文章编号: 1000-8152(2002)04-06-0521

# 一类受控闭排队网络基于性能势的最优性方程*

周亚平[1], 奚宏生[2], 殷保群[2], 孙德敏[2]
(1. 中国科技大学 管理科学系,合肥 230026; 2. 中国科技大学 自动化系,合肥 230026)

**摘要:** 研究一类受控闭排队网络系统的性能优化问题.文章引进了两个基本概念:折扣代价 $\alpha$-性能势和平均代价性能势,并且讨论了这两个性能势之间的一个关系式.在一般的假设条件下,我们应用性能势的基本性质直接建立了无限时间水平平均代价模型的最优性方程,并且证明了在紧致集上最优解的存在性.最后给出了一个策略优化的迭代算法并通过一个实际算例以说明该算法的效果.

**关键词:** 闭排队网络系统; 性能势; 最优性方程; 最优解的存在性
**中图分类号:** O226　　**文献标识码:** A

# Optimality equations based performance potentials for a class of controlled closed queueing networks

ZHOU Ya-ping[1], XI Hong-sheng[2], YIN Bao-qun[2], SUN De-min[2]
(1. Department of Management Science, China University of Science and Technology, Hefei 230026, China;
2. Department of Automation, China University of Science and Technology, Heifei 230026, China)

**Abstract:** This paper deals with the performance optimization problem of a class of controlled closed queueing network systems (CQNS). We introduce two fundamental concepts: the discounted cost $\alpha$- performance potentials and average cost performance potentials, and consider a fundamental relation between the two potentials. Under a general assumption, we establish directly the optimality equation for infinite time horizon average cost model and prove the existence of optimal solution in a compact action set by using properties of the performance potentials, suggest an policy-optimality algorithm and give a numerical example to illustrate the application of the proposed algorithm.

**Key words:** closed queueing network systems; performance potentials; optimality equation; existence of optimal solution

## 1 引言(Introduction)

闭排队网络系统(CQNS)是随机离散事件动态系统(DEDS)的主要数学模型之一.它被广泛地用于模拟计算机和通讯网络、柔性制造和公共服务等实际系统.所以 CQNS 的性能分析和优化问题一直受到控制界的广泛关注.鉴于 CQNS 通常能被描述成一类具有有限状态空间的不可约、正常返的连续时间 Markov 向量过程,这样一类过程的性能优化问题通常被归结为 Markov 决策过程,此优化方法在理论上具有较强的约束条件,有些假设条件对于多数实际系统是难以满足或验证的[1,2].最近文献[3,4]提出了一种新的概念: Markov 性能势,并揭示了 Markov 势论,摄动分析和 Markov 决策过程三者之间的紧密联系,文献[5,6]将文献[3]的结果推广到了排队网络系统,建立了排队网络性能势理论.本文主要研究一类在平稳控制策略驱动下的受控闭排队网络系统的性能优化问题.我们引进了无限水平折扣代价 $\alpha$- 性能势和平均代价性能势的概念,在一个相对较弱的假设条件下直接导出了基于性能势的最优性方程及其在紧致行动集上的最优解的存在性定理,并给出了策略优化迭代算法,通过一个实际算例以说明该算法的应用.

## 2 问题的描述(Problem description)

我们考虑一个具有 $M$ 个单类服务节点和 $N$ 个顾客的闭排队网络系统(CQNS).假设每个服务节点拥有无限容量的缓冲器并且服从先到先服务规则,顾客对每个服务节点的服务要求服从均值为 1 的指数分布.顾客从服务者 $i$ 转移到服务者 $j$ 的概率记为 $q_{i,j}$,则路径概率矩阵记为 $Q=[q_{i,j}]_{M\times M}$,它是一个不可分随机矩阵.这样一个 CQNS 的状态过程 $N(t), t\geqslant 0$ 是一个不可约、正常返的 Markov 向量过程.

* 基金项目:国家自然科学基金(69974037)和国家高性能计算基金(00212)资助项目.
收稿日期:2000-08-29; 收修改稿日期:2001-08-06.

