第22卷 第3期 自 动 化 学 报 Vol. 22, No. 3
1996年5月 ACTA AUTOMATICA SINICA May, 1996

# 一种新的非线性规划神经网络模型[1)]

杨若黎　吴沧浦
（北京理工大学自动控制系　北京　100081）

**摘　　要**

提出一种新型的求解非线性规划问题的神经网络模型. 该模型由变量神经元、Lagrange 乘子神经元和 Kuhn-Tucker 乘子神经元相互连接构成. 通过将 Kuhn-Tucker 乘子神经元限制在单边饱和工作方式,使得在处理非线性规划问题中不等式约束时不需要引入松弛变量,避免了由于引入松弛变量而造成神经元数目的增加, 有利于神经网络的硬件实现和提高神经网络的收敛速度. 可以证明,在适当的条件下,文中提出的神经网络模型的状态轨迹收敛到与非线性规划问题的最优解相对应的平衡点.

**关键词**: 神经网络,非线性规划,最优化.

## 1　引言

Hopfield 神经网络模型[1] 的提出为快速求解各种有约束优化问题开辟了一条新途径. 在 Hopfield 神经网络模型的基础上,已发展了多种不同形式的求解非线性规划问题的神经网络模型,所采用的方法主要是罚函数法和 Lagrange 乘子法. 基于罚函数法的神经网络模型[2-3] 的主要缺点是不能求得非线性规划问题的精确最优解. 虽然采用较大的罚因子可以提高解的精度, 但会给神经网络的硬件实现带来一定的困难. 尽管在基于罚函数法的神经网络模型的基础上已发展了一些变种的形式[4-5],但都不能从根本上解决基于罚函数法的神经网络模型存在的主要缺点. 基于 Lagrange 乘子法的神经网络模型不需要使用很大的罚因子, 有利于神经网络的硬件实现. 不同形式的基于 Lagrange 乘子法的神经网络模型的主要区别在于对非线性规划问题中的不等式约束采用不同的处理方法. 采用类似于罚函数的方法处理不等式约束[6], 显然不能求得具有不等式约束的非线性规划问题的精确最优解. 而采用松弛变量的方法处理不等式约束[7], 虽然可以求得具有不等式约束的非线性规划问题的精确最优解,但需要增加神经元的数目,这不利于神经网络的硬件实现,而且也会影响神经网络的收敛速度. 文[8]提出的神经网络模型在处理不等式约束时不需要引入松弛变量,但对 Kuhn-Tucker 乘子神经元的工作方式要求过于苛刻, 一旦某个神经元的状态超过了与非线性规划问题的最优解相对应的 Kuhn-Tucker 乘子的值,就无法使 Kuhn-Tucker 最优性条件满足, 而且有可能导致神经网络的状态轨

1）国家自然科学基金资助项目.
本文于1994年3月8日收到.

迹发散。

本文提出一种新型的求解非线性规划问题的神经网络模型。它与文[6—8]中的神经网络模型的不同之处在于采用限制 Kuhn-Tucker 乘子神经元工作方式的方法处理非线性规划问题中的不等式约束，避免了引入松弛变量。可以证明，在适当的条件下，本文提出的神经网络模型的状态轨迹可收敛到与非线性规划问题的最优解相对应的平衡点。

## 2 问题描述

考虑如下的一般非线性规划问题：

$$\min_{x} f(\boldsymbol{x}), \tag{1a}$$

$$\text{s.t.}\ \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{g}(\boldsymbol{x}) \leqslant 0. \tag{1b}$$

其中 $\boldsymbol{x} \in R^n; f: R^n \to R; \boldsymbol{h}: R^n \to R^p; p < n; \boldsymbol{g}: R^n \to R^m$。问题(1)的可行解集记为 $X$。

**定义 1.** 问题(1)的 Lagrange 函数 $L: R^n \times R^p \times R^m \to R$ 定义为

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \equiv f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}), \tag{2}$$

式中 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_p)$ 是 Lagrange 乘子向量; $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_m)$ 是 Kuhn-Tucker 乘子向量。

