第22卷 第6期　　自 动 化 学 报　　Vol. 22, No. 6
1996年 11月　　ACTA AUTOMATICA SINICA　　Nov., 1996

一类二层决策问题的性质分析和基于 Frank-Walfe 方法的神经网络算法[1)]

盛昭瀚
(东南大学经济管理学院　南京　210018)
吕庆喆　　徐南荣
(国家统计局国际中心　北京　100826)　(东南大学经济管理学院　南京　210018)

摘　　要

提出一类二层决策问题的模型. 在某些凸性假设下,研究它的一些性质,提出了一种基于 Frank-Walfe 和人工神经网络方法的算法. 并以算例说明该算法可行性.

关键词: 二层决策, Frank-Walfe 算法,人工神经网络.

1　引言

地位不同的许多人或多个机构做为一个整体做出的决策称为多层决策问题. 其中个人或机构之间有着一定的隶属关系,并均有各自的决策目标.

这里描述的一类二层决策问题的数学模型如下:

$$
\begin{aligned}
&\min_{\boldsymbol{x}} p(\boldsymbol{x}) \triangleq F(\boldsymbol{x}, \boldsymbol{q}(\boldsymbol{x})),\\
&\text{s.t. } \boldsymbol{x} \in X = \{\boldsymbol{x} \in R^n \mid \boldsymbol{h}(\boldsymbol{x}) \leqslant \boldsymbol{0}\},\\
&q_i(\boldsymbol{x}) \triangleq \min_{\boldsymbol{y}_i} f_i(\boldsymbol{x}, \boldsymbol{y}_i),\\
&\text{s.t. } \boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{y}_i) \leqslant \boldsymbol{0},\ i = 1, 2, \cdots, m.
\end{aligned}
\tag{P$_1$}
$$

这里 $\boldsymbol{q}(\boldsymbol{x}) = (q_1(\boldsymbol{x}), \cdots, q_m(\boldsymbol{x}))$; 变量 $\boldsymbol{x}, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m$ 分别为 $n, l_1, \cdots, l_m$ 维向量; $F: R^n \times R^m \to R^1$; $\boldsymbol{h} = (h_1, \cdots, h_n): R^n \to R^h$; $f_i: R^n \times R^{l_i} \to R^1$; $\boldsymbol{g}_i = (g_{i1}, \cdots, g_{ik_i}): R^n \times R^{l_i} \to R^{k_i}$.

这类模型的决策机制是: 上层首先宣布它的决策为 $\boldsymbol{x}$, 这一决策将影响下级各决策问题的约束集和目标函数值; 然后下层决策者在这种限制下选取使自己目标函数 f_i 最优的决策 $\boldsymbol{y}_i (i = 1, \cdots, m)$; 下层各决策者的目标最优值也影响上层决策问题的约束集和

1) 国家自然科学基金资助课题.
本文于1994年9月2日收到

目标函数值,上层决策者再调整它的决策变量 $\boldsymbol{x}$ 的取值，通过如此替代,直至其目标函数达最优为止。

这里假设下层各决策者之间不进行合作,且

(1) 函数 $f_i(\boldsymbol{x},\boldsymbol{y}_i)$ 关于 $(\boldsymbol{x},\boldsymbol{y}_i)$ 是凸的和可微的，$i=1,\cdots,m$;

(2) 函数 $g_{ij}(\boldsymbol{x},\boldsymbol{y}_i)$ 关于 $(\boldsymbol{x},\boldsymbol{y}_i)$ 是凸的和可微的，$j=1,\cdots,k_i$，$i=1,\cdots,m$;

(3) $\boldsymbol{h}(\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 是凸的和可微的，且对 $\boldsymbol{x}\in X$，集合 $\{(\boldsymbol{x},\boldsymbol{y}_i)|\boldsymbol{g}_i(\boldsymbol{x},\boldsymbol{y}_i)\leqslant \boldsymbol{0}\}$ 的内点非空，$i=1,\cdots,m$;

(4) 函数 $F(\boldsymbol{x},\boldsymbol{q})$ 关于 $(\boldsymbol{x},\boldsymbol{q})$ 是凸的和可微的,且关于 $\boldsymbol{q}$ 单调非减;

(5) 对任意 $\boldsymbol{x}\in X$ 和满足条件 $q_i(\boldsymbol{x})=f_i(\boldsymbol{x},\boldsymbol{y}_i)$，$\boldsymbol{g}_i(\boldsymbol{x},\boldsymbol{y}_i)\leqslant\boldsymbol{0}$ 的 $(\boldsymbol{x},\boldsymbol{y}_i)$ 都有 $\{\nabla \boldsymbol{y}_i g_{ij}(\boldsymbol{x},\boldsymbol{y}_i),\ j\in I_i(\boldsymbol{x},\boldsymbol{y}_i)\}$ 是线性独立的，这里 $I_i(\boldsymbol{x},\boldsymbol{y}_i)=\{j|g_{ij}(\boldsymbol{x},\boldsymbol{y}_i)=0,\ j=1,\cdots,k_i\}$，$i=1,\cdots,m$.

