

# 解最小一乘问题的递归神经网络

李智勇

(集美大学理学院, 福建 厦门 361021)

[摘要] 应用鞍点理论和投影算子的性质, 给出了一种递归神经网络求解具有线性约束的最小一乘问题, 证明了此神经网络全局收敛于一个最优解. 数值实验表明, 用本文的方法求解最小一乘问题是切实可行的.

[关键词] 递归神经网络; 最小一乘问题; 线性约束

[中图分类号] O 221; TP 181

[文献标志码] A

## A Recurrent Neural Network for Solving Least Absolute Deviation Problem

LI Zhi-yong

(School of Science, Jimei University, Xiamen 361021, China)

**Abstract:** By using the saddle theorem and the properties of projection mapping, a recurrent neural network is proposed for solving least absolute deviation with linear constraints. It is shown that the proposed neural network is globally convergent to an optimal solution. The example given in the paper demonstrates that the proposed approach provides a promising alternative for solving least absolute deviation problem.

**Key words:** recurrent neural network; least absolute deviation problem; linear constraints

## 0 引言

本文主要讨论具有线性约束的最小一乘问题:

$$\begin{aligned} \min \quad & \|Dx - d\|_1, \\ \text{s. t. } \quad & p \leq Ax \leq q, l \leq x \leq h, \end{aligned} \quad (1)$$

其中:  $D \in \mathbf{R}^{m \times n}$ ;  $x, l, h \in \mathbf{R}^n$ ;  $d \in \mathbf{R}^m$ ;  $A \in \mathbf{R}^{r \times n}$ ;  $p, q \in \mathbf{R}^r$ ;  $\|d\|_1 = \sum_{i=1}^m |d_i|$ .

在线性回归模型中, 经常用最小二乘估计来估计参数的值. 但是, 当个别异常点有较大偏离时, 其误差的平方比其误差的绝对值要大得多. 所以最小二乘估计的鲁棒性不如最小一乘估计. 因此最小一乘估计被广泛地应用到线性回归和工程领域, 尤其是信号和图像处理领域<sup>[1-4]</sup>. 但是, 由于最小一乘问题的目标函数不是光滑的, 所以求解最小一乘问题是比较复杂的. 因此研究求解它的算法是有必要和有意义的. 有时人们需要实时求解最小一乘问题, 但是经典的数值算法如下降算法<sup>[5]</sup>、线性规划方法<sup>[6]</sup>都很难做到这一点. 由于神经网络计算具有并行计算和实时求解的特点, 因此, 文献 [3-4]、[7-9] 提出用神经网络的方法来求解, 它们的主要内容都是用递归神经网络的方法求解一些具有线性约束的最小一乘问题, 都是问题 (1) 的特殊情况, 所以问题 (1) 具有一般性. 文献 [3]、[7] 不能求解问题 (1). 必须对问题 (1) 进行转化后, 文献 [4]、[8-9] 才能求解问题 (1),

但是, 这将导致问题的规模变大, 计算效率降低. 本文应用与文献 [9] 类似的技巧, 应用鞍点理论和投影算子的性质, 给出了一种具有全局收敛的求解问题 (1) 的递归神经网络.

## 1 预备知识

定义 1<sup>[10]</sup> 设  $\Omega_0 \subset \mathbf{R}^n$  是闭的凸集,  $\|\mathbf{x}\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$ , 若  $P_{\Omega_0}(\mathbf{x}) = \arg \min_{\mathbf{v} \in \Omega_0} \|\mathbf{x} - \mathbf{v}\|_2$ , 则称  $P_{\Omega_0}(\cdot)$  是  $\Omega_0$  上的一个投影算子.