记

$$\Phi = \{n = (n_1, n_2, \cdots, n_M): \sum_{i=1}^{M} n_i = N\}$$

为系统的状态空间,其中 $n_i$ 表示第 $i$ 个服务节点处的顾客数,我们将 $\Phi$ 中的元素以字典序排列,为了表达简洁起见以有序自然数列 $n = 1,2,\cdots,K$ 来表示这些元素,

$$K = \binom{N+M+1}{N}.$$

设 $\Lambda(n)$ 为在状态 $n$ 所采用的行动集,它是一个紧致集,并记 $\Lambda = \bigcup_{n=1}^{K}\Lambda(n)$ 为行动空间.一平稳控制策略是一个确定性映射 $v:\Phi \to \Lambda$,对任意状态 $n$,有

$$v(n) = (\mu_{1,n}, \mu_{2,n}, \cdots, \mu_{M,n}) = \beta_n \in \Lambda(n) \subset \Lambda,$$

其中 $\mu_{i,n}$ 是服务者 $i$ 在状态 $n$ 的服务率(控制决策变量).满足当 $n_i = 0$ 时,$\mu_{i,n} = 0$;当 $n_i \neq 0$ 时,$\mu_{i,n} > 0$, $i = 1,2,\cdots,M$,并记 $\Omega_s$ 是全体平稳控制策略集.

设 $P^v(t) = [P_{n,m}(t, v(n))]_{n=1,\ m=1}^{K,\ K}$ 是 CQNS 的状态过程在给定的平稳控制策略 $v$ 下的状态转移概率矩阵.令 $\{T_l\}_{l=0}^{\infty}$ 是系统状态转移时刻,记 $\{X_l\}_{l=0}^{\infty}$ 是 $N(t), t \geq 0$ 的一嵌入 Markov 链,其中 $X_l = N(T_l + 0)$,它的状态转移矩阵记为(参看文献[7])

$$P^v = [P_{n,m}(v(n))]_{n=1,\ m=1}^{K,\ K},$$

与之对应的无穷小生成矩阵为

$$A^v = [a_{n,m}(v(n))]_{n=1,\ m=1}^{K,\ K},$$

其中 $A^v = U^v(P^v - I)$, $U^v = \mathrm{diag}(\mu(1), \mu(2), \cdots, \mu(K))$, $\mu(n) = \sum_{i=1}^{M}\mu_{i,n}$,是 $M$ 个服务节点在状态 $n$ 的总服务率.记 $\pi^v = (\pi_1^v, \cdots, \pi_K^v)$ 是稳态概率,$e = (1,1,\cdots,1)^\tau$,则有

$$\pi^v e = 1,\ A^v e = 0,\ \pi^v A^v = 0. \qquad (2.1)$$

**假设 1** 对任意的 $n, m \in \Phi$, $P_{n,m}(t, \beta_n)$ 是定义在 $\Lambda(n)$ 上的连续函数.

**假设 2** 设 $f: \Phi \times \Lambda \to \mathbb{R}$ 是实映射,记

$$f^v = (f(1, v(1)), \cdots, f(K, v(K)))^\tau,$$

对每一 $n \in \Phi$, $f(n, \beta_n)$ 是定义在 $\Lambda(n)$ 上的有界连续函数.

我们称 $\widetilde{N} = (N(t), \Phi, \Lambda, P^v(t), f^v)$ 为约束在 $\Omega_s$ 上的受控 CDNS 过程.在这一过程中,系统的状态转移规律与控制决策所选用的方案相互作用决定了过程的发展,问题是如何选择控制决策方案,使系统运行的全过程在某种准则下达到最优的运行效果.

## 3 性能势(Performance potentials)

设 $\widetilde{N}$ 的无限时间水平总折扣代价期望值为

$$\Psi_\alpha^v(n) = E\{\int_0^\infty e^{-\alpha t} f(N(t), v(N(t)))\mathrm{d}t \mid N(0) = n\} = E_n\{\int_0^\infty e^{-\alpha t} f(N(t), v(N(t)))\mathrm{d}t\}, \qquad (3.1)$$

其中 $v \in \Omega_s$, $n \in \Phi$, $\alpha > 0$ 是给定的折扣因子.它的向量形式为

$$\Psi_\alpha^v = (\Psi_\alpha^v(1), \Psi_\alpha^v(2), \cdots, \Psi_\alpha^v(K))^\tau.$$

文献[8]给出了下面的定义 1 及引理 1.

**定义 1** 对于给定的 $\alpha > 0$,记

$$R^\alpha = \int_0^\infty \mathrm{e}^{-\alpha t} P(t)\mathrm{d}t,$$

称 $R^\alpha$ 为 CQNS 的状态过程 $N(t), t \geq 0$ 的 $\alpha$-势,其中 $P(t)$ 是 $N(t)$ 的状态转移矩阵.

**引理 1** 在定义 1 下,$R^\alpha$ 具有下述性质:

1) $(\alpha I - A)R^\alpha = R^\alpha(\alpha I - A) = I$,其中 $A$ 是 $N(t)$ 的无穷小生成矩阵;

2) $R^\alpha = (\alpha I - A)^{-1}$;

3) $\alpha R^\alpha = I + AR^\alpha$.