在实际问题中,例如 $F(\boldsymbol{x},\boldsymbol{q})$ 和 $f_i(\boldsymbol{x},\boldsymbol{y}_i)$ 为凸二次函数，$\boldsymbol{h}(\boldsymbol{x})$ 和 $\boldsymbol{g}_i(\boldsymbol{x},\boldsymbol{y}_i)$ 为线性函数 $(i=1,\cdots,m)$ 的问题 (P_1) 均满足假设(1)—(4),而解优化问题常常有假设(5).

本文针对上面提出的模型 (P_1)，首先分析它的若干性质，然后结合 Frank-Wolfe 方法和人工神经网络,提出一种新的算法.

2 二层决策问题 (P_1) 的性质分析

定义 2.1. 设 $X\subseteq R^n$，点集映射 $G:X\to 2^{R^n}$ 称为凸的，如果对任何 $\boldsymbol{x}_1$，$\boldsymbol{x}_2\in X$，$0\leqslant\lambda\leqslant 1$，则有 $\lambda G(\boldsymbol{x}_1)+(1-\lambda)G(\boldsymbol{x}_2)\subseteq G(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)$.

引理 2.1. 设 $X\subseteq R^n$ 为凸集，$f(\boldsymbol{x},\boldsymbol{y})$ 为 $X\times R^m$ 上的凸函数，$G:X\to 2^{R^m}$ 为 X 上的凸点集映射,对任何 $\boldsymbol{x}\in X$，$f(\boldsymbol{x},\boldsymbol{y})$ 关于 $\boldsymbol{y}$ 在 $G(\boldsymbol{x})$ 上的极小点存在,则

$$\boldsymbol{q}(\boldsymbol{x})\triangleq\min_{\boldsymbol{y}\in G(\boldsymbol{x})} f(\boldsymbol{x},\boldsymbol{y}) \text{ 为}X\text{上的凸函数.}$$

证明. $\forall\boldsymbol{x}_1,\boldsymbol{x}_2\in X$，$0\leqslant\lambda\leqslant 1$，由于$G$为$X$上的凸点集映射,则有

$$\lambda G(\boldsymbol{x}_1)+(1-\lambda)G(\boldsymbol{x}_2)\subseteq G(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2),$$

$$\begin{aligned}
\boldsymbol{q}(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2) &= \min_{\boldsymbol{y}\in G(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2)} f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2,\boldsymbol{y})\\
&\leqslant \min_{\boldsymbol{y}\in\lambda G(\boldsymbol{x}_1)+(1-\lambda)G(\boldsymbol{x}_2)} f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2,\boldsymbol{y})\\
&= \min_{\boldsymbol{y}_1\in G(\boldsymbol{x}_1)\boldsymbol{y}_2\in G(\boldsymbol{x}_2)} f(\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_2,\lambda\boldsymbol{y}_1+(1-\lambda)\boldsymbol{y}_2)\\
&\leqslant \min_{\substack{\boldsymbol{y}_1\in G(\boldsymbol{x}_1)\\ \boldsymbol{y}_2\in G(\boldsymbol{x}_2)}} [\lambda f(\boldsymbol{x}_1,\boldsymbol{y}_1)+(1-\lambda)f(\boldsymbol{x}_2,\boldsymbol{y}_2)] = \lambda\min_{\boldsymbol{y}_1\in G(\boldsymbol{x}_1)} f(\boldsymbol{x}_1,\boldsymbol{y}_1)\\
&\quad +(1-\lambda)\min_{\boldsymbol{y}_2\in G(\boldsymbol{x}_2)} f(\boldsymbol{x}_2,\boldsymbol{y}_2)=\lambda\boldsymbol{q}(\boldsymbol{x}_1)+(1-\lambda)\boldsymbol{q}(\boldsymbol{x}_2).
\end{aligned}$$

所以 $\boldsymbol{q}(\boldsymbol{x})$ 是X上的凸函数.

定理 2.1. 在假设(1)—(3)的条件下,函数 $q_i(\boldsymbol{x})$ 是凸的 $(i=1,\cdots,m)$.

定理 2.2. 在假设(1)—(4)的条件下,函数 $p(\boldsymbol{x})\triangleq F(\boldsymbol{x},\boldsymbol{q}(\boldsymbol{x}))$ 为X上的凸函数.

定理 2.3. 在满足定理 2.1 的条件下，若 $f_i(\boldsymbol{x},\boldsymbol{y}_i)$ 还满足它是关于 $(\boldsymbol{x},\boldsymbol{y}_i)$ 的严格

凸函数,则 $q_i(\boldsymbol{x})$ 为 $\boldsymbol{x}$ 的严格凸函数 $(i=1,\cdots,m)$.

定理 2.4. 在满足定理 2.2 和 2.3 的条件下,若 $F(\boldsymbol{x},\boldsymbol{q}(\boldsymbol{x}))$ 还满足它是关于 $(\boldsymbol{x},\boldsymbol{q})$ 为严格凸函数或者它是关于 $\boldsymbol{q}$ 严格增的,则 $p(\boldsymbol{x})\triangleq F(\boldsymbol{x},\boldsymbol{q}(\boldsymbol{x}))$ 为严格凸函数.(以上定理证明从略.)

下面进一步讨论隐式函数 $\boldsymbol{q}(\boldsymbol{x})$ 与 $p(\boldsymbol{x})$ 的连续性和可微性. 为此，引入如下定义.