**假定 1.** $X$ 是 $R^n$ 中的非空紧子集，$f$ 是二次连续可微的凸函数，其 Hesse 阵是正定阵，$h_i (i = 1, 2, \cdots, p)$ 是线性仿射函数，所有 $h_i$ 的系数向量线性无关，$g_j (j = 1, 2, \cdots, m)$ 是二次连续可微的凸函数。

**假定 2.** 存在一点 $\hat{\boldsymbol{x}} \in R^n$ 满足 $\boldsymbol{h}(\hat{\boldsymbol{x}}) = 0$ 和 $\boldsymbol{g}(\hat{\boldsymbol{x}}) < 0$。

## 3 神经网络模型及其稳定性分析

求解问题(1)的神经网络模型由下列带有适当约束的微分方程描述：

$$C_x \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{3a}$$

$$C_\lambda \frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} = \boldsymbol{h}(\boldsymbol{x}), \tag{3b}$$

$$C_\mu \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t} = \boldsymbol{g}(\boldsymbol{x}),\ \boldsymbol{\mu} \geqslant 0. \tag{3c}$$

式中 $C_x$, $C_\lambda$ 和 $C_\mu$ 分别为适当维数的对角阵，其对角线上的元素均大于零；$\nabla$ 是梯度算子。

神经网络模型 (3) 包含三种类型的神经元：变量神经元、Lagrange 乘子神经元和 Kuhn-Tucker 乘子神经元。其动态特性分别用 (3a), (3b) 和 (3c) 描述。由于 Kuhn-Tucker 神经元的动态特性是用有边界约束的微分方程描述，所以，Kuhn-Tucker 乘子神经元是在单边饱和方式下工作。通过限制 Kuhn-Tucker 乘子神经元的工作方式，使得相应的 Kuhn-Tucker 乘子在神经网络模型(3)的状态演变过程中始终保持非负性，从而

在处理不等式约束时可以避免引入松弛变量.

在模型(3)中，由于 Kuhn-Tucker 乘子神经元是用有边界约束的微分方程描述. 所以，可采用与文[9]类似的方法给出模型(3)的解的定义. 为此，需要先引入下列符号，以表示处于饱和状态的 Kuhn-Tucker 乘子神经元的分布情况. 令

$$\Lambda \equiv \{\boldsymbol{\xi} \in R^m \mid \xi_i = 1 \text{ or } 0,\ i = 1,2,\cdots,m\}, \tag{4}$$

对于每个 $l \in \{0,1,\cdots,m\}$，又令

$$\Lambda_l \equiv \{\boldsymbol{\xi} \in \Lambda \mid \exists \sigma \in \mathrm{Sym}(m) \Longrightarrow \xi_{\sigma(k)} = 1,\ 1 \leqslant k \leqslant l;\ \xi_{\sigma(k)} = 0,\ l < k \leqslant m\}. \tag{5}$$

其中 $\mathrm{Sym}(m)$ 表示 $m$ 阶置换群. 需要指出，当 $l = 0$ 或 $l = m$ 时，$\Lambda_l$ 中只有一个向量，其分量全为零或全为 1. 对于每个 $\boldsymbol{\xi} \in \Lambda_l, 0 \leqslant l \leqslant m$，令

$$C(\boldsymbol{\xi}) \equiv \{\boldsymbol{\mu} \in R^m \mid \mu_k > 0 \text{ if } \xi_k = 1;\ \mu_k = 0 \text{ if } \xi_k = 0\}. \tag{6}$$

**定义 2.** 对于给定的 $\delta > 0$，连续可微函数 $\boldsymbol{\phi}:(0,\delta) \to R^n, \boldsymbol{\omega}:(0,\delta) \to R^p$ 和 $\boldsymbol{\varphi}:(0,\delta) \to C(\boldsymbol{\xi}),(\boldsymbol{\xi} \in \Lambda_l, 0 \leqslant l \leqslant m)$，称为神经网络模型(3)的一个局部解，如果对于所有的 $t \in (0,\delta)$，它们满足下述条件：

$$C_x \frac{\mathrm{d}\boldsymbol{\phi}(t)}{\mathrm{d}t} = -\nabla_x L(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t)), \tag{7a}$$