我们将现代 Markov 过程的 $\alpha$-势论引入 $\widetilde{N}$,给出下面的定义.

**定义 2** 对给定的 $\alpha > 0$ 和 $v \in \Omega_s$,记

$$g_\alpha^v = R^\alpha f^v = \int_0^\infty e^{-\alpha t} P^v(t) f^v \mathrm{d}t,$$

称 $g_\alpha^v$ 为 $\widetilde{N}$ 关于性能代价函数 $f^v$ 在控制策略下的一个 $\alpha$-性能势.

注意到,$g_\alpha^v$ 是 $K$ 维向量,它的第 $n$ 个分量为

$$g_\alpha^v(n) = \int_0^\infty e^{-\alpha t} \sum_{m=1}^{K} P_{n,m}^v(t) f(m, v(m))\mathrm{d}t = E_n\{\int_0^\infty e^{-\alpha t} f(N(t), v(N(t)))\mathrm{d}t\},\ n \in \Phi, \qquad (3.2)$$

我们有,$g_\alpha^v(n) = \Psi_\alpha^v(n), n \in \Phi$, $\alpha$-性能势向量 $g_\alpha^v = \Psi_\alpha^v$.

定义 2 表明,在控制策略 $v$ 下 $\widetilde{N}$ 的 $\alpha$-性能势就是折扣代价准则函数.由引理 1,并应用类似于文献[1]和[2]的方法能直接得到下述结论.

**定理 1** 对任意给定的 $\alpha > 0$,设 $V$ 是定义在 $\Omega_s$ 上的 $K$ 维有界实向量函数集,存在唯一的平稳最优控制策略 $v^* \in \Omega_s$ 和 $g_\alpha^{v^*} \in V$,满足最优性方程

$$0 = \min_{v \in \Omega_s}\{f^v - (\alpha I - A^v) g_\alpha^{v^*}\}. \qquad (3.3)$$

等价地,设

$$(v(1), v(2), \cdots, v(K)) = \beta \in \Lambda,$$

$$(v^*(1), v^*(2), \cdots, v^*(K)) = \delta^\infty \in \Lambda,$$

并令

$$\delta = \arg\min_{\beta \in \Lambda}\{f^\beta - (\alpha I - A^\beta) g_\alpha^{\delta^\infty}\}, \quad (3.4)$$

则有

$$0 = f^\delta - (\alpha I - A^\delta) g_\alpha^{\delta^\infty} \leqslant$$

$$f^v - (\alpha I - A^v) g_\alpha^{\delta^\infty}, \ v \in \Omega_s. \quad (3.5)$$

$\tilde{N}$ 的另一个重要的准则是无限时间水平的平均代价期望值,它被定义为

$$\eta^v = \lim_{T \to \infty} \frac{1}{T} E\{\int_0^T f(N(t), v(N(t))) \mathrm{d}t\}, \ v \in \Omega_s. \quad (3.6)$$

由 Markov 过程的遍历和稳定性知[8]

$$\eta^v = \sum_{m=1}^{K} \pi_m^v f(m, v(m)) = \pi^v f^v, \ v \in \Omega_s. \quad (3.7)$$

文献[5]和[6]给出了 CQNS 无穷小矩阵及其群逆 $A^\#$ 的性质,对于一个 $\tilde{N}$ 也有类似的结论.

**引理 2**　对任意的给定的 $v \in \Omega_s$, $\tilde{N}$ 的无穷小矩阵 $A^v$ 和它的群逆 $A^{v\#}$,具有下述性质:

1) $(-A^v + e\pi^v)$ 是非奇异的;

2) $-A^{v\#} = (-A^v + e\pi^v)^{-1} - e\pi^v$,且 $A^v A^{v\#} = A^{v\#} A^v = I - e\pi^v$.

证　1) 对任意给定的 $v \in \Omega_s$, $\mathrm{rank} A^v = K - 1$,故存在一依赖 $v$ 的非奇异的矩阵 $W^v$,使

$$A^v = (W^v)^{-1} \begin{bmatrix} A_1^v & 0 \\ 0 & 0 \end{bmatrix} W^v,$$

其中 $A_1^v$ 是非奇异矩阵,且它的特征值 $\lambda(A_1^v) < 0$.

$$e\pi^v = \lim_{t \to \infty} P^v(t) = \lim_{t \to \infty} e^{A^v t} =$$

$$\lim_{t \to \infty} (W^v)^{-1} \begin{bmatrix} e^{A_1^v t} & 0 \\ 0 & 1 \end{bmatrix} W^v = (W^v)^{-1} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W^v,$$

故 $-A^v + e\pi^v$ 非奇异.

$$2)\ -(A^v)^\# = (W^v)^{-1} \begin{bmatrix} -(A_1^v)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W^{v^v} =$$

$$(-A^v + e\pi^v)^{-1} - e\pi^v.$$

易验证,$A^v A^{v\#} = A^{v\#} A^v = I - e\pi^v$.

