarrow +\infty} \|F(\mathbf{x})\| = +\infty, i = 1, 2$. 因此定理3(i)的条件成立, 从任意初始点 $\mathbf{x}_0 \in \Omega$ 出发的解轨道 $\mathbf{x}(t)$ 收敛到 Ω^e . 见图1(a).



(a) 例 1

(b) 例 2

图 1 闭凸约束的连续可微最优化问题的数值模拟

(x_1, x_2 分别表示解轨道的两个坐标)

例 2 Minimize

$$E(\mathbf{x}) = e^{x_1^2 + x_2^2 - 1} \arctan(x_1 + x_2 - 1)^2 + \cos(x_1^2 + x_2^2) \pi, \tag{27}$$

约束条件为

$$\mathbf{x} \in \Omega = \{ \mathbf{x} \in \mathcal{R}^2 \mid x_1 \geq 0, x_2 \geq 0 \}. \tag{28}$$

我们有 $\Omega^* = \{ (0, 1), (1, 0) \}$ 以及 $\Omega^e = \Omega^* \cup \{ 0.6503, 0.6503 \}$. 根据定理 3(ii), RNN(A) 对任意的初始点 $\mathbf{x}_0 \in \Omega$ 是整体收敛的. 图 1(b) 给出了神经网络解轨道的收敛情况.

从这两个例子里, 可以看出从任意的初始点 $\mathbf{x}_0 \in \Omega$, RNN(A) 都可以迅速地收敛到(A)式的平衡点集. 因此, RNN(A) 是解决这类最优化问题的有效方法.

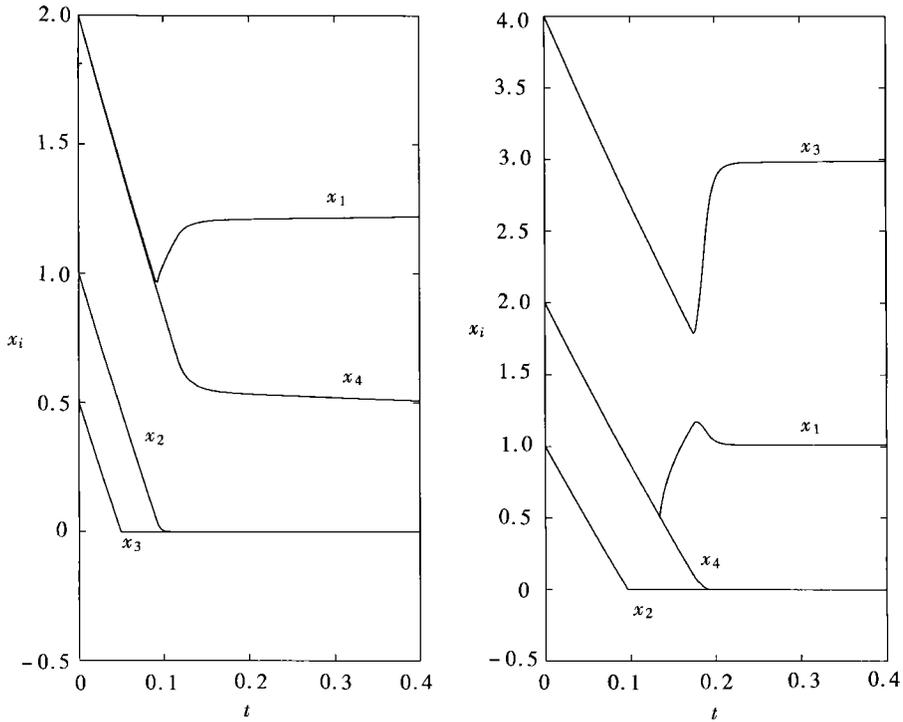
非线性互补问题的例子(Kojima_Shindo nonlinear complementarity test problem^[15])

例 3 退化例子

$$\mathbf{g}(\mathbf{x}) = \mathbf{x}, \mathbf{F}(\mathbf{x}) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

这个问题有两个解 $\mathbf{x}^* = (\sqrt{6}/2, 0, 0, 1/2)$ 和 $\mathbf{x}^{**} = (1, 0, 3, 0)$, 第 1 个解是退化的(由于 $\mathbf{F}(\mathbf{x}^*) = (0, 2 + \sqrt{6}/2, 0, 0)$), 第 2 个解是非退化的(由于 $\mathbf{F}(\mathbf{x}^{**}) = (0, 31, 0, 4)$). 图 2 给出了用神经网络模型解此问题的解轨道的收敛情况.