$$C_\lambda \frac{\mathrm{d}\boldsymbol{\omega}(t)}{\mathrm{d}t} = \boldsymbol{h}(\boldsymbol{\phi}(t)), \tag{7b}$$

$$C_{\mu,\sigma(k)} \frac{\mathrm{d}\boldsymbol{\varphi}_{\sigma(k)}(t)}{\mathrm{d}t} = g_{\sigma(k)}(\boldsymbol{\phi}(t)), \varphi_{\sigma(k)}(t) > 0,\quad 1 \leqslant k \leqslant l, \tag{7c}$$

$$g_{\sigma(k)}(\boldsymbol{\phi}(t)) \leqslant 0,\ \ \varphi_{\sigma(k)}(t) = 0,\ \ l < k \leqslant m. \tag{7d}$$

式中 $\varphi_{\sigma(k)}(t)$ 和 $g_{\sigma(k)}(\boldsymbol{\phi}(t))$ 分别表示函数 $\boldsymbol{\varphi}(t)$ 和 $\boldsymbol{g}(\boldsymbol{\phi}(t))$ 的第 $\sigma(k)$ 个分量；$C_{\mu,\sigma(k)}$ 表示对角阵 $C$ 的第 $\sigma(k)$ 个对角线元素.

**定义 3.** 对于给定的 $\boldsymbol{x}^0 \in R^n, \boldsymbol{\lambda}^0 \in R^p$ 和 $\boldsymbol{\mu}^0 \in C(\boldsymbol{\xi})$ $(\boldsymbol{\xi} \in \Lambda_l, 0 \leqslant l \leqslant m)$，以及 $\tilde{t} > 0$，连续函数 $\boldsymbol{x}:[0,\tilde{t}) \to R^n, \boldsymbol{\lambda}:[0,\tilde{t}) \to R^p$ 和 $\boldsymbol{\mu}:[0,\tilde{t}) \to R_+^m$，其中 $R_+ \equiv [0, +\infty)$，称为神经网络模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)$ 为初始条件的解，如果：1) $\boldsymbol{x}(0) = \boldsymbol{x}^0, \boldsymbol{\lambda}(0) = \boldsymbol{\lambda}^0, \boldsymbol{\mu}(0) = \boldsymbol{\mu}^0$；2)存在至多可列个互不相交的开区间 $(t_i, t_i + \delta_i)$，使得 $\bigcup(t_i, t_i + \delta_i)$ 的闭包等于 $(0,\tilde{t})$ 的闭包，并且当把函数 $\boldsymbol{x}(t), \boldsymbol{\lambda}(t)$ 和 $\boldsymbol{\mu}(t)$ 限制在每个开区间 $(t_i, t_i + \delta_i)$ 上时，它们是神经网络模型(3)的一个局部解.

对于神经网络模型(3)，做如下假定.

**假定 3.** 存在某个正数 $\delta_m$，使 (7a)—(7d) 式成立的每个开区间 $(t_i,\ t_i + \delta_i)$ 都满足 $\delta_i \geqslant \delta_m$.

根据常微分方程理论[10]，可得如下定理.

**定理 1.** 在假定 1 和 3 下，存在 $R^n \times R^p \times R_+^m$ 中的某个区域 $D$，使得对于任意给定的初始条件 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0) \in D$，神经网络模型(3)有唯一解.

**定义 4.** 向量 $\tilde{\boldsymbol{x}} \in R^n, \tilde{\boldsymbol{\lambda}} \in R^p$ 和 $\tilde{\boldsymbol{\mu}} \in C(\boldsymbol{\xi})(\boldsymbol{\xi} \in \Lambda_l,\ 0 \leqslant l \leqslant m)$ 称为神经网络模型(3)的平衡点或平衡解，如果它们满足下列条件：

$$\nabla_x L(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}}) = 0, \tag{8a}$$
$$\boldsymbol{h}(\tilde{\boldsymbol{x}}) = 0, \tag{8b}$$
$$g_{\sigma(k)}(\tilde{\boldsymbol{x}}) = 0, \tilde{\mu}_{\sigma(k)} > 0, 1 \leqslant k \leqslant l, \tag{8c}$$
$$g_{\sigma(k)}(\tilde{\boldsymbol{x}}) \leqslant 0, \tilde{\mu}_{\sigma(k)} = 0, l < k \leqslant m. \tag{8d}$$

根据问题(1)的 Kuhn-Tucker 最优性条件和模型(3)的平衡点的定义，可得如下定理.