引理 1<sup>[10]</sup> 设  $\Omega_0 \subset \mathbf{R}^n$  是闭的凸集,  $P_{\Omega_0}(\cdot)$  是  $\Omega_0$  上的一个投影算子,  $F: \Omega_0 \rightarrow \mathbf{R}^n$  是  $\Omega_0$  上的连续函数, 则: 1) 对任意的  $\mathbf{u} \in \mathbf{R}^n, \mathbf{v} \in \Omega_0$ , 都有  $[\mathbf{u} - P_{\Omega_0}(\mathbf{u})]^T [P_{\Omega_0}(\mathbf{u}) - \mathbf{v}] \geq 0$ ; 2) 对任意的  $\mathbf{u}, \mathbf{w} \in \mathbf{R}^n$ , 都有  $\|P_{\Omega_0}(\mathbf{u}) - P_{\Omega_0}(\mathbf{w})\|_2 \leq \|\mathbf{u} - \mathbf{w}\|_2$ ; 3)  $\mathbf{x}^* = P_{\Omega_0}(\mathbf{x}^* - F(\mathbf{x}^*)) \Leftrightarrow (\mathbf{x} - \mathbf{x}^*)^T F(\mathbf{x}^*) \geq 0, \forall \mathbf{x} \in \Omega_0$ .

## 2 神经网络模型和全局收敛性

易知问题 (1) 等价于如下问题:

$$\begin{aligned} & \min \left\{ \max_{\mathbf{y} \in \Omega} \mathbf{y}^T (\mathbf{D}\mathbf{x} - \mathbf{d}) \right\}, \\ & \text{s. t. } \mathbf{A}\mathbf{x} = \boldsymbol{\zeta}, \mathbf{x} \in \Omega_1, \boldsymbol{\zeta} \in \Omega_2, \end{aligned} \quad (2)$$

其中:  $\Omega = \{\mathbf{y} \mid -1 \leq y_i \leq 1 (i = 1, 2, \dots, m)\}$ ;  $\Omega_1 = \{\mathbf{x} \mid \mathbf{l} \leq \mathbf{x} \leq \mathbf{h}\}$ ;  $\Omega_2 = \{\boldsymbol{\zeta} \mid \mathbf{p} \leq \boldsymbol{\zeta} \leq \mathbf{q}\}$ .

定理 1  $\mathbf{x}^* \in \mathbf{R}^n$  是问题 (1) 的最优解当且仅当存在  $(\mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^m \times \mathbf{R}^r$ , 使得  $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$  满足方程组:

$$\begin{cases} \mathbf{x}^* = P_{\Omega_1}[\mathbf{x}^* - (\mathbf{D}^T \mathbf{y}^* + \mathbf{A}^T \mathbf{z}^*)], \\ \mathbf{y}^* = P_{\Omega}[\mathbf{y}^* + (\mathbf{D}\mathbf{x}^* - \mathbf{d})], \\ \mathbf{A}\mathbf{x}^* = P_{\Omega_2}(\mathbf{A}\mathbf{x}^* + \mathbf{z}^*). \end{cases} \quad (3)$$

其中:  $P_{\Omega_1}(\mathbf{x}) = [P_{\Omega_1}(x_1), \dots, P_{\Omega_1}(x_n)]^T$ ,  $P_{\Omega_1}(x_i) = \begin{cases} l_i, & x_i < l_i \\ x_i, & l_i \leq x_i \leq h_i \\ h_i, & x_i > h_i \end{cases}$ .

证明 令问题 (2) 的拉格朗日函数为  $L(\mathbf{x}, \boldsymbol{\zeta}, \mathbf{y}, \mathbf{z}) = \mathbf{y}^T (\mathbf{D}\mathbf{x} - \mathbf{d}) + \mathbf{z}^T (\mathbf{A}\mathbf{x} - \boldsymbol{\zeta})$ ,  $\mathbf{z} \in \mathbf{R}^r$ . 由著名的鞍点理论<sup>[11]</sup>可知:  $\mathbf{x}^* \in \mathbf{R}^n$  是问题 (1) 的最优解  $\Leftrightarrow (\mathbf{x}^*, \boldsymbol{\zeta}^*) = (\mathbf{x}^*, \mathbf{A}\mathbf{x}^*)$  是问题 (2) 的最优解  $\Leftrightarrow$  存在  $(\mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^m \times \mathbf{R}^r$ , 使得  $(\mathbf{x}^*, \boldsymbol{\zeta}^*, \mathbf{y}^*, \mathbf{z}^*)$  是一个鞍点, 即任给  $\mathbf{x} \in \Omega_1, \boldsymbol{\zeta} \in \Omega_2, \mathbf{y} \in \Omega, \mathbf{z} \in \mathbf{R}^r$ , 有

$$L(\mathbf{x}^*, \boldsymbol{\zeta}^*, \mathbf{y}, \mathbf{z}) \leq L(\mathbf{x}^*, \boldsymbol{\zeta}^*, \mathbf{y}^*, \mathbf{z}^*) \leq L(\mathbf{x}, \boldsymbol{\zeta}, \mathbf{y}^*, \mathbf{z}^*). \quad (4)$$