例 4 非退化例子



(a) 初始点为(2, 1, 0.5, 2), 收敛到解($\sqrt{6}/2, 0, 0, 1/2$) (b) 初始点为(2, 1, 4, 2), 收敛到解(1, 0, 3, 0)

图2 退化问题例子解轨道的收敛情况

$$g(x) = x, F(x) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 3x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

这个问题只有一个非退化解 $x^* = (\sqrt{6}/2, 0, 0, 1/2)$ (由于 $F(x^*) = (0, 2 + \sqrt{6}/2, 5, 0)$).

图3给出了用神经网络模型解此问题的解轨道的收敛情况.

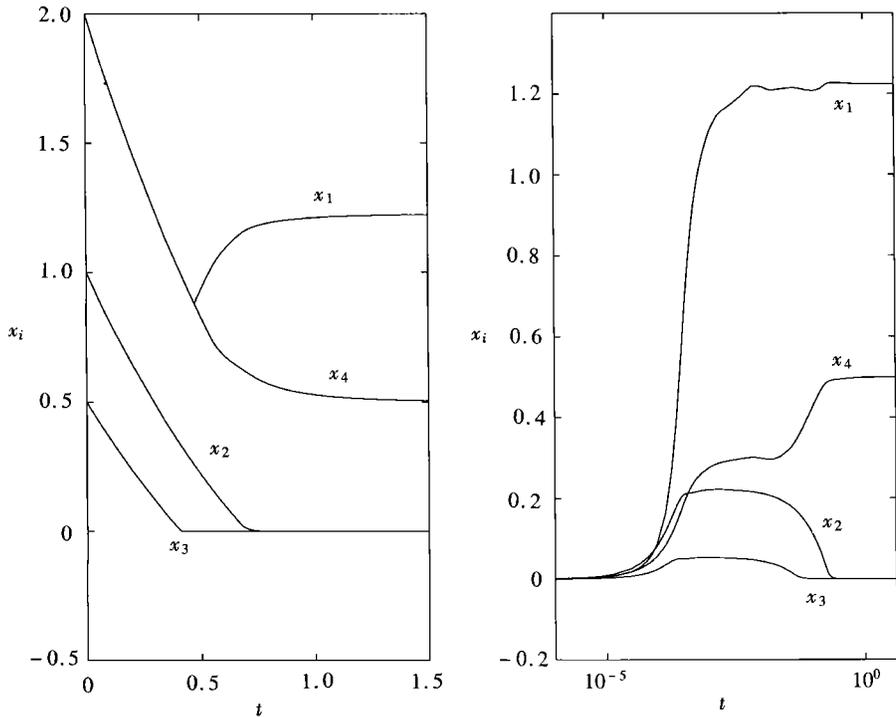
从这些例子中都可以看出, 我们所考虑的神经网络模型从任意的初始点 $x_0 \in \Omega$ 出发, 都可以迅速地收敛到平衡点集, 这也说明了该方法解决非线性互补问题的有效性.

4 结 论

本文研究了基于投影算子的回归神经网络模型, 其值域为 \mathcal{R} 中一般的闭凸子集. 我们证明了该神经网络模型具有整体解轨道, 以及当目标函数满足某些条件时解轨道的整体收敛性. 这样, 我们就将文献[8]中关于紧凸约束连续可微优化问题的结论和文献[9]中关于闭凸约束连续可微凸优化问题的结论, 推广到闭凸约束(未必有界)连续可微优化(未必凸)的问题中去. 此外, 我们还将所研究的模型应用于闭凸约束极小化问题以及非线性互补问题和隐互补问题中, 并证明了隐互补问题的正则解的 Liapunov 渐进稳定性. 最后, 我们通过具体的例子进行数值模拟, 说明了该神经网络方法的有效性.