考虑线性方程组

$$(-A^v + e\pi^v) g^v = f^v, \quad (3.8)$$

其唯一解向量为

$$g^v = (-A^v + e\pi^v)^{-1} f^v, \quad (3.9)$$

用 $\pi^v$ 左乘(3.8)式两边得

$$\pi^v g^v = \pi^v f^v = \eta^v. \quad (3.10)$$

**定义 3**　对任意平稳策略 $v \in \Omega_s$,称 $g^v + ec$ 为 $\tilde{N}$ 的一个性能势向量,$c$ 是任意给定的常数.它的第 $n$ 个分量 $g^v(n) + c$ 为 $\tilde{N}$ 在状态 $n$ 的性能势.

注意到当 $c = 0$ 时,$g^v$ 是一个性能势向量,当 $c = -\eta^v$ 时,由引理 2 和(3.9)式得

$$g^v - e\eta^v = -A^{v\#} f^v = \tilde{g}^v. \quad (3.11)$$

在策略 $v$ 下 $\tilde{g}^v$ 也是一个性能势向量,任意两个给定的性能势向量之间仅差一个常向量.

记任意两个状态 $n, m$ 的性能势之差为 $d_{n,m}^v = g^v(m) - g^v(n) = \tilde{g}^v(m) - \tilde{g}^v(n)$. 我们有

$$d_{n,m}^v = \lim_{T \to \infty} \{E_m[\int_0^T f(N(t), v(N(t))) \mathrm{d}t] -$$

$$E_n[\int_0^T f(N(t), v(N(t))) \mathrm{d}t]\}.$$

上式左边相当于 $\alpha$-性能势(3.3)式中,当 $\alpha = 1$ 时两个分量 $g_1^v(m)$ 和 $g_1^v(n)$ 之差.虽然 $g_1^v(m)$ 和 $g_1^v(n)$ 均是无限的,然而它们的差收敛到一有限值.这一结果在 CQNS 的摄动分析理论中具有十分重要的意义[5,6]. $g_\alpha^v$ 和 $g^v$ 的一个重要关系式被给出在下述定理中.

**定理 2**　设 $g_\alpha^v$ 是 $\tilde{N}$ 的折扣代价模型的一个 $\alpha$-性能势,则 $g_\alpha^v$ 在 $\alpha = 0$ 附近能被表示为

$$g_\alpha^v = \left(\frac{1-\alpha}{\alpha}\right) e\eta^v + g^v + F^v(\alpha). \quad (3.12)$$

其中 $\eta^v$ 和 $g^v$ 分别是平均代价模型的性能指标和性能势,$F^v(\alpha) = O(\alpha)$.

证　记 $\rho = \dfrac{1}{\alpha + 1}$, $\alpha > 0, 0 < \rho < 1$,则

$$(\alpha I - A^v)^{-1} = \frac{\rho}{1-\rho}[I + \frac{\rho}{1-\rho} A^v]^{-1} =$$

$$\frac{\rho}{1-\rho}\{e\pi^v + (W^v)^{-1} \begin{bmatrix} \left[I_{K-1} + \dfrac{\rho}{\rho - 1} A_1^v\right]^{-1} & 0 \\ 0 & 0 \end{bmatrix} W^v\}.$$

其中

$$[I_{K-1} + \frac{\rho}{1-\rho} A_1^v]^{-1} =$$

$$[\frac{\rho - 1}{\rho}(A_1^v)^{-1} + I_{K-1}]^{-1} \frac{\rho - 1}{\rho}(A_1^v)^{-1},$$

注意到

$$\lambda\left(\frac{\rho - 1}{\rho}(A_1^v)^{-1}\right) = \frac{\rho - 1}{\rho} \Big/ \lambda(A_1^v) > 0,$$

当 $\alpha$ 适当小时,$\dfrac{\rho - 1}{\rho} \Big/ \lambda(A_1^v) < 1$,由文献[9]及引理 2 知

$$(\alpha I - A^v)^{-1} =$$

$$\frac{\rho}{1-\rho} e\pi^v - \sum_{n=0}^{\infty} \left(\frac{1-\rho}{\rho}\right)^n (W^v)^{-1} \begin{bmatrix} (A_1^v)^{-1} & 0 \\ 0 & 0 \end{bmatrix}^{n+1} W^v =$$

$$\frac{\rho}{1-\rho}e\pi^v - \sum_{n=0}^{\infty}\left(\frac{1-\rho}{\rho}\right)^n[(A^v)^{\#}]^{n+1},$$

$$g_\alpha^v = (\alpha I - A^v)^{-1}f^v =$$

$$\frac{1-\alpha}{\alpha}e\eta^v + g^v - \sum_{n=1}^{\infty}\alpha^n(A^v)^{\#^{n+1}}f^v.$$

由(3.13)式在 $\rho = 1$ 附近是一致收敛的,等价地,上式第3项在 $\alpha = 0$ 附近也是一致收敛的.