即

$$\begin{aligned} \mathbf{y}^T (\mathbf{D}\mathbf{x}^* - \mathbf{d}) + \mathbf{z}^T (\mathbf{A}\mathbf{x}^* - \boldsymbol{\zeta}^*) & \leq (\mathbf{y}^*)^T (\mathbf{D}\mathbf{x}^* - \mathbf{d}) + (\mathbf{z}^*)^T (\mathbf{A}\mathbf{x}^* - \boldsymbol{\zeta}^*) \leq \\ & (\mathbf{y}^*)^T (\mathbf{D}\mathbf{x} - \mathbf{d}) + (\mathbf{z}^*)^T (\mathbf{A}\mathbf{x} - \boldsymbol{\zeta}). \end{aligned} \quad (5)$$

下面先证: 若式 (5) 成立, 则方程组 (3) 成立. 由式 (5) 左端的不等式可知:  $\mathbf{A}\mathbf{x}^* = \boldsymbol{\zeta}^*$ , 任给  $\mathbf{y} \in \Omega$  有  $(\mathbf{y}^* - \mathbf{y})^T (\mathbf{D}\mathbf{x}^* - \mathbf{d}) \geq 0$ . 分别取  $\boldsymbol{\zeta} = \boldsymbol{\zeta}^*$  和  $\mathbf{x} = \mathbf{x}^*$ , 则由式 (5) 右端的不等式分别可知: 任给  $\mathbf{x} \in \Omega_1$  有  $(\mathbf{x} - \mathbf{x}^*)^T (\mathbf{D}^T \mathbf{y}^* + \mathbf{A}^T \mathbf{z}^*) \geq 0$ , 任给  $\boldsymbol{\zeta} \in \Omega_2$  有  $(\mathbf{z}^*)^T (\boldsymbol{\zeta} - \boldsymbol{\zeta}^*) \leq 0$ . 由引理 1 可知方程组 (3) 成立.

若方程组 (3) 成立, 应用引理 1 容易得出式 (5) 成立. 定理 1 得证.

为了解问题 (1), 本文用下列递归神经网络模型:

$$\begin{cases} d\mathbf{x}(t)/dt = \lambda \{ \mathbf{E}(t) - \mathbf{D}^T \mathbf{F}_1(t) - \mathbf{A}^T \mathbf{F}_2(t) \}, \\ d\mathbf{y}(t)/dt = \lambda \{ \mathbf{D}\mathbf{E}(t) + \mathbf{F}_1(t) \}, \\ d\mathbf{z}(t)/dt = \lambda \{ \mathbf{A}\mathbf{E}(t) + \mathbf{F}_2(t) \}. \end{cases} \quad (6)$$

输出方程为  $w(t) = P_{\Omega_1}(x(t))$ . 其中  $\lambda > 0$  是一个常数,  $E(t) = P_{\Omega_1}[x(t) - (D^T y(t) + A^T z(t))] - x(t)$ ,  $F_1(t) = P_{\Omega}[y(t) + (Dx(t) - d)] - y(t)$ ,  $F_2(t) = Ax(t) - P_{\Omega_2}(Ax(t) + z(t))$ .

**定理 2** 若问题 (1) 有解, 则网络模型 (6) 将全局收敛于问题 (1) 的一个最优解.

**证明** 由引理 1 可知  $P_{\Omega_0}(x)$  在  $\mathbf{R}^n$  上满足利普希茨条件而且是连续的, 所以由微分方程的存在性定理可知: 任取一初始点, 方程 (6) 存在唯一的连续解. 设这个解为  $u(t) = (x(t), y(t), z(t))$ , 设  $u^* = (x^*, y^*, z^*)$  是方程组 (3) 的解,  $V(u) = \|u - u^*\|_2^2/2$ , 则  $dV(u(t))/dt = (u(t) - u^*)^T du/dt = \lambda \{ (x(t) - x^*)^T (E(t) - D^T F_1(t) - A^T F_2(t)) + (y(t) - y^*)^T (DE(t) + F_1(t)) + (z(t) - z^*)^T (F_2(t) + AE(t)) \}$ .