**定理 2.** 在假定 1 和 2 下，1) 若 $\boldsymbol{x}^*$ 是问题(1)的最优解，则存在向量 $\boldsymbol{\lambda}^* \in R^p$ 和 $\boldsymbol{\mu}^* \in R^m$，使得 $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 是模型(3)的平衡点；2)若 $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 是模型(3)的平衡点，则 $\boldsymbol{x}^* \in X$ 是问题(1)的最优解.

**定义 5.** 设 $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}})$ 是模型(3)的平衡点，模型(3)的能量函数 $E: R^n \times R^p \times R_+^m \to R$ 定义为

$$E(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \equiv E_1(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) + \frac{1}{2} \nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})^T C_x^{-1} \nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$
$$+ \frac{1}{2} \boldsymbol{h}(\boldsymbol{x})^T C_\lambda^{-1} \boldsymbol{h}(\boldsymbol{x}) + \frac{1}{2} \boldsymbol{q}(\boldsymbol{x}, \boldsymbol{\mu})^T C_\mu^{-1} \boldsymbol{q}(\boldsymbol{x}, \boldsymbol{\mu}). \tag{9}$$

式中

$$E_1(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \equiv \frac{1}{2} (\boldsymbol{x} - \tilde{\boldsymbol{x}})^T C_x (\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \frac{1}{2} (\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})^T C_\lambda (\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})$$
$$+ \frac{1}{2} (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}})^T C_\mu (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}); \tag{10}$$

向量值函数 $\boldsymbol{q}(\boldsymbol{x}, \boldsymbol{\mu}) \in R^m$ 的各分量 $q_k: R^n \times R \to R (k = 1, 2, \cdots, m)$ 定义为

$$q_k(\boldsymbol{x}, \mu_k) \equiv \begin{cases} 0, & \text{if} \quad \mu_k = 0, g_k(\boldsymbol{x}) \leqslant 0, \\ g_k(\boldsymbol{x}), & \text{otherwise}. \end{cases} \tag{11}$$

**定理 3.** 在假定 1—3 下，设 $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}})$ 是模型(3)的平衡点，则对于任意给定的实数 $\eta > 0$，存在实数 $\varepsilon > 0$，使得当 $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0, \boldsymbol{\mu}^0) \in N_\varepsilon(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}}) \equiv \{(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in R^n \times R^p \times R_+^m \mid \|(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}})\| < \varepsilon\}$ 时（其中$\|\cdot\|$表示 Euclidean 范数)，模型(3)以 $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0, \boldsymbol{\mu}^0)$ 为初始条件的解满足对于所有的 $t \geqslant 0$，有 $(\boldsymbol{x}(t), \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) \in N_\eta(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}})$.

证明. 设连续可微函数 $\boldsymbol{\phi}: (t_i, t_i + \delta_i) \to R^n$，$\boldsymbol{\omega}: (t_i, t_i + \delta_i) \to R^p$ 和 $\boldsymbol{\varphi}: (t_i, t_i + \delta_i) \to C(\boldsymbol{\xi}) (\boldsymbol{\xi} \in \Lambda_l, 0 \leqslant l \leqslant m)$ 是模型(3)的一个非平衡局部解，在该局部解的轨迹上函数 $E_1$ 变为

$$E_1(\boldsymbol{\phi}(t), \boldsymbol{\omega}(t), \boldsymbol{\varphi}(t)) = \frac{1}{2} (\boldsymbol{\phi}(t) - \tilde{\boldsymbol{x}})^T C_x (\boldsymbol{\phi}(t) - \tilde{\boldsymbol{x}})$$
$$+ \frac{1}{2} (\boldsymbol{\omega}(t) - \tilde{\boldsymbol{\lambda}})^T C_\lambda (\boldsymbol{\omega}(t) - \tilde{\boldsymbol{\lambda}})$$
$$+ \frac{1}{2} \sum_{k=1}^{l} c_{\mu, \sigma(k)} (\varphi_{\sigma(k)}(t) - \tilde{\mu}_{\sigma(k)})^2$$
$$+ \frac{1}{2} \sum_{k=l+1}^{m} c_{\mu, \sigma(k)} \tilde{\mu}_{\sigma(k)}^2. \tag{12}$$

