

一种新型的 Lagrange 非线性规划神经网络

黄远灿

(北京理工大学自动控制系, 北京 100081)

摘 要: 通过改造 Lagrange 乘子项,对不等式约束和等式约束用同样的方法处理,构造出一种新型的 Lagrange 非线性规划神经网络,并对网络的稳定性和收敛性进行严格的理论分析.而且,用增加惩罚项的方法,放宽网络收敛的条件.

关键词: 神经优化;非线性规划;稳定性

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2002) 01-0027-03

A New Type of Lagrange Nonlinear Programming Neural Network

HUANG Yuan-can

(Department of Automatic Control, Beijing Institute of Technology, Beijing 100081, China)

Abstract: By modifying the Lagrange multipliers, both equality and inequality constraints can be treated identically. Based on this idea, a new type of Lagrange nonlinear programming neural network is constructed, and the stability and convergence of the neural network are analyzed rigorously. Furthermore, the convergence condition of the network is relaxed by the approach of adding the penalty terms in the Lagrangian function.

Key words: neural optimization; nonlinear programming; stability

1 引言

自从 Hopfield 提出神经优化的思想以来^[1,2], 有很多人对此问题进行了深入研究. 文献[3,4]介绍了用神经网络求解线性规划和非线性规划问题的方法以及电路实现. 然而, 文献[4]只讨论了不等式约束的非线性规划问题. 尽管在数学上等式约束可用两个耦合的不等式代替, 但应用文献[4]的方法处理等式约束问题, 其电路由两个反向的二极管耦合而成, 不能正常工作. 另外, 文献[4]中的方法基于罚函数法, 如果没有选取罚因子和初始点的信息, 罚函数法在理论上必须通过极小化一系列罚因子逐渐增加的无约束问题来求问题的解, 或采用试凑法确定罚因子和初始点, 这两种方法在实际中都不可行. 而且, 罚函数法总会碰到病态问题^[7]. 针对这些缺点, 文献[5]提出用 Lagrange 乘子法构造神经网络解决一般非线性规划问题. 文献[6]通过改造 Lagrange 乘子项的方法直接处理不等式约束, 不用增加“松弛变量”, 可减少网络的复杂程度和实现难度. 本文根据文献[6]提出的方法, 使用数学规划理论, 严格地分析如何处理不等式约束的问题, 并用在 Lagrange 函数中增加惩罚项的方法, 放宽网络收敛的条件.

2 非线性规划神经网络

考虑如下的一般非线性规划问题 (GNP):

$$\begin{aligned} \min_{x \in X} f(x) \\ \text{(GNP)} \quad X = \{x \mid h(x) = 0, g(x) \leq 0\} \end{aligned} \quad (1)$$

式中, $x \in R^n$, $f: R^n \rightarrow R$, $h: R^n \rightarrow R^m$ 和 $g: R^n \rightarrow R^r$ 是给定的函数, 且 $m \leq n$. 假设 $h(x)$ 和 $g(x)$ 二次连续可微.

定义 1: 设 x^* 满足约束条件, 且 $I(x^*)$ 是由 $g_i(x^*) = 0$ 的 i 组成的指标集, 即

$$I(x^*) = \{i \mid g_i(x^*) = 0, i = 1, 2, \dots, r\}$$

如果梯度向量 $\nabla h_j(x^*)$, $1 \leq j \leq m$, $\nabla g_i(x^*)$, $i \in I(x^*)$ 线性无关, 则称 x^* 是正则点.

定义 2: Lagrange 函数 $L: R^{n+m+r} \rightarrow R$ 定义为

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu_+^T g(x) \quad (2)$$

式中, $\mu \in R^r$ 和 $\lambda \in R^m$ 是 Lagrange 乘子, $\mu_+ = (\mu_1^2/2, \mu_2^2/2, \dots, \mu_r^2/2)^T$.