由方程组 (3) 和引理 1 可知:

$$[P_{\Omega}(y(t) + Dx(t) - d) - y^*]^T [-(Dx^* - d)] \geq 0, \quad (7)$$

$$[(y(t) + Dx(t) - d) - P_{\Omega}(y(t) + Dx(t) - d)]^T [P_{\Omega}(y(t) + Dx(t) - d) - y^*] \geq 0. \quad (8)$$

式 (7) 和式 (8) 相加后整理可得:  $-(x(t) - x^*)^T D^T F_1(t) + (y(t) - y^*)^T F_1(t) \leq -\|F_1(t)\|_2^2 + (y(t) - y^*)^T D(x(t) - x^*)$ .

由方程组 (3) 和引理 1 可知:

$$(P_{\Omega_2}(Ax(t) + z(t)) - Ax^*)^T (-z^*) \geq 0, \quad (9)$$

$$(Ax(t) + z(t) - P_{\Omega_2}(Ax(t) + z(t)))^T (P_{\Omega_2}(Ax(t) + z(t)) - Ax^*) \geq 0. \quad (10)$$

式 (9) 和式 (10) 相加后整理可得:  $-(x(t) - x^*)^T A^T F_2(t) + (z(t) - z^*)^T F_2(t) \leq -\|F_2(t)\|_2^2 + (x(t) - x^*)^T A^T (z(t) - z^*)$ .

由方程组 (3) 和引理 1 可知:

$$[P_{\Omega_1}(x(t) - (D^T y(t) + A^T z(t))) - x^*]^T (D^T y^* + A^T z^*) \geq 0, \quad (11)$$

$$[x(t) - (D^T y(t) + A^T z(t)) - P_{\Omega_1}(x(t) - (D^T y(t) + A^T z(t)))]^T [P_{\Omega_1}(x(t) - (D^T y(t) + A^T z(t))) - x^*] \geq 0. \quad (12)$$

式 (11) 和式 (12) 相加后整理可得:

$$(x(t) - x^*)^T E(t) + (y(t) - y^*)^T DE(t) + (z(t) - z^*)^T AE(t) \leq -\|E(t)\|_2^2 - (y(t) - y^*)^T D(x(t) - x^*) - (x(t) - x^*)^T A^T (z(t) - z^*). \quad (13)$$

所以,

$$dV(u(t))/dt \leq \lambda \{ -\|E(t)\|_2^2 - \|F_1(t)\|_2^2 - \|F_2(t)\|_2^2 \} \leq 0. \quad (14)$$

所以若在  $\bar{u}$  处  $dV/dt = 0 \Rightarrow E(t) = F_1(t) = F_2(t) = 0 \Rightarrow \bar{u}$  是方程组 (3) 的解.

由拉萨尔不变性原理可知: 设  $E = \{u \mid dV(u(t))/dt = 0, \|u - u^*\|_2^2 \leq \|u_0 - u^*\|_2^2\}$ ,  $M$  是  $E$  关于方程 (6) 的最大不变子集, 则存在  $\tilde{u} \in M$ , 数列  $\{t_n\}$ , 使得  $\lim_{n \rightarrow \infty} \|u(t_n) - \tilde{u}\| = 0$ ,  $\lim_{n \rightarrow \infty} t_n = \infty$ . 因为  $\tilde{u} \in M$ , 所以  $\tilde{u}$  是方程组 (3) 的解. 若把证明过程中的  $u^*$  换成  $\tilde{u}$ , 则式 (14) 仍然成立. 所以  $V(u(t)) = \|u(t) - \tilde{u}\|_2^2/2$  关于  $t$  是单调递减的. 所以当  $t > t_n$  时,  $\|u(t) - \tilde{u}\|_2^2/2 \leq \|u(t_n) - \tilde{u}\|_2^2/2$ . 所以  $\lim_{t \rightarrow \infty} \|u(t) - \tilde{u}\| = 0$ . 定理 2 证毕.