定义 2.2. 设 $f(\boldsymbol{x})$ 是 $X\subseteq R^n$ 上的有限凸函数，$\boldsymbol{x}_0\in X$. 集合

$$\partial f(\boldsymbol{x}_0)=\{\boldsymbol{v}\in X\,|\,f(\boldsymbol{x})-f(\boldsymbol{x}_0)\geqslant(\boldsymbol{v},\boldsymbol{x}-\boldsymbol{x}_0),\forall\boldsymbol{x}\in X\}$$

称为函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}_0$ 处的次微分. 元素 $\boldsymbol{v}_0\in\partial f(\boldsymbol{x}_0)$ 称为函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}_0$ 处的次梯度.

引理 2.2. 假设函数 $h(\boldsymbol{u},\boldsymbol{v}):R^s\times R^t\to\bar{R}$ 是关于 $(\boldsymbol{u},\boldsymbol{v})$ 的凸函数,令

$$r(\boldsymbol{u})=\inf_{\boldsymbol{v}}h(\boldsymbol{u},\boldsymbol{v}).$$

若 $r(\bar{\boldsymbol{u}})=h(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$,则 $\partial r(\bar{\boldsymbol{u}})=\{\boldsymbol{\xi}\in R^s\,|\,(\boldsymbol{\xi},\boldsymbol{0})\in\partial h(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\}$，这里 $\bar{R}=\{+\infty\}\cup R$. (证明参见文 [5] Theorem 24.)

引理 2.3. 设 $f:R^n\to\bar{R}$ 和 $g:R^n\to\bar{R}$ 是下半连续真凸函数. 若

$$(\mathrm{dom}\,f)\cap\mathrm{int}(\mathrm{dom}\,g)\neq\varnothing,$$

则 $\partial(f+g)=\partial f+\partial g$. 这里 $\mathrm{dom}\,f=\{\boldsymbol{x}\in R^n\,|\,f(\boldsymbol{x})<+\infty\}$，$\mathrm{int}\,s$ 表示集合 s 的内点集合.(证明参见文 [6] Theorem 5c.)

引理 2.4. 凸函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}_0\in X\subseteq R^n$ 处可微,当且仅当它的次微分 $\partial f(\boldsymbol{x}_0)$ 由唯一的一点组成,即 $\partial f(\boldsymbol{x}_0)=\{\nabla f(\boldsymbol{x}_0)\}$. (证明参见文 [7] Theorem 15.4.)

对于问题 (P_1) 的下层优化问题

$$q_i(\boldsymbol{x})=\min_{\boldsymbol{y}_i}f_i(\boldsymbol{x},\boldsymbol{y}_i),$$

$$\text{s.t.}\ \ \boldsymbol{g}_i(\boldsymbol{x},\boldsymbol{y}_i)\leqslant\boldsymbol{0}\quad(i=1,\cdots,m),$$

可以写成另一种形式

$$q_i(\boldsymbol{x})=\inf_{\boldsymbol{y}_i}\{f_i(\boldsymbol{x},\boldsymbol{y}_i)+\delta_{\boldsymbol{c}_{i1}}(\boldsymbol{x},\boldsymbol{y}_i)+\cdots+\delta_{\boldsymbol{c}_{ik_i}}(\boldsymbol{x},\boldsymbol{y}_i)\},$$

这里 $\boldsymbol{c}_{ij}=\{(\boldsymbol{x},\boldsymbol{y}_i)\,|\,g_{ij}(\boldsymbol{x},\boldsymbol{y}_i)\leqslant0\}$，$j=1,\cdots,k_i$;

$$\delta_{\boldsymbol{c}_{ij}}(\boldsymbol{x},\boldsymbol{y}_i)=\begin{cases}0, & \text{若 }(\boldsymbol{x},\boldsymbol{y}_i)\in\boldsymbol{c}_{ij},\\ +\infty, & \text{其它.}\end{cases}$$

若 $q_i(\boldsymbol{x}^*)=f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)$，$\boldsymbol{g}_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\leqslant\boldsymbol{0}$，且假设 (1)，(2) 成立，从引理 2.2 知，$\boldsymbol{\xi}_i\in\partial q_i(\boldsymbol{x}^*)\Longleftrightarrow(\boldsymbol{\xi}_i,0)\in\partial(f_i+\delta_{\boldsymbol{c}_{i1}}+\cdots+\delta_{\boldsymbol{c}_{ik_i}})(\boldsymbol{x}^*,\boldsymbol{y}_i^*)$.

因为 $\mathrm{int}(\mathrm{dom}\,f_i)=R^n\times R^{l_i}$ 和 $\mathrm{dom}\,\delta_{\boldsymbol{c}_{ij}}=\boldsymbol{c}_{ij}\neq\varnothing$，$j=1,\cdots,k_i$. 由引理 2.3 知，$\boldsymbol{\xi}_i\in\partial q_i(\boldsymbol{x}^*)\Longleftrightarrow(\boldsymbol{\xi}_i,\boldsymbol{0})\in\partial f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)+\partial\delta_{\boldsymbol{c}_{i1}}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)+\cdots+\partial\delta_{\boldsymbol{c}_{ik_i}}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)$.