这里需要说明的是, 我们定义的 Lagrange 函数与经典优化理论不同, 将不等式约束的乘子取为 $\mu_+ = (\mu_1^2/2, \mu_2^2/2, \dots, \mu_r^2/2)^T$, 当然, 也可以取为分量的其它正定可微函数. 这样做的目的是避免对不等式约束乘子附加非负约束, 直接对不等式约束进行处理. 如果将这些正定函数看成经典优化理论中不等式约束的乘子, 那么相关的最优性条件都可以移植过来.

与经典优化理论相似, 有以下定理成立^[6,7].

令

$$\begin{aligned} Z(x^*) = \{z \mid z^T \nabla g_i(x^*) = 0, i \in I(x^*), \\ z^T \nabla h_j(x^*) = 0, j = 1, 2, \dots, m\} \end{aligned}$$

定理 1: 假设 x^* 是 GNP 的局部极小点, 且 x^* 是正则点,

那么存在 λ^*, μ^* , 使得下式成立:

$$\begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= 0 \\ \mu_i^* g_i(x^*) &= 0, i=1, \dots, r \end{aligned} \quad (3)$$

且对任何 $z \neq 0, z \in Z(x^*)$, 有

$$z^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z \geq 0 \quad (4)$$

定义 3: 称满足定理 1 的 (x^*, λ^*, μ^*) 为 GNP 的 Kuhn-Tucker 点.

定理 2: 假设 x^* 满足约束条件, 存在 λ^*, μ^* , 使得下式成立:

$$\begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= 0 \\ \mu_i^* g_i(x^*) &= 0, i=1, \dots, r \end{aligned} \quad (5)$$

且满足严格互补性条件, 即, $\mu_i^* \neq 0, \forall i \in I(x^*)$. 如果对任何 $z \neq 0, z \in Z(x^*)$, 有

$$z^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z > 0 \quad (6)$$

那么 x^* 是 GNP 的局部严格极小点.

我们的目的是, 构造收敛于 GNP 的 Kuhn-Tucker 点的神经网络, 其暂态行为由下列方程决定:

$$\begin{cases} \frac{dx}{dt} = -\nabla_x L(x, \lambda, \mu) \\ \frac{d\lambda}{dt} = \nabla_{\lambda} L(x, \lambda, \mu) \\ \frac{d\mu}{dt} = \nabla_{\mu} L(x, \lambda, \mu) \end{cases} \quad (7)$$

按照在求解过程中的作用不同, 可将网络的神经元分为变量神经元 x 和 Lagrange 神经元 λ, μ . Lagrange 神经元使网络的轨迹落入可行区域内, 而变量神经元使 Lagrange 函数减少.

定义 4: 如果对任何 $x \in R^n, \lambda \in R^m, \mu \in R^r$, 有

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \quad (8)$$

成立, 则称 Lagrange 函数 $L(x, \lambda, \mu)$ 在 (x^*, λ^*, μ^*) 具有鞍点性质, 且 (x^*, λ^*, μ^*) 称为 $L(x, \lambda, \mu)$ 的鞍点.

定理 3: 设 (x^*, λ^*, μ^*) 为 GNP 的 Lagrange 函数的鞍点, 则 x^* 是 GNP 的最优解. 换句话说, Lagrange 函数的鞍点性质是最优性的充分条件^[8].

文中定义的网络类似于梯度系统. 当 λ, μ 保持常值时, $L(x, \lambda, \mu)$ 随 x 减小; 当 x 保持常值时, 随 λ, μ 增大. 因此, 网络的稳定点 (x^*, λ^*, μ^*) 就是 Lagrange 函数的鞍点.

3 收敛性和稳定性分析

定理 4: 设 (x^*, λ^*, μ^*) 是 GNP 的 Kuhn-Tucker 点, x^* 是正则点, 且严格互补性条件满足, $\nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) > 0$, 那么 (x^*, λ^*, μ^*) 是网络的渐近稳定点.