将 $E_1(\boldsymbol{\phi}(t), \boldsymbol{\omega}(t), \boldsymbol{\varphi}(t))$ 对 $t$ 求导，可得

$$\frac{\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))}{\mathrm{d}t}=-(\boldsymbol{\phi}(t)-\tilde{\boldsymbol{x}})^T\nabla_x L(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))$$
$$+(\boldsymbol{\omega}(t)-\tilde{\boldsymbol{\lambda}})^T\boldsymbol{h}(\boldsymbol{\phi}(t))+(\boldsymbol{\varphi}(t)-\tilde{\boldsymbol{\mu}})^T\boldsymbol{g}(\boldsymbol{\phi}(t))$$
$$+\sum_{k=l+1}^{m}\tilde{\mu}_{\sigma(k)}g_{\sigma(k)}(\boldsymbol{\phi}(t)). \tag{13}$$

由假定 1 知 $L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$ 是关于变量 $\boldsymbol{x}$ 的严格凸函数,利用严格凸函数的性质以及 $L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$ 在 $(\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))$ 处关于 $(\boldsymbol{\lambda},\boldsymbol{\mu})$ 的泰勒级数展开式,可将上式变为

$$\frac{\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))}{\mathrm{d}t}<L(\tilde{\boldsymbol{x}},\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))-L(\boldsymbol{\phi}(t),\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$$
$$+\sum_{k=l+1}^{m}\tilde{\boldsymbol{\mu}}_{\sigma(k)}g_{\sigma(k)}(\boldsymbol{\phi}(t)). \tag{14}$$

因为 $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 满足 Kuhn-Tucker 最优性条件,并且由假定 1 可得 $\nabla_x^2 L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})>0$,所以,根据泰勒级数展开式可得不等式:

$$L(\tilde{\boldsymbol{x}},\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))\leqslant L(\boldsymbol{\phi}(t),\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}}). \tag{15}$$

考虑到 $\tilde{\mu}_{\sigma(k)}g_{\sigma(k)}(\boldsymbol{\phi}(t))\leqslant 0$, $l<k\leqslant m$, 可得: 当 $\boldsymbol{\phi}(t)\neq\tilde{\boldsymbol{x}}$ 时, $\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))/\mathrm{d}t<0$; 当 $\boldsymbol{\phi}(t)=\tilde{\boldsymbol{x}}$ 时,由(13)式可得 $\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))/\mathrm{d}t=0$. 因此,对任意的 $t\in(t_i,t_i+\delta_i)$, 有 $\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))/\mathrm{d}t\leqslant 0$.

考虑到 $E_1$ 在模型(3)的任意两个相邻局部解的连接点处的连续性,可得对所有 $t\geqslant 0$, 模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)\in N_\varepsilon(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 为初始条件的解满足 $E_1(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))\leqslant E_1(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)$. 由此可得,对于任意给定的实数 $\eta>0$, 只要取足够小的 $\varepsilon>0$, 就可得到对于所有的 $t\geqslant 0$, 模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)\in N_\varepsilon(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 为初始条件的解满足 $(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))\in N_\eta(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$. 证毕.

定理 3 表明,在一定条件下模型(3)的解是有界的. 根据常微分方程理论[10], 可得如下定理.

**定理 4.** 在定理 3 的假设下,存在实数 $\varepsilon>0$, 使得当 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)\in N_\varepsilon(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 时,模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)$ 为初始条件的解可以唯一地延拓到无限区间 $[0,+\infty)$ 上.