### 3 数值实验

例 1 考虑最小一乘问题 (1), 其中  $D = \begin{bmatrix} 5 & 9 & 6 & 9 & 3 & 8 & 1 & 3 & 0 \\ 3 & 7 & 6 & 9 & 0 & 1 & 9 & 1 & 9 \\ 4 & 3 & 0 & 7 & 1 & 8 & 8 & 1 & 3 \\ 12 & 19 & 12 & 25 & 4 & 17 & 18 & 5 & 12 \\ 4 & 13 & 12 & 11 & 2 & 1 & 2 & 3 & 6 \end{bmatrix}$ ,

$$A = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$
$$p = \begin{bmatrix} -100 \\ 1 \end{bmatrix}, q = \begin{bmatrix} 7.5 \\ 1 \end{bmatrix}, d = [45 \ 46 \ 36 \ 125 \ 55]^T,$$
$$l = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T,$$
$$h = [2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2]^T.$$

例1与文献[9]的例1类似. 因为在 $D$ 的各列的和中, 第四列的和最大, 所以在理论上, 可以得出此例的最优解为 $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ . 取 $\lambda = 10$ , 取初始点 $x(0) = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$ , 用MATLAB 7.0 求解网络模型(6)可得解为 $[0.0000 \ 0 \ 0 \ 1.0000 \ 0.0000 \ 0 \ 0 \ 0.0000 \ 0.0000]^T$ . 图1显示了神经网络(6)收敛于最优解的轨线状态. 由此可见, 用本文所提的方法求解问题(1)是可行的.

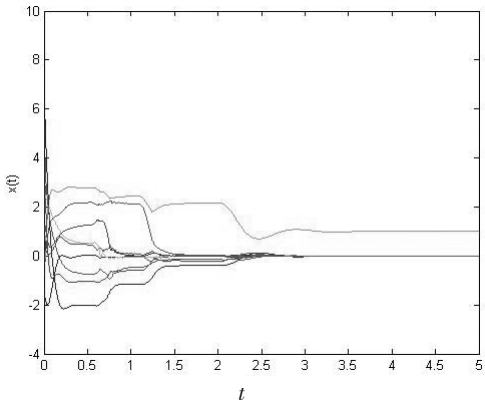


图1 例题中基于神经网络(6)的 $x(t)$ 的轨线状态  
Fig.1 Convergence behavior of the state trajectory  $x(t)$  based on the neural network(6) in example

[ 参 考 文 献 ]

[1] KUO S S, MAMMONE R J. Image restoration by convex projections using adaptive constraints and the  $L_1$  norm [J]. IEEE Trans Signal Processing, 1992, 40(1): 159-168.

[2] 占美全, 邓志良. 基于  $L_1$  范数的总变分正则化超分辨率图像重建 [J]. 科学技术与工程, 2010, 10(28): 6903-6906.

[3] XIA Y S, KAMEL M S. Novel cooperative neural fusion algorithms for image restoration and image fusion [J]. IEEE Trans Image Process, 2007, 16(2): 367-381.

[4] XIA Y S, SUN C Y, ZHENG W X. Discrete-time neural network for fast solving large linear  $L_1$  estimation problems and its application to image restoration [J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(5): 812-820.

[5] BARTELS R H, CONN A R, SINCLAIR J W. Minimization techniques for piecewise differentiable functions: the  $L_1$  solution to an overdetermined linear system [J]. SIAM J Numer Anal, 1978, 15(2): 224-241.

[6] RUZINSKY S A, OLSEN E T.  $L_1$  and  $L_\infty$  minimization via a variant of Karmarkar's algorithm [J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1989, 37(2): 245-253.

[7] XIA Y S, KAMEL M S. Cooperative recurrent neural networks for the constrained  $L_1$  norm estimator [J]. IEEE Trans Signal Process, 2007, 55(7): 3192-3205.

[8] XIA Y S, KAMEL M S. A cooperative recurrent neural network for solving  $L_1$  estimation problems with linear constraints [J]. Neural Comput, 2008, 20(3): 844-872.

[9] XIA Y S. A compact cooperative recurrent neural network for computing general constrained  $L_1$  norm estimators [J]. IEEE Trans Signal Process, 2009, 57(9): 3693-3697.

[10] KINDERLEHRER D, STAMPACCHIA G. An introduction to variational inequalities and their applications [M]. New York: Academic Press, 1980.

[11] BERTSEKAS D P, NEDIC A, OZDAGLAR A E. Convex analysis and optimization [M]. 北京: 清华大学出版社, 2006.

(责任编辑 马建华 英文审校 黄振坤)