证明: 根据一次近似稳定原理, 在 (x^*, λ^*, μ^*) 处将系统 (7) 线性化, 那么平衡点的局部性质由线性化系统决定. 由定理 1, $g_i(x^*) \neq 0$, 有 $\mu_i^* = 0$. 从式 (7) 看出, 在 μ_i^* 的邻域, μ_i 对其它神经元没有影响, 构成一个稳定的自闭环. 因此, 可以将其删除而不影响结果. 令 μ 是由 $\mu_i, i \in I(x^*)$ 组成的向量, 定义 $\mu^s = (\mu_1, \mu_2, \dots, \mu_s)^T$, 其中, s 是 $I(x^*)$ 中指标的个数. $\nabla g^s(x^*) = (\mu_1^* \nabla g_1(x^*), \mu_2^* \nabla g_2(x^*), \dots, \mu_s^* \nabla g_s(x^*))$. 考虑到 $\nabla_x L(x^*, \lambda^*, \mu^*) = 0, h(x^*) = 0, g_i(x^*) =$

$0, i \in I(x^*)$, 则线性化系统为

$$\begin{aligned} \begin{bmatrix} \frac{dx}{dt} \\ \frac{d\lambda}{dt} \\ \frac{d\mu}{dt} \end{bmatrix} &= - \begin{bmatrix} \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) & \nabla h(x^*) & \nabla g^s(x^*) \\ -\nabla h(x^*)^T & 0 & 0 \\ -\nabla g^s(x^*)^T & 0 & 0 \end{bmatrix} \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \mu^s - (\mu^s)^* \end{bmatrix} \\ &= G \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \mu^s - (\mu^s)^* \end{bmatrix} \end{aligned} \quad (9)$$

用 v^H 表示复向量 v 的共轭转置, $\text{Re}(y)$ 表示复数 y 的实部. 令 β 是 G 的特征值, $(z, w, y) \neq (0, 0, 0)$ 为相应的特征向量, 于是

$$\begin{aligned} \text{Re} \left\{ \begin{bmatrix} z^H & w^H & y^H \end{bmatrix} G \begin{bmatrix} z \\ w \\ y \end{bmatrix} \right\} &= \text{Re} \left\{ \beta \begin{bmatrix} z^H & w^H & y^H \end{bmatrix} \begin{bmatrix} z \\ w \\ y \end{bmatrix} \right\} \\ &= \text{Re}(\beta) (|z|^2 + |w|^2 + |y|^2) \end{aligned} \quad (10)$$

又

$$\begin{aligned} \text{Re} \left\{ \begin{bmatrix} z^H & w^H & y^H \end{bmatrix} G \begin{bmatrix} z \\ w \\ y \end{bmatrix} \right\} &= \text{Re} \{ z^H \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z - w^H \nabla h(x^*) z \\ &\quad - y^H \nabla g^s(x^*) z + z^H \nabla h(x^*) w + z^H \nabla g^s(x^*) y \} \end{aligned} \quad (11)$$

由于 $\text{Re}(w^H \nabla h(x^*) z) = \text{Re}(z^H \nabla h(x^*) w)$ 和 $\text{Re}(y^H \nabla g^s(x^*) z) = \text{Re}(z^H \nabla g^s(x^*) y)$, 所以

$$\text{Re}(\beta) (|z|^2 + |w|^2 + |y|^2) = \text{Re} \{ z^H \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z \} > 0, \quad \forall z \neq 0$$

根据上式推出, 或者 $\text{Re}(\beta) > 0$, 或者 $z = 0$. 若 $z = 0$, 由

$$G \begin{bmatrix} z & w & y \end{bmatrix}^T = \beta \begin{bmatrix} z & w & y \end{bmatrix}^T$$

$$\text{有} \quad \begin{bmatrix} \nabla h(x^*) & \nabla g^s(x^*) \end{bmatrix} \begin{bmatrix} w \\ y \end{bmatrix} = 0$$

又因 $\begin{bmatrix} \nabla h(x^*) & \nabla g^s(x^*) \end{bmatrix}$ 的秩为 $m+s$, 所以 $w=0, y=0$, 与假设 $(z, w, y) \neq (0, 0, 0)$ 矛盾, 于是 $\text{Re}(\beta) > 0$. 因此, (x^*, λ^*, μ^*) 是网络的渐近稳定点.