[参 考 文 献]

[1] Hopfield J J, Tank D W. Neural computation of decision in optimization problem[J]. Biol Cybern ,



(a) 初始点为 $(2, 1, 0.5, 2)$, 收敛到唯一解 $(\sqrt{6}/2, 0, 0, 0.5)$ (b) 初始点为 $(0, 0, 0, 0)$, 收敛到唯一解 $(\sqrt{6}/2, 0, 0, 0.5)$

图 3 非退化问题例子解轨道的收敛情况

1985, **52**(1): 141—152.

- [2] Tank D W, Hopfield J J. Simple 'neural' optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit[J]. IEEE Trans Circuits Syst (I), 1988, **35**(5): 554—562.
- [3] Bouzerdoum A, Pattison T R. Neural network for quadratic optimization with bound constraints[J]. IEEE Transactions on Neural Networks, 1993, **4**(2): 293—303.
- [4] Perez Ilzarbe M J. Convergence analysis of a discrete time recurrent neural network to perform quadratic real optimization with bound constraints[J]. IEEE Transactions on Neural Networks, 1998, **9**(6): 1344—1351.
- [5] Liang X B, Wang J. A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints[J]. IEEE Transactions on Neural Networks, 2000, **11**(6): 1251—1262.
- [6] XIA You_shen, Leung Henry, WANG Jun. A projection neural network and its application to constrained optimization problems[J]. IEEE Trans Circuits Syst (I), 2002, **49**(4): 447—458.
- [7] XIA You_shen, WANG Jun. A recurrent neural network for solving linear projection equations[J]. Neural Networks, 2000, **13**(3): 337—350.
- [8] Liang X B. A recurrent neural network for nonlinear continuously differentiable optimization over a compact convex subset[J]. IEEE Transactions on Neural Networks, 2001, **12**(6): 1487—1490.
- [9] Liang X B. Qualitative analysis of a recurrent neural network for nonlinear continuously differentiable convex minimization over a nonempty closed convex subset[J]. IEEE Transactions on Neural Networks, 2001, **12**(6): 1521—1525.

- [10] Kinderlehrer D, Stampcchia G. An Introduction to Variational Inequalities and Their Applications [M]. New York Academic, 1980.
- [11] Courant R, John F. Introduction to Calculus and Analysis [M]. Vol 1. New York: Wiley, 1989.
- [12] Fischer A. An NCP_function and its use for the solution of complementarity problems[A] In: Du D, Qi L, Womersley R, Eds. Recent Advance in Nonsmooth Optimization [C]. New Jersey: World Scientific Publishers, 1995, 88—105.
- [13] Fischer A. New constrained optimization reformulation of complementarity problems[J]. Journal of Optimization Theory and Applications, 1998, **97**(1): 105—117.
- [14] Qi H, Liao L. A smoothing Newton method for general nonlinear complementarity problems[J]. Computational Optimization Applications, 2000, **17**(2/3): 231—253.
- [15] Kojima M, Shindo S. Extensions of Newton and quasi_Newton methods to systems of PC_1 equations [J]. J Oper Res Soc Jpn, 1986, **29**: 352—374.

Recurrent Neural Network Model Based on Projective Operator and Its Application to Optimization Problems

MA Ru_ning^{1,2}, CHEN Tian_ping²

(1. Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, P. R. China ;

2. Institute of Mathematics, Fudan University, Shanghai 200433, P. R. China)

Abstract: The recurrent neural network(RNN) model based on projective operator is studied. Different from the former study, the value region of projective operator in the neural network which they study was a general closed convex subset of n dimensional Euclidean space and it wasn't a compact convex set in general, that is, the value region of projective operator was probably unbounded. They prove that the network has a global solution and its solution trajectory converges to some equilibrium set whenever objective function satisfies some conditions. After that, the model was applied to continuously differentiable optimization and nonlinear or implicit complementarity problems. In addition, simulation experiments confirm the efficiency of the RNN.

Key words: recurrent neural network model; projective operator; global convergence; optimization; complementarity problem