由于 f_i 可微，则 $\partial f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=\{(\nabla_x f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*),\ \nabla_{\boldsymbol{y}_i}f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*))\}$. 又由假设 (4) 知 $\mathrm{int}\,\boldsymbol{c}_{i1}\cap\cdots\cap\mathrm{int}\,\boldsymbol{c}_{ik_i}\neq\varnothing$，则由文[8] Corollary 23.7.1 得

$$\partial\delta_{c_{ij}}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=\begin{cases}\bigcup\{\lambda_{ij}\partial g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\mid\lambda_{ij}\geqslant 0\}, & \text{若 } g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=0,\\ \{\boldsymbol{0}\}, & \text{若 } g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)<0,\\ \varnothing, & \text{若 } g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)>0,\end{cases}$$

即 $\partial\delta_{c_{i1}}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)+\cdots+\partial\delta_{c_{ik_i}}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)$

$$=\left\{\sum_{j=1}^{k_i}\lambda_{ij}^*(\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*),\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*))\mid\lambda_{ij}^*\geqslant 0,\ \lambda_{ij}^*=0,\ \text{若 } j\notin I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\right\}.$$

综上所述,得出如下定理.

定理 2.5. 若假设(1)—(3)成立,且有 $q_i(\boldsymbol{x}^*)=f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)$ 和 $\boldsymbol{g}_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\leqslant\boldsymbol{0}$,那么 $\partial q_i(\boldsymbol{x}^*)=\left\{\nabla_x f_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)+\sum_{j=1}^{k_i}\lambda_{ij}^*\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\mid\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)+\sum_{j=1}^{k_i}\lambda_{ij}^*\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=\boldsymbol{0},\ \lambda_{ij}^*\geqslant 0,\ \lambda_{ij}^*=0,\ \text{若 } j\notin I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)\right\}$.

由定理 2.5 知,在满足此定理的条件下,$\boldsymbol{\xi}_i\in\partial q_i(\boldsymbol{x}^*)$ 的充要条件为

$$\begin{cases}\boldsymbol{\xi}_i=\nabla_x f_i(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)+\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}^*\nabla_x g_{ij}(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*),\\ \boldsymbol{0}=\nabla_{y_i}f_i(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)+\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}^*\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*),\\ \lambda_{ij}^*\geqslant 0,\\ \lambda_{ij}^*=0,\ \text{若 } j\notin I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*).\end{cases}\tag{1}$$

假如有另外的 $\boldsymbol{\xi}_i'=\boldsymbol{\xi}_i+\boldsymbol{\xi}_i''\in\partial q_i(\boldsymbol{x}^*)$,由定理 2.5 知存在 $\lambda_i'=\lambda_i^*+\lambda_i''$,使得

$$\begin{cases}\boldsymbol{\xi}_i'=\nabla_x f_i(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)+\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}'\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*),\\ \boldsymbol{0}=\nabla_{y_i}f_i(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)+\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}'\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*),\\ \lambda_{ij}'\geqslant 0,\\ \lambda_{ij}'=0,\ \text{若 } j\notin I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*).\end{cases}\tag{2}$$

把(1)式与(2)式中对应式子相减得

$$\boldsymbol{\xi}_i''=\sum_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*),$$

$$\boldsymbol{0}=\sum_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*),\ \lambda_{ij}''\geqslant 0,\ \text{若 } \lambda_{ij}^*=0.$$

因此,由引理 2.4 知 $q_i(\boldsymbol{x}^*)$ 在点 $\boldsymbol{x}^*$ 处可微的充要条件为 $\boldsymbol{\xi}''=\boldsymbol{0}$,即 $q_i(\boldsymbol{x}^*)$ 在点 $\boldsymbol{x}^*$ 处可微的充要条件是,若 $\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)=\boldsymbol{0}$ 且 $\lambda_{ij}''\geqslant 0$,若 $\lambda_{ij}^*=0$,能推出 $\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=\boldsymbol{0}$. 显然,若 $\{\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*),\ j\in I_i(\boldsymbol{x}^*,\ \boldsymbol{y}_i^*)\}$ 线性独立,由 $\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_{y_i}g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=\boldsymbol{0}$ 可推出 $\sum\limits_{j\in I_i(\boldsymbol{x}^*,\boldsymbol{y}_i^*)}\lambda_{ij}''\nabla_x g_{ij}(\boldsymbol{x}^*,\boldsymbol{y}_i^*)=0$.

由此可得出以下定理.

定理 2.6. 若假设(1)—(3)和(5)成立, $\forall \boldsymbol{x}^* \in X$, 且有 $q_i(\boldsymbol{x}^*) = f_i(\boldsymbol{x}^*, \boldsymbol{y}_i^*)$, $g_{ij}(\boldsymbol{x}^*, \boldsymbol{y}_i^*) \leqslant 0$, 则 $q_i(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处可微, $\nabla_x q_i(\boldsymbol{x}^*)$ 由(1)式确定.

定理 2.7. 若假设(1)—(5)成立, $\forall \boldsymbol{x}^* \in X$, 且有 $q_i(\boldsymbol{x}^*) = f_i(\boldsymbol{x}^*, \boldsymbol{y}_i^*)$, $g_{ij}(\boldsymbol{x}^*, \boldsymbol{y}_i^*) \leqslant 0 (j = 1, \cdots, k_i)$, 则 $p(\boldsymbol{x}) \triangleq F(\boldsymbol{x}, q(\boldsymbol{x}))$ 在点 $\boldsymbol{x}^*$ 处可微, 且

$$\nabla_x p(\boldsymbol{x}^*) = \nabla_x F(\boldsymbol{x}^*, q(\boldsymbol{x}^*)) + \sum_{i=1}^{m} \frac{\partial F}{\partial q_i}(\boldsymbol{x}^*, q(\boldsymbol{x}^*)) \cdot \nabla_x q_i(\boldsymbol{x}^*).$$

证明. 因为 $F(\boldsymbol{x}, \boldsymbol{q})$ 和 $\boldsymbol{q}(\boldsymbol{x})$ 在定理条件下是可微的,由链式规则即可推出.