令 $F^v(\alpha) = -\sum_{n=1}^{\infty}\alpha^n(A^v)^{\#^{n+1}}f^v$, 显然有 $\lim_{\alpha\to 0}F^v(\alpha) = 0$.

## 4 平均代价模型的最优性方程(Optimality equation of average cost model)

本节我们将给出平均代价模型的最优性方程,为此,我们首先给出

**定理 3** $v^* \in \Omega_s$ 是使(3.6)式达到最小的平稳最优控制策略的充分必要条件为对任意平稳控制策略 $v \in \Omega_s$,

$$f^{v^*} + A^{v^*}g^{v^*} \leqslant f^v + A^v g^{v^*}. \tag{4.1}$$

其中,$\leqslant$表示左边向量中的每个分量都小于或等于右边向量中的每个对应分量.

证 对任意 $v \in \Omega_s, \pi_n^v > 0, n \in \Phi$. 将 $\pi^v$ 左乘在策略 $v^*$ 下的(3.8)式,得 $\pi^{v^*}g^{v^*} = \pi^v A^{v^*}g^{v^*} + \pi^v f^{v^*}$,又由 $\pi^v A^v f^{v^*} = 0$,故

$$\eta^{v^*} - \eta^v = \pi^v[(f^{v^*} + A^{v^*}g^{v^*}) - (f^v + A^v g^{v^*})], \tag{4.2}$$

故当(4.1)式成立时,$\eta^{v^*} \leqslant \eta^v$. 反之,若 $\eta^{v^*} \leqslant \eta^v$,而(4.1)式不成立,则必存在一策略 $v' \in \Omega_s$,且至少有一个状态 $n_0 \in \Phi$,使

$$f(n_0, v'(n_0)) + \sum_{m=1}^{K}a_{nm}(v'(n_0))g^{v^*}(m) <$$

$$f(n_0, v^*(n_0)) + \sum_{m=1}^{K}a_{nm}(v^*(n_0))g^{v^*}(m).$$

我们定义一个策略 $v''$,对 $n \in \Phi$ 为

$$v''(n) = \begin{cases} v'(n_0), & n = n_0, \\ v^*(n), & n \neq n_0, \end{cases}$$

则有 $f^{v''} + A^{v''}g^{v^*} < f^{v^*} + A^{v^*}g^{v^*}$(其中,$<$ 表示 $f$ 左边向量中的每个分量都小于或等于右边向量中的每个对应分量,且至少有一个分量小于右边向量中的对应分量). 由(4.2)式,$\eta^{v''} < \eta^{v^*}$,这与 $v^*$ 是最优策略矛盾.

注意到 $(-A^{v^*} + e\pi^{v^*})g^{v^*} = f^{v^*}$, $e\pi^{v^*}g^{v^*} = f^{v^*} + A^{v^*}g^{v^*}$, $e\eta^{v^*} = f^{v^*} + A^{v^*}g^{v^*}$, 则最优性定理4.1可以等价地表示为 $v^*$ 是最优控制策略的充分必要条件为

$$e\eta^{v^*} = \min_{v\in\Omega_s}\{f^v + A^v g^{v^*}\}. \tag{4.3}$$

(4.3)式被称为 $\tilde{N}$ 的平均代价模型基于性能势的最优性方程.

**引理 3** 设 $\{B_k\}$ 是一矩阵序列,且以模收敛到一可逆矩阵 $B$,则对充分大的 $k$, $B_k^{-1}$ 存在,且 $B_k^{-1}$ 以模收敛到矩阵 $B^{-1}$,记为 $\lim_{k\to\infty}B_k^{-1} = B^{-1}$.

**定理 4** 在假设条件 1,2 下,存在实数 $\eta$ 和一个 $K$ 维向量 $g$,满足

$$0 = \min_{v\in\Omega_s}\{f^v - e\eta + A^v g\}. \tag{4.4}$$

其次,若 $(\eta', g')$ 是满足(4.4)式的任意一个解,则 $\eta' = \eta$.