与定理 2 的充分条件相比, 上述定理给出的网络局部渐近收敛的条件要求 $\nabla_{xx}^2 L(x, \lambda, \mu)$ 在 (x^*, λ^*, μ^*) 点正定. 如果增加惩罚项, 只要满足定理 2 的充分条件, 网络可获得同样的收敛性质. 为了证明这个断言, 先给出下列引理^[7]:

引理: 令 P 是 $n \times n$ 对称阵, Q 是 $n \times n$ 正半定对称阵. 若 $\forall x \neq 0, x^T P x > 0$ 满足 $x^T Q x = 0$, 那么存在标量 $c > 0$, 使

$$P + cQ > 0$$

我们定义推广 Lagrange 函数 $L_c: R^{n+m+r} \rightarrow R$ 为

$$\begin{aligned} L_c(x, \lambda, \mu) &= f(x) + \lambda^T h(x) + \mu^T_+ \\ &\quad g(x) + \frac{c}{2} [h(x)^T h(x) + g(x)^T \text{diag}(\mu_1^4/4, \mu_2^4/4, \dots, \mu_r^4/4) g(x)] \end{aligned}$$

其对 x 的梯度和 Hess 矩阵为

$$\begin{aligned} \nabla_x L_c(x, \lambda, \mu) &= \nabla f(x) + \nabla h(x) [\lambda + ch(x)] + \nabla g \mu_+ \\ &\quad + c \nabla g(x) \text{diag}(\mu_1^4/4, \mu_2^4/4, \dots, \mu_r^4/4) g(x) \end{aligned}$$

和

$$\nabla_{xx}^2 L_c(x, \lambda, \mu) = \nabla^2 f(x) + \sum_{j=1}^m [\lambda_j + ch_j(x)] \nabla^2 h_j(x)$$

$$+ c \nabla h(x) \nabla h(x)^T + \sum_{i=1}^r \frac{\mu_i^2}{2} \left[1 + c \frac{\mu_i^2}{2} g_i(x) \right] \nabla^2 g_i(x) \\ + c \nabla g(x) \text{diag}(\mu_1^4/4 \ \mu_2^4/4 \ \cdots \ \mu_r^4/4) \nabla g(x)^T$$

如果 x^* 满足定理 2 的充分条件, 由引理可知, 存在 $\bar{c}$, 使 $\nabla_{xx} L(x^*, \lambda^*, \mu^*) + \bar{c} [\nabla h(x^*) \nabla h(x^*)^T + \nabla g(x^*) \text{diag}((\mu_1^*)^4/4 \ \cdots \ (\mu_r^*)^4/4) \nabla g(x^*)^T] > 0$

因此, 对任意 $c \geq \bar{c}$, 有

$$\nabla_{xx} L_c(x^*, \lambda^*, \mu^*) = \nabla_{xx} L(x^*, \lambda^*, \mu^*) + c [\nabla h(x^*) \nabla h(x^*)^T + \nabla g(x^*) \text{diag}((\mu_1^*)^4/4 \ \cdots \ (\mu_r^*)^4/4) \nabla g(x^*)^T] > 0$$

如果根据增广 Lagrange 函数构造神经网络, 与定理 4 的证明一样, 可得下列定理:

定理 5 如果定理 2 的假设成立, 那么存在 $\bar{c} > 0$, 对任意 $c \geq \bar{c}$, 有 $\nabla_{xx} L_c(x^*, \lambda^*, \mu^*) > 0$, 且 (x^*, λ^*, μ^*) 是网络的渐近稳定点.