定理 2.8. 若假设(1)—(5)成立,则 $q_i(\boldsymbol{x})$, $F(\boldsymbol{x}, \boldsymbol{q}(\boldsymbol{x}))(i = 1, \cdots, m)$ 在X的内部连续可微.

证明. 若假设(1)—(5)成立,由定理 2.6 和 2.7 知 $q_i(\boldsymbol{x}), F(\boldsymbol{x}, \boldsymbol{q}(\boldsymbol{x}))(i = 1, \cdots, m)$ 在X上为凸的和可微的,则由文 [8] Corollary 25.5.1 知 $q_i(\boldsymbol{x})$, $F(\boldsymbol{x}, q(\boldsymbol{x}))$ 在X内部连续可微.

3　二层决策问题 (P_1) 的神经网络算法

由上面分析可知,在满足假设(1)—(5)的条件下, 二层决策问题 (P_1) 可转变成如下的单层规划问题

$$\begin{aligned} &\min p(\boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{q}(\boldsymbol{x})), \\ &\text{s.t. } \boldsymbol{x} \in X = \{\boldsymbol{x} \in R^n \mid \boldsymbol{h}(\boldsymbol{x}) \leqslant \boldsymbol{0}\}. \end{aligned} \tag{P_2}$$

由于 $p(\boldsymbol{x}) = F(\boldsymbol{x}, q(\boldsymbol{x})), \boldsymbol{h}(\boldsymbol{x})$ 分别为凸的和连续可微的,所以问题 (P_2) 可以用 Frank-Wolfe 方法求解. 下面再结合神经网络方法提出问题 (P_1) 的求解算法.

假设凸规划问题

$$\begin{aligned} &\min f(\boldsymbol{x}), \\ &\text{s.t. } g_1(\boldsymbol{x}) \leqslant 0, \\ &\qquad \vdots \\ &\qquad g_m(\boldsymbol{x}) \leqslant 0. \end{aligned} \tag{P_3}$$

这里 $\boldsymbol{x} \in R^n$, $f(\boldsymbol{x})$, $g_i(\boldsymbol{x})(i = 1, \cdots, m)$ 为连续可微的凸函数. 优化问题 (P_3) 可转化为解辅助微分方程系统问题

$$\frac{dx_j}{dt} = -\mu\left[\frac{\partial f(\boldsymbol{x})}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(\boldsymbol{x})}{\partial x_j}\right], \quad x_j(\boldsymbol{0}) = x_j^{(0)}, \tag{3}$$

$$\frac{d\lambda_i}{dt} = \begin{cases} \mu g_i(\boldsymbol{x}), & \text{若 } g_i(\boldsymbol{x}) > 0, \\ 0, & \text{若 } g_i(\boldsymbol{x}) \leqslant 0. \end{cases} \tag{4}$$

这里 $\mu > 0$; $\lambda_i \geqslant 0$; $i = 1, \cdots, m$; $j = 1, \cdots, n$. (3)式表示问题 (P_3) 的 Lagrange 函数沿 $x_j(t)(j = 1, \cdots, n)$ 轨道减小,(4)式表示问题 (P_3) 的 Lagrange 函数沿 $\lambda_i(t)$ $(i = 1, \cdots, m)$ 轨道增加. 对于微分方程(3),(4),很容易建立如图 1 所示的人工神经网络.

若 $f(\boldsymbol{x})$, $g(\boldsymbol{x})$ 为严格凸的和连续可微的,则系统(3),(4)收敛到问题 (P_3) 的最优解;否则有可能在问题 (P_3) 的平衡点附近振荡. 因此,把微分方程(3),(4)修正为