证 选择一折扣因子序列 $\alpha_k, \alpha_k \downarrow 0$,由定理 1 知,对每一固定的 $\alpha_k$,存在唯一平稳最优控制策略 $v_k^*$,其对应的行动 $\delta_k^\infty \in \Lambda$ 及 $\delta_k \in \Lambda$(见(3.4)式),由于 $\Lambda = \bigcup_{n\in\Phi}\Lambda(n)$ 是紧致的,故存在一子序列 $\{\delta_{k_l}\}$ 收敛到 $\delta \in \Lambda$,且 $\{\delta_{k_l}^\infty\}$ 收敛到 $\delta^\infty \in \Lambda$. 为表达简单起见,我们将 $\{k_l\}$ 仍记为 $\{k\}$. 由(3.5)式,

$$0 = f^{\delta_k} - (\alpha_k I - A^{\delta_k})g_{\alpha_k}^{\delta_k^\infty} \leqslant$$

$$f^v - (\alpha_k I - A^v)g_{\alpha_k}^{\delta_k^\infty}, \ v \in \Omega_s. \tag{4.5}$$

由假设条件 1,易证得 $\alpha_{n,m}(\beta_n)$ 和 $\pi_m(\beta_n)$ 是 $\Lambda(n)$ 上的连续函数,并由假设条件 2 和引理 3,我们有

$$\lim_{k\to\infty}\eta^{\delta_k^\infty} = \lim_{k\to\infty}\pi^{\delta_k^\infty}f^{\delta_k^\infty} = \pi^{\delta^\infty}f^{\delta^\infty} = \eta^{\delta^\infty},$$

$$\lim_{k\to\infty}g^{\delta_k^\infty} = \lim_{k\to\infty}(-A^{\delta_k^\infty} + e\pi^{\delta_k^\infty})^{-1}f^{\delta_k^\infty} =$$

$$(-A^{\delta^\infty} + e\pi^{\delta^\infty})^{-1}f^{\delta^\infty} = g^{\delta^\infty}.$$

当 $\alpha_k$ 适当小时,将定理 2 的(3.12)式代入(4.5)得

$$0 = f^{\delta_k} - e\eta^{\delta_k^\infty} - (\alpha_k I - A^{\delta_k})g^{\delta_k^\infty} +$$

$$(\alpha_k I - A^{\delta_k})e\eta^{\delta_k^\infty} - (\alpha_k I - A^{\delta_k})F^{\delta_k^\infty}(\alpha_k) \leqslant$$

$$f^v - e\eta^{\delta_k^\infty} - (\alpha_k I - A^v)g^{\delta_k^\infty} + (\alpha_k I -$$

$$A^v)e\eta^{\delta_k^\infty} - (\alpha_k I - A^v)F^{\delta_k^\infty}(\alpha_k), \ v \in \Omega_s.$$

令 $k\to\infty, \alpha_k \downarrow 0$,并注意到(2.1)式,有

$$0 = f^\delta - e\eta^{\delta^\infty} + A^\delta g^{\delta^\infty} \leqslant$$

$$f^v - e\eta^{\delta^\infty} + A^v g^{\delta^\infty}, \ v \in \Omega_s.$$

即存在 $(\eta^{\delta^\infty}, g^{\delta^\infty})$ 满足(4.4)式. 若 $(\eta', g')$ 是(4.4)式的另一解,则由(4.2)式直接可得,$\eta' \leqslant \eta^{\delta^\infty}$,同理可得,$\eta^{\delta^\infty} \leqslant \eta'$,故 $\eta^{\delta^\infty} = \eta'$.

定理4同时也表明,存在一个控制策略($v^*(1)$,

$v^*(2),\cdots,v^*(k)) = \delta^\infty \in \Lambda$. 满足(4.3)式,也即对任意 $v \in \Omega_s, f^{v^*}+A^{v^*}g^{v^*} \leqslant f^v + A^v g^{v^*}$,故 $v^*$ 是最优策略,$v^*$ 能由(4.1)导出的一个策略迭代算法而被获得.在通常情况下,平均代价模型的最优控制策略并不唯一,因而性能势向量也不唯一,然而最优准则函数值是唯一的.

## 5 优化算法与算例(Optimality algorithm and numerical example)

本节中,我们给出基于平均代价模型最优性方程的策略优化算法,并通过一个实际算例说明该算法的优越性.为了便于比较,我们采用了文献[10]中的算例.该算例是一个很小规模的问题,但是本文算法完全可以适用于包含多个状态的问题.我们在文献[11]中应用该方法解决呼叫接入控制问题.该服务网络的结构与本文讨论的网络有许多相似之处,它含有 140 个状态.对于该问题,本算法仍然具有相当快的收敛速度.