**定理 5.** 在定理 3 的假设下,存在实数 $\varepsilon>0$, 使得当 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)\in N_\varepsilon(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 时,沿着模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)$ 为初始条件的非平衡解的轨迹,能量函数 $E$ 单调下降.

证明. 令 $\hat{t}>0$ 表示在模型(3)的解的轨迹上某两个相邻局部解相互连接的时刻,由定义 2 可知存在 $\rho>0$, 使得 $\boldsymbol{\mu}(t)\in C(\boldsymbol{\xi}),t\in(\hat{t}-\rho,\hat{t}),\boldsymbol{\xi}\in\Lambda_l,0\leqslant l\leqslant m$. 当 $t=\hat{t}$ 时,处于饱和状态的 Kuhn-Tucker 乘子神经元的分布情况将可能发生变化. 如果对于 $1\leqslant k\leqslant l$ 下述条件成立:

$$\begin{cases}\mu_{\sigma(k)}(t)>0,\ t\in(\hat{t}-\rho,\hat{t}),\\ \mu_{\sigma(k)}(\hat{t})=0,\\ g_{\sigma(k)}(\boldsymbol{x}(\hat{t}))<0,\end{cases} \tag{16}$$

则存在某个 $1\leqslant k\leqslant l$, 使得 $q_{\sigma(k)}(\boldsymbol{x}(t),\mu_{\sigma(k)}(t))$ 在 $t=\hat{t}$ 时从 $g_{\sigma(k)}(\boldsymbol{x}(\hat{t}))$ 跳变到零. 由于

$$E(\boldsymbol{x}(\hat{t}^+),\boldsymbol{\lambda}(\hat{t}^+),\boldsymbol{\mu}(\hat{t}^+))<E(\boldsymbol{x}(\hat{t}^-),\boldsymbol{\lambda}(\hat{t}^-),\boldsymbol{\mu}(\hat{t}^-)), \tag{17}$$

所以，当(16)式的条件成立时，能量函数 $E$ 在 $t=\hat{t}$ 处将出现一个负跳变. 如果(16)式的条件不成立，则由定义 5 可知，能量函数 $E$ 在 $t=\hat{t}$ 处是连续的. 因此，要证明该定理，只需证明沿着模型(3)的任意一个非平衡局部解的轨迹，能量函数 $E$ 是单调下降的. 设连续可微函数 $\boldsymbol{\phi}:(t_i,t_i+\delta_i)\to R^n,\boldsymbol{\omega}:(t_i,t_i+\delta_i)\to R^p$ 和 $\boldsymbol{\varphi}:(t_i,t_i+\delta_i)\to C(\boldsymbol{\xi}),\boldsymbol{\xi}\in\Lambda_l$, $0\leqslant l\leqslant m$ 是模型(3)的一个非平衡局部解，将能量函数$E(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))$对 $t$ 求导，可得

$$\begin{aligned}\frac{\mathrm{d}E(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))}{\mathrm{d}t}=&\frac{\mathrm{d}E_1(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))}{\mathrm{d}t}\\&-\frac{\mathrm{d}\boldsymbol{\phi}(t)^T}{\mathrm{d}t}\nabla_x^2L(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))\frac{\mathrm{d}\boldsymbol{\phi}(t)}{\mathrm{d}t}.\end{aligned}\tag{18}$$

根据定理 3 的证明过程可得，对于任意的 $t\in(t_i,t_i+\delta_i)$，有 $\mathrm{d}E(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))/\mathrm{d}t\leqslant 0$，并且 $\mathrm{d}E(\boldsymbol{\phi}(t),\boldsymbol{\omega}(t),\boldsymbol{\varphi}(t))/\mathrm{d}t=0$ 当且仅当 $\boldsymbol{\phi}(t)=\tilde{\boldsymbol{x}},\boldsymbol{\omega}(t)=\tilde{\boldsymbol{\lambda}}$ 和 $\boldsymbol{\varphi}(t)=\tilde{\boldsymbol{\mu}}$. 这表明能量函数 $E$ 在区间 $(t_i,t_i+\delta_i)$ 上是单调下降的. 证毕.