4 数值仿真

为了说明上述方法的有效性, 我们对下列例子进行数值仿真:

$$\begin{aligned} \min \quad & x_1 \\ \text{s. t.} \quad & 3(x_1 - 3)^2 + x_2 \geq 0 \\ & (x_1 - 3)^2 + x_2^2 - 10 = 0 \end{aligned}$$

Lagrange 函数为 $L(x, \lambda, \mu) = x_1 - \lambda^2 [3(x_1 - 3)^2 + x_2] + \mu [(x_1 - 3)^2 + x_2^2 - 10]$, $\nabla_{xx} L(x, \lambda, \mu)$ 为 $\begin{bmatrix} -6\lambda^2 + 2\mu & 0 \\ 0 & 2\mu \end{bmatrix}$, 在 Kuhn-Tucker 点 $(3 - \sqrt{10}, 0, 0, 1/2\sqrt{10})^T$ 上正定, 满足定理 4 的条件. 按照式(7)构造网络, 随机选取动态方程的初值, 用四阶/五阶 Runge-Kutta-Fehlberg 法仿真, 结果如图 1 所示.

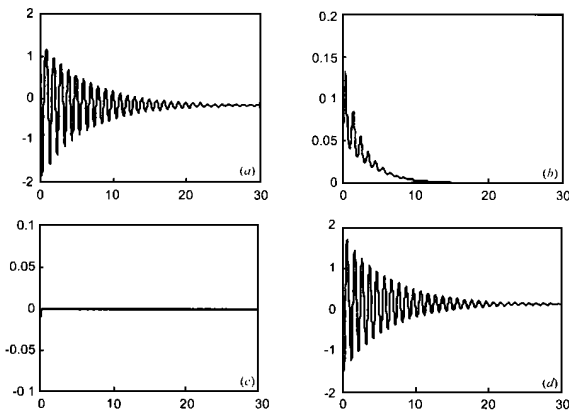


图 1 (a)、(b)、(c) 和 (d) 分别表示状态 x_1 , x_2 , λ 和 μ 的演化轨迹

5 结束语

通过改造 Lagrange 乘子项, 可对不等式约束和等式约束用同样的方法处理, 不用增加“松弛变量”. 基于这个思想构造神经网络, 可减少复杂程度和实现难度, 并且只需满足定理 2 的充分条件, 通过增加惩罚项, 能使网络局部渐近稳定. 为了验证文中方法的有效性, 我们对例子进行数值仿真, 结果表明文中的方法可行. 关于神经网络的物理实现, 我们在文献[6]中已有详细的论述, 这里不再重复.

参考文献:

- [1] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities [J]. Proc. Natl. Acad. Sci. USA., 1982, 79(4): 2554 - 2558.
- [2] J J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons [J]. Proc. Natl. Acad. Sci. USA., 1984, 81(5): 3088 - 3092.
- [3] D W Tank, J J Hopfield. Simple 'neural' optimization networks: an A/D converter, signal decision circuit and a linear programming circuit [J]. IEEE Trans. on CAS, 1986, 33(5): 533 - 541.
- [4] M P Kennedy, L O Chua. Neural networks for nonlinear programming [J]. IEEE Trans. on CAS, 1988, 35(5): 554 - 562.
- [5] S W Zhang, A G Constantinides. Lagrange programming neural networks [J]. IEEE Trans. on CAS, 1992, 39(7): 441 - 451.
- [6] 黄远灿, 孙圣和, 韩京清. 基于 Lagrange 乘子法的非线性规划神经网络 [J]. 电子学报, 1998, 26(1): 24 - 28.
- [7] D P Bertsekas. Constrained optimization and Lagrange multiplier methods [M]. Academic Press, New York, 1982.
- [8] M Avriel. Nonlinear programming analysis and methods [M]. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.

作者简介:



黄远灿 男. 1969 年 12 月生于贵州盘县. 1995 于哈尔滨工业大学自动测试与控制系获博士学位. 1996 年至 1998 年在中科研系统所控制论与运筹学博士后流动站工作. 现为北京理工大学自动控制系统副教授. 目前感兴趣的研究方向有神经网络在非线性控制中的应用、系统辨识以及多传感器数据融合等.