考虑本文第二节所介绍的闭排队网络,我们这个算例考虑的是负荷相关的策略(即每个服务器的服务率取决于该服务器处的顾客总数,我们的目标是选择一个策略 $v = (\mu_{1,1},\cdots,\mu_{1,N},\cdots,\mu_{M,1},\cdots,\mu_{M,N})$,使性能指标函数 $J(v)$ 达到最小.

基于平均代价模型最优性方程的策略优化算法如下:

第一步 对于每一个状态 $n = 1,2,\cdots,K$,选择一个策略 $v^0(n)$,使得

$$\sum_{m=1}^{K} a_{nm}(v^0(n))f(n,v^0(n)) + f(n,v^0(n)) \leqslant \sum_{m=1}^{K} a_{nm}(\tilde{v}(n))f(n,\tilde{v}(n)) + f(n,\tilde{v}(n))$$

对所有 $\tilde{v}(n) \in v(n)$ 均成立,由此得到策略 $v^0$.

第二步 计算在策略 $v^0$ 下所对应的性能指标势向量 $g^{v^0}$ 及稳态概率向量 $\pi^{v^0}$,并令 $k = 0$.

第三步 对于每一个状态 $n = 1,2,\cdots,K$,选择一个策略 $v^{k+1}(n)$,使得

$$\sum_{m=1}^{K} a_{nm}(v^{k+1}(n))g^{v^k}(m) + f(n,v^{k+1}(n)) \leqslant \sum_{m=1}^{K} a_{nm}(\tilde{v}(n))g^{v^k}(m) + f(n,\tilde{v}(n))$$

对所有 $\tilde{v}(n) \in v(n)$ 均成立,由此得到策略 $v^{k+1}$.

第四步 计算在策略 $v^{k+1}$ 下所对应的性能指标势向量 $g^{v^{k+1}}$ 及稳态概率向量 $\pi^{v^{k+1}}$,并计算在策略 $v^{k+1}$ 下所对应的性能指标值 $J(v^{k+1})$.

第五步 若 $J(v^{k+1}) < J(v^k)$,则置 $k+1 = k$,返回步骤 3.否则,$v^k$ 即为最优策略.

表 1 两种方法的比较

Table 1 Comparing of two methods

| | $v^{(0)}$ | | | $v^*$ | | | $J(v^*)$ | Run-time |
|---|---|---|---|---|---|---|---|---|
| 直接梯度搜索结果 ($\varepsilon = 10^{-5}$) | 1.00 | 1.00 | 1.00 | 1.22494 | 0.67927 | 0.583042 | 0.99023 | 36″ |
| | 1.00 | 1.00 | 1.00 | 1.12494 | 0.67927 | 0.583042 | | |
| | 0.40 | 0.45 | 0.30 | 1.12512 | 0.67927 | 0.583248 | 0.99023 | 55″ |
| | 0.20 | 0.80 | 0.60 | 1.12541 | 0.67927 | 0.583042 | | |
| | 0.32 | 0.73 | 0.54 | 1.12495 | 0.67927 | 0.583042 | 0.99023 | 40″ |
| | 0.86 | 0.64 | 0.27 | 1.12504 | 0.67927 | 0.583042 | | |
| MDP 方法优化结果 ($\varepsilon = 10^{-7}$) | 1.00 | 1.00 | 1.00 | 1.12518 | 0.67927 | 0.583042 | 0.99023 | 2″ |
| | 1.00 | 1.00 | 1.00 | 1.12518 | 0.67927 | 0.583042 | | |
| | 0.40 | 0.45 | 0.30 | 1.12518 | 0.67927 | 0.582795 | 0.99023 | 4″ |
| | 0.20 | 0.80 | 0.60 | 1.12518 | 0.67927 | 0.582792 | | |
| | 0.32 | 0.73 | 0.54 | 1.12519 | 0.67927 | 0.583267 | 0.99023 | 2″ |
| | 0.86 | 0.64 | 0.27 | 1.12518 | 0.67927 | 0.582780 | | |

考虑文献[10]中所讨论的闭排队网络的例子,即 $M = 2, N = 3, q_{1,1} = q_{2,2} = 0.3, q_{1,2} = 0.7, 0.01 \leqslant \mu_{i,j} \leqslant 10$,性能指标函数

$$f(n,v(n)) = c(n,v(n)) + h(n,v(n)),$$

其中

$$c(n,v) = \begin{cases} \ln(1 + n_i/N)\mu_{i,j}, & n_i \neq 0, \\ 0, & n_i = 0. \end{cases}$$