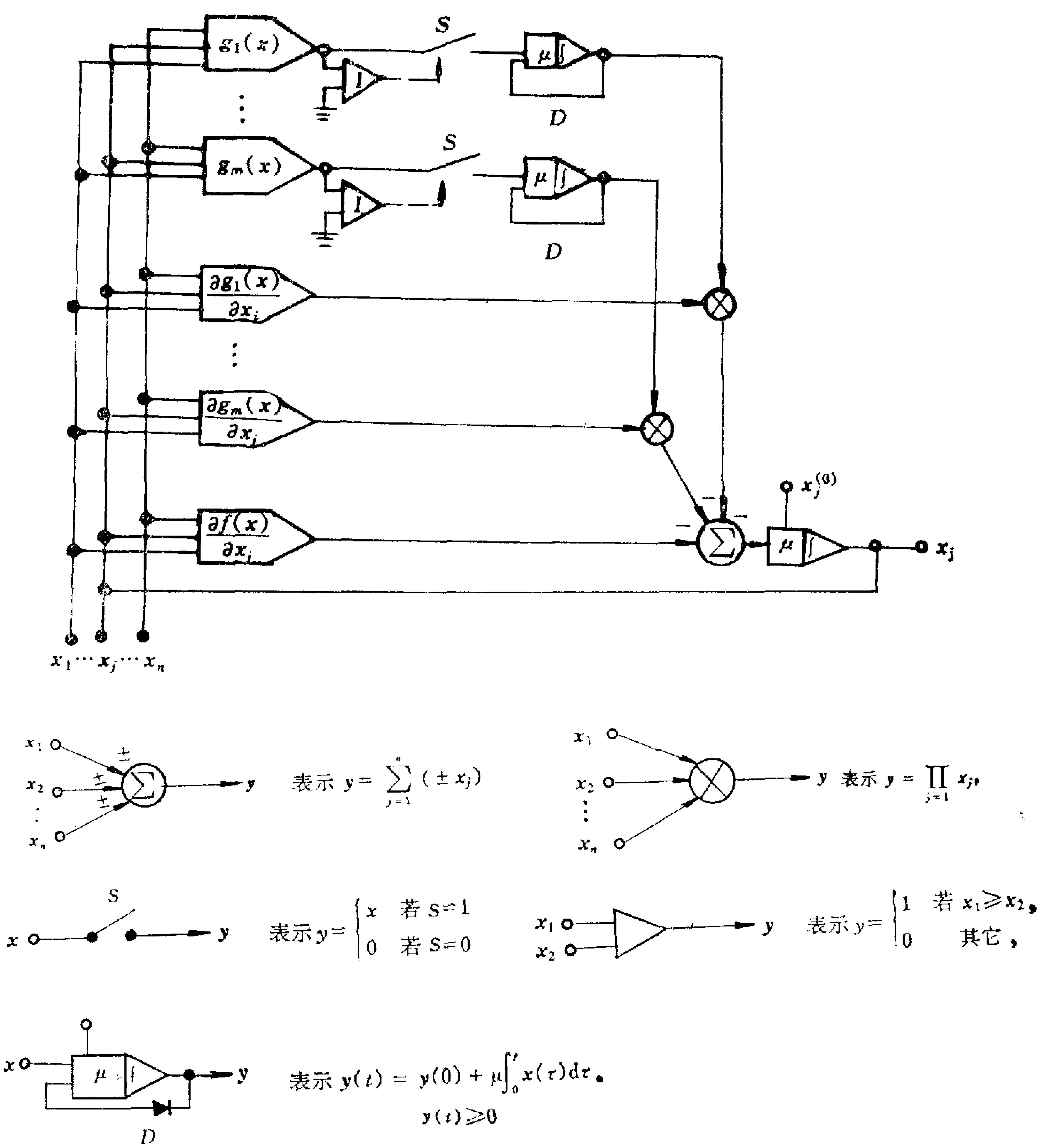

图1 基于 Lagrange 乘子法具有不等式约束的非线性优化问题的人工神经网络示意图

$$\frac{dx_j}{dt}=-\mu\left[\frac{\partial f(\boldsymbol{x})}{\partial x_j}+\sum_{i=1}^{m}\lambda_i\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}\right], \tag{5}$$

$$\frac{d\lambda_i}{dt}=\begin{cases}\mu[g_i(\boldsymbol{x})-\alpha\lambda_i], & 若\ g_i(\boldsymbol{x})>0,\\ 0, & 若\ g_i(\boldsymbol{x})\leqslant 0.\end{cases} \tag{6}$$

这里 $\mu>0$, $\alpha>0$, $\lambda_i\geqslant 0$.

对于微分方程(3),(4),很容易被转化为离散时间的迭代算法

$$x_j^{(k+1)}=x_j^{(k)}-\eta\frac{\partial f(\boldsymbol{x}^{(k)})}{\partial x_j}-\eta\sum_{i=1}^{m}\lambda_i^{(k)}\frac{\partial g_i(\boldsymbol{x}^{(k)})}{\partial x_j},\ j=1,\cdots,n, \tag{7}$$

$$\lambda_i^{(k+1)}=[\lambda_i^{(k)}+\eta g_i(\boldsymbol{x}^{(k)})]_+,\ i=1,\cdots,m. \tag{8}$$

这里 $\eta>0$, $x_j^{(0)}=x_j(0)$, $\lambda_i^{(0)}=\lambda_i(0)$, $[y]_+=\max\{0,y\}$.

如上所述的算法是并行方式，它适宜于用并行的模拟电路来实现. 为了便于在串行

数字式计算机上模拟，可把(7)，(8)式改写成串行方式

$$\begin{cases} x_j^{(k+1)} = x_j^{(k)} + \eta \dfrac{\partial f(\boldsymbol{x}^{(k)})}{\partial x_j} - \eta \sum\limits_{i=1}^{m} \lambda_i^{(k)} \dfrac{\partial g_i(\boldsymbol{x}^{(k)})}{\partial x_j}, \\ x_l^{(k+1)} = x_l^{(k)},\ l = 1,\cdots,n,\ l \neq j, \\ \lambda_i^{(k+1)} = [\lambda_i^{(k)} + \eta g_i(\boldsymbol{x}^{(k)})]_+,\ j = 1,\cdots,n,\ i = 1,\cdots,m. \end{cases} \tag{9}$$

这里 $\eta > 0$，$x_j^{(0)} = x_j(0)$，$\lambda_i^{(0)} = \lambda_i(0)$，$[y]_+ = \max\{0, y\}$. 显然，这在计算机上很容易使用迭代算法. 由(7)—(9)式可以看出，并行方式要比串行方式的计算速度快得多，这就是神经网络的一大特点.