**定理 6.** 在定理 3 的假设下，模型(3)的平衡点 $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 是渐近稳定的.

证明. 设 $(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))$ 是模型(3)以 $(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)\neq(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 为初始条件的解. 根据定理 3 可得，平衡点 $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})$ 是稳定的. 下面证明当 $t\to+\infty$ 时，$\|(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))-(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})\|\to 0$. 因为沿着模型(3)的解的轨迹，能量函数 $E$ 是关于时间 $t$ 的单调下降函数，所以，当 $t\to+\infty$ 时，有 $E\to a$，并且在任何时刻都有 $E>a$. 下面证明 $a=0$. 假定 $a\neq 0$，由 $E\geqslant 0$ 可得 $a>0$. 由定理 3 知，存在实数 $0<\beta<\eta$，使得对所有的 $t\geqslant 0$，有 $(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))\in S=\{(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})\in R^n\times R^p\times R_+^m\mid\beta\leqslant\|(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})-(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}})\|\leqslant\eta\}$. 考虑到在模型(3)的解的轨迹上任意两个相邻局部解的连接时刻 $\hat{t}$ 处，左导数 $\mathrm{d}E(\boldsymbol{x}(\hat{t}^-),\boldsymbol{\lambda}(\hat{t}^-),\boldsymbol{\mu}(\hat{t}^-))/\mathrm{d}t$ 和右导数 $\mathrm{d}E(\boldsymbol{x}(\hat{t}^+),\boldsymbol{\lambda}(\hat{t}^+),\boldsymbol{\mu}(\hat{t}^+))/\mathrm{d}t$ 都是存在的. 若令

$$\begin{aligned}\frac{\mathrm{d}E(\boldsymbol{x}(\hat{t}),\boldsymbol{\lambda}(\hat{t}),\boldsymbol{\mu}(\hat{t}))}{\mathrm{d}t}=\max\Bigg\{&\frac{\mathrm{d}E(\boldsymbol{x}(\hat{t}^-),\boldsymbol{\lambda}(\hat{t}^-),\boldsymbol{\mu}(\hat{t}^-))}{\mathrm{d}t},\\&\frac{\mathrm{d}E(\boldsymbol{x}(\hat{t}^+),\boldsymbol{\lambda}(\hat{t}^+),\boldsymbol{\mu}(\hat{t}^+))}{\mathrm{d}t}\Bigg\},\end{aligned}\tag{19}$$

则可以使函数 $\mathrm{d}E(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))/\mathrm{d}t$ 成为区间 $[0,+\infty)$ 上的上半连续函数. 由于在集 $S$ 上 $\mathrm{d}E(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))/\mathrm{d}t<0$，并且可以在集 $S$ 上达到上确界，所以，对于所有的 $t\geqslant 0$，有 $\mathrm{d}E(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))/\mathrm{d}t\leqslant -b$，其中 $b>0$. 由此可得

$$\begin{aligned}E(\boldsymbol{x}(t),\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))&=E(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)+\int_0^t\frac{\mathrm{d}E(\boldsymbol{x}(\tau),\boldsymbol{\lambda}(\tau),\boldsymbol{\mu}(\tau))}{\mathrm{d}\tau}\mathrm{d}\tau\\&\leqslant E(\boldsymbol{x}^0,\boldsymbol{\lambda}^0,\boldsymbol{\mu}^0)-bt.\end{aligned}\tag{20}$$

显然这是不可能的，因为对于充分大的 $t$ 值，上述不等式的右端是负的，这与 $E\geqslant 0$ 矛盾. 因此，当 $t\to+\infty$ 时，有 $E\to 0$. 由此可得，当 $t\to+\infty$ 时，有 $\boldsymbol{x}(t)\to\tilde{\boldsymbol{x}}$, $\boldsymbol{\lambda}(t)\to\tilde{\boldsymbol{\lambda}}$, $\boldsymbol{\mu}(t)\to\tilde{\boldsymbol{\mu}}$. 证毕.

定理 2 和定理 6 表明，在适当的条件下，本文提出的神经网络模型的非平衡解的轨迹收敛到与原问题的最优解相对应的平衡点.