及

$$h(n,v)=\begin{cases}\dfrac{\sqrt{n_i}}{2\mu_{i,j}N}, & n_i\neq 0,\\ 0, & n_i=0.\end{cases}$$

表1给出了应用梯度搜索方法[10]与应用MDP方法进行优化的结果与运行时间,从中可以看出两者的速度差别是很明显的.因为该方法是将对一个 $M*N$ 维向量 $v=(\mu_{1,1},\cdots,\mu_{1,N},\cdots,\mu_{M,1},\cdots,\mu_{M,N})$ 的整体寻优问题转化为分别对 $K$ 个($K$ 为状态数)$M$ 维向量的寻优.如果考虑的是状态相关的策略[11],则是将对一个 $M*K$ 维向量的整体寻优问题转化为分别对$K$个$M$维向量的寻优.当 $M,N$ 较大时,$K\gg N$.所以,对于状态相关的策略,当状态数 $K$ 较大时,该方法的优越性将非常明显(在这种情况下,基于梯度搜索方法对 $M*K$ 维向量进行寻优运算时间非常长,实际上已不可能).

需要说明的是,本文虽然是针对闭排队网络进行的策略优化,但是其适用范围仍然是很广泛的.首先,实际系统中许多问题本身就是闭排队网络的优化问题.第二,有些实际网络可以通过适当的建模转化为闭排队网络问题(如ATM通讯网络中的节点在饱和输入的情况下),我们在另一篇论文中讨论了将该方法应用于呼叫接入控制问题中的策略优化.该问题中的网络从本质上来说是一个开排队网络,但由于系统总带宽一定,因此它等价于本文中所讨论的闭排队网络[11].

## 6 结论(Conclusions)

本文主要在CQNS性能势基础上,以一种简明的方法直接导出了有关无限水平折扣代价和平均代价模型在平稳策略集上的两个最优性方程,证明了在紧致行动集上最优解的存在性,并给出了基于最优性方程的策略优化算法与实际算例,通过与直接梯度方法的比较,进一步说明了该算法的优越性.在对实际问题的优化中常常会碰到状态维数很高,甚至模型参数不是完全清楚的情况,这时需要考虑基于仿真的方法获得性能势的估计值,我们在文献[12]中进行了通过并行仿真的方法估计性能势的研究.我们将对这一问题进行进一步的研究,以建立一种新的并行优化算法.

## 参考文献(References)


[1] Song J S. Continuous-time Markov decision programming with non-uniformly bounded transfer rates [J]. Science in China (Series A), 1987, 17(12):1258-1267 (in Chinese)

[2] Raul Montes-de-Oca. The average cost optimality equation for Markov control processes on Borel spaces [J]. Systems & Control Letters, 1994, 22(2):351-357

[3] Cao X R, Chen H F. Perturbation realization, potentials, and sensitivity analysis of Markov processes [J]. IEEE Trans. Automat. Control, 1997,42(10):1382-1393

[4] Cao X R. The relations among potentials, perturbation analysis, and Markov decision processes [J]. Discrete Event Dynamic Systems: Theory and Applications, 1998,8(1):71-78

[5] Yin B Q, Zhou Y P, Yang X X, et al. Sensitivity formulas of performance in closed state-dependent queueing networks [J]. Control Theory and Applications, 1999,16(2):255-257 (in Chinese)

[6] Yin B Q, Zhou Y P, Xi H S, et al. Sensitivity formulas of performance in two-server cyclic queueing networks with phase-type distributed service times [J]. International Trans. in Operation Research, 1999,6(6):649-663

[7] Cao X R. Realization Probabilities: The Dynamics of Queueing Systems [M]. New York: Springer-Verlag, 1994

[8] Cinlar E. Introduction to Stochastic Processes [M]. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1975

[9] Lamond F, Puterman M L. Generalized inverses in discrete time Markov decision processes [J]. SIAM J Matrix Anal. Appl., 1989, 10(1):351-357

[10] Zhou Y P, Yin B Q, Xi H S, et al. Algorithms of decentralized optimization for a class of closed queueing network by using performance potentials [J]. Journal of China University of Science and Technology, 2000,30(2):151-157 (in Chinese)

[11] Zhou Y P, Xi H S, Yin B Q, et al. Continuous-time Markov decision processes to call admission control problem [J]. Control & Decision, 2001,16(suppl.):795-799 (in Chinese)

[12] Zou C C, Xi H S, Yin B Q, et al. Derivative estimates parallel simulation algorithms based on performance potentials theory [A]. Proceedings of IFAC 14th World Congress [C]. Beijing, 1