下面结合上述的神经网络，提出问题 (P_1) 的神经网络算法的具体步骤

1）选择一初始点 $\boldsymbol{x} \in X^q$，令 $q = 1$.

2）计算 $\nabla_x q_i(\boldsymbol{x}^q)$ $(i = 1,\cdots,m)$ 和 $\nabla_x p(\boldsymbol{x}^q)$.

由定理 2.5 知，$\nabla_x q_i(\boldsymbol{x}^q) = \nabla_x f_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q) + \sum\limits_{j=1}^{k_i} \lambda_{ij}^q \nabla_x g_{ij}(\boldsymbol{x}^q, \boldsymbol{y}_i^q)$，且 $\boldsymbol{y}_i^q, \lambda_{ij}^q$ 满足

$$\begin{aligned} & q_i(\boldsymbol{x}^q) = f_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q),\ g_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q) \leqslant \boldsymbol{0}, \\ & \nabla_{y_i} f_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q) + \sum_{j=1}^{k_i} \lambda_{ij}^q \nabla_{y_i} g_{ij}(\boldsymbol{x}^q, \boldsymbol{y}_i^q) = \boldsymbol{0}, \\ & \lambda_{ij}^q \geqslant 0, \\ & \lambda_{ij}^q = 0,\ 若\ j \notin I_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q). \end{aligned}$$

所以 $\boldsymbol{y}_i^q$，λ_{ij}^q $(j = 1,\cdots,k_i)$ 是优化问题

$$\begin{aligned} & q_i(\boldsymbol{x}^q) = \min_{y_i} f_i(\boldsymbol{x}^q, \boldsymbol{y}_i), \\ & \text{s.t. } \boldsymbol{g}_i(\boldsymbol{x}^q, \boldsymbol{y}_i) \leqslant \boldsymbol{0},\ (i = 1,\cdots,m) \end{aligned} \tag{P_4}$$

的最优解和对应的 Kuhn-Tucker 乘子.

因为 $f_i(\boldsymbol{x}, \boldsymbol{y}_i)$，$\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{y}_i)$ 是关于 $(\boldsymbol{x}, \boldsymbol{y}_i)$ 连续可微的凸函数，所以当 $\boldsymbol{x} = \boldsymbol{x}^q$ 时 $f_i(\boldsymbol{x}^q, \boldsymbol{y}_i)$，$\boldsymbol{g}_i(\boldsymbol{x}^q, \boldsymbol{y}_i)$ 是关于 $\boldsymbol{y}_i$ 的连续可微凸函数. 这样，由上面提出的神经网络可求解优化问题 (P_4)，得最优解 $\boldsymbol{y}_i^q$，λ_{ij}^q，$j = 1,\cdots,k_i$，则

$$\nabla_x q_i(\boldsymbol{x}^q) = \nabla_x f_i(\boldsymbol{x}^q, \boldsymbol{y}_i^q) + \sum_{j=1}^{k_i} \lambda_{ij}^q \nabla_x g_{ij}(\boldsymbol{x}^q, \boldsymbol{y}_i^q),\ i = 1,\cdots,m,$$

$$\nabla_x p(\boldsymbol{x}^q) = \nabla_x F(\boldsymbol{x}^q, q(\boldsymbol{x}^q)) + \sum_{i=1}^{m} \frac{\partial F}{\partial q_i}(\boldsymbol{x}^q, q(\boldsymbol{x}^q)) \cdot \nabla_x q_i(\boldsymbol{x}^q).$$

3）求解优化问题

$$\begin{aligned} & \min \nabla_x p(\boldsymbol{x}^q) \cdot (\boldsymbol{z} - \boldsymbol{x}^q), \\ & \text{s.t. } \boldsymbol{h}(\boldsymbol{z}) \leqslant 0. \end{aligned} \tag{P_5}$$

由于 $\nabla_x p(\boldsymbol{x}^q) \cdot (\boldsymbol{z} - \boldsymbol{x}^q)$ 是 $\boldsymbol{z}$ 的线性函数，当然是连续可微的凸函数；由已知 $\boldsymbol{h}(\boldsymbol{z})$ 也是连续可微的凸函数. 所以，优化问题 (P_5) 也可用上述的神经网络求解.

4）寻找步长，亦即求优化问题

$$\begin{aligned} & \min F(\boldsymbol{x}^q + s(\boldsymbol{z}^q - \boldsymbol{x}^q), q(\boldsymbol{x}^q + s(\boldsymbol{z}^q - \boldsymbol{x}^q))), \\ & \text{s.t. } 0 \leqslant s \leqslant 1. \end{aligned}$$

由于 $q(\cdot)$ 为隐函数,不能直接求解,可采取试探法. 在此可把区间 [0, 1.0] 分成若干等份,例如分成 10 等份. 让 s 分别取值 0, 0.1, ……, 0.9, 1.0, 并由神经网络计算出 $q(\cdot)$ 的值,再由 $q(\cdot)$ 的值计算出 F 的值,选取使 F 为最小的 s 值 s^q, 令

$$x^{q+1} = x^q + s^q(z^q - x^q).$$

5) 若 $F(x^q, q(x^q)) - F(x^{q+1}, q(x^{q+1})) \leqslant \alpha$, 计算结束; 否则把 x^{q+1} 当做初始点,转到 1), 继续迭代.