# 4 结束语

本文所提出的求解非线性规划问题的神经网络模型克服了基于罚函数法的优化神经网络模型的缺点，能够求得原问题的精确最优解．通过限制 Kuhn-Tucker 乘子神经元在单边饱和方式下工作，可避免引入松弛变量，将问题中的不等式约束化成等式约束的形式，从而在处理不等式约束时不必增加神经元的数目，有利于神经网络的硬件实现和加快神经网络的收敛速度．分析表明，在适当的条件下，本文提出的神经网络模型的非平衡解的轨迹收敛到与原问题的最优解相对应的平衡点．

这里只考虑了求解凸规划问题的情况，若使用增广 Lagrange 函数构造相应的神经网络模型，则可将本文的结果推广到求解非凸规划问题．

## 参　考　文　献


[1] Hopfield J J. Neurons with graded response have collective computational properties like those of two-state neurons. In Proc. Natl. Acad. Sci. USA, 1984, **81**: 3088—3092.

[2] Kennedy M P, Chua L O. Neural networks for nonlinear programming. *IEEE Trans. on Circuits and Systems*, 1988, **35**: 554—562.

[3] Maa Chia Yiu, Shanblatt M A. Linear and quadratic programming neural network analysis. *IEEE Trans. on Neural Networks*, 1992, **3**: 580—594.

[4] Rodriguez Vazguez A, Dominguez Castro R, Rueda A, Huertas J L, Sanchez Sinencio E. Nonlinear switched-capacitor "Neural" networks for optimization problems. *IEEE Trans. on Circuits and Systems*, 1990, **37**: 384—398.

[5] Wang Jun. On the asymptotic properties of recurrent neural networks for optimization. *International Journal of Pattern Recognition and Artificial Intelligence*, 1991, **5**: 581—601.

[6] Cichocki A, Unbehauen R. Switched-capacitor neural networks for differential optimization. *International Journal of Circuit Theory and Applications*. 1991, **19**: 161—187.

[7] Zhang Shengwei, Constantinides A G. Lagrange programming neural networks. *IEEE Trans. on Circuits and Systems*, 1992, **39**: 441—452.

[8] Maa Chia Yiu, Shanblatt M A. A two-phase optimization neural network. *IEEE Trans. on Neural Networks*, 1992, **3**: 1003—1009.

[9] Li Jian Hua, Michel A N, Porod W. Analysis and synthesis of a class of neural networks: Linear systems operating on a closed hypercube. *IEEE Trans. on Circuits and Systems*, 1989, **36**: 1405—1422.

[10] Miller P K, Michel A N. Ordinary differential equations. New York: Academic-Press, 1982.

# A NOVEL NEURAL NETWORK MODEL FOR NONLINEAR PROGRAMMING

YANG RUOLI WU CANGPU
(*Dept. of Automatic Control, Beijing Inst. of Tech., Beijing* 100081)

ABSTRACT

A novel neural network model for solving nonlinear programming problems is proposed in this paper. It is composed of variable neurons, Lagrange multiplier neurons and Kuhn-Tucker multiplier neurons which are interconnected. By making the Kuhn-Tucker multiplier neurons operate in an one-sided saturated mode, the introduction of the slack variables is no more necessary in dealing with the inequality constraints of nonlinear programming problems. This method can avoid the increase in the number of neurons caused by the slack variables. This is advantageous to the hardware implementation and the convergence rate improvement. It can be shown that under suitable conditions the trajectory of the proposed neural network model converges to the equilibrium point corresponding to the optimal solution of the nonlinear programming problem.

**Key words:** Neural network, nonlinear programming, optimization.

**杨若黎** 1959年生. 1982年和1985年分别在北京理工大学获学士和硕士学位. 毕业后留校任教. 1994年任北京理工大学副教授，1995年在北京理工大学获博士学位. 现在中国科学院系统所做博士后研究工作. 目前主要研究方向为全局优化与递阶优化、系统优化、控制与决策中的神经网络技术等.

**吴沧浦** 简介及照片见本刊第20卷第5期.