本文提出的二层决策问题 (P_1) 的神经网络算法是收敛的. 事实上, 此算法是以 Frank-Wolfe 算法为基础的,序列 $\{x^q\}$ 的任意极限点 x^0 是决策问题 (P_1) 的上层最优决策;下层各决策最优解可用神经网络求优化问题

$$q_i(x^0) = \min_{y_i} f_i(x^0, y_i),$$

$$\text{s.t. } g_i(x^0, y_i) \leqslant 0, \quad (i = 1, 2, \cdots, m).$$

4 算例

例. $$\min_x F(x, q) = x_1^2 + (x_2 - 5)^2 + \frac{1}{2} q^2.$$

$$\text{s.t. } 0 \leqslant x_1 \leqslant 2, \quad 1 \leqslant x_2 \leqslant 5,$$

$$q(x) = \min_y (y-1)^2 + (2x_1 + x_2 - 4)^2;$$

$$\text{s.t. } x_1 + 3x_2 - y \geqslant 3, \quad 2x_1 - x_2 + 0.5y \geqslant -4,$$

$$-x_2 - y \geqslant -7, \quad y \geqslant 0.$$

容易验证,本例满足假设 (1)—(5), 且 $q(x)$ 为严格凸函数.

取 $x^1 = (x_1^1, x_2^1) = (1, 2)$ 为初始点,第一轮迭代求出 $x^2 = (0.1, 4.7)$; 第二轮迭代以 x^2 为初始点,求出 $x^3 = (0.096, 4.552)$. 此时 $F(x^2, q(x^2)) - F(x^3, q(x^3)) = 0.428 - 0.353 = 0.075 < 0.1$.

当取 $\alpha = 0.1$ 时,则迭代停止,此时上层决策变量 $x = (0.096, 4.552)$, 下层决策变量 $y = 1.0$; 若取 $\alpha = 0.05$, 可继续迭代.

5 结论

本文分析了一类二层决策问题的若干解析性质, 并在此基础上提出了基于 Frank-Wolfe 方法的人工神经网络算法. 由于 Frank-Wolfe 方法迭代过程简单、迅速, 而人工神经网络用于求解优化问题又具有快速收敛的特点. 因此, 把它们结合起来用于求解二层决策问题是有意义的.

参 考 文 献

[1] 郑应平. 多人多级递阶决策的几个问题—鼓励性对策及模型简化. 自动化学报, 1985, 11(4):372—378.

[2] Bard J F. Convex Two-level optimization. *Mathematical Programming*, 1988, 40: 15—27.

[3] Outrata J V. Necessary optimality conditions for stackelberg problems. *Journal of optimization theory and applications*, 1993, 76(2): 305—320.

[4] Bard J F. An algorithm for solving the general bilevel programming problem. *Mathe of Operations Research*, 1983, 8: 260—272.

[5] Rockafella. The conjugate duality and optimization. Philadelphia: SIAM, 1974.
[6] Rockafella. The theory of subgradients and its applications to problems of optimization: convex and nonconvex functions. Berlin: Heldermann Verlag, 1981.
[7] 金洪臻等译. 不可微量优化. 大连：大连理工大学出版社，1991:154—160.
[8] Rockafella. Convex analysis. Princeton: Princeton University Press, 1970: 213—260.
[9] 吕庆喆，盛昭瀚，徐南荣. 一类二层决策问题的 Frank-Wolfe 算法. 东南大学学报，1996，26(1):74—80.
[10] Hopfield J J, Tank D W. "Neural" computation of decision in optimization problems. *Biol. Cybern.*, 1985, **52:** 141—152.
[11] Kennedy M P, Chua L O. Neural networks for linear and nonlinear programming. *IEEE Trans on Circuits and systems*. 1988, **35:** 554—562.
[12] Rodriguez Vazquez. Nonlineal switched——capacitor neural networks for optimization problems. *IEEE Trans on Circuits and Systems*. 1990, **37:** 384—395.

A NEW ALGORITHM BASED ON THE FRANK-WOLFE METHOD AND NEURAL NETWORK FOR A CLASS OF BILEVEL DECISSION MAKING PROBLEMS

SHENG ZHAOHAN LÜ QINGZHE XU NANRONG
(*Economics and Management College, Southeast University, Nanjing* 210018)

ABSTRACT

In this paper a model for a class of bilevel decision-making problems is presented in the form of a static Stackelberg game. Under certain convexity assumptions, some properties of the model is studied. According to these properties, this paper presents a new algorithm based on the Frank-Wolfe method and neural network method. An example demonstrates the feasibility of the proposed algorithm.

Key words: Bilevel decision-making, the Frank-Wolfe method, artificial neural networks.

盛昭瀚 简介及照片见本刊第 19 卷第 6 期。

吕庆喆 1995 年 5 月在东南大学经济管理学院获博士学位。已发表学术论文 10 余篇。现在国家统计局国际统计信息中心工作，主要从事世界经济、统计方法等方面的研究。

徐南荣 简介及照片见本刊第 19 卷第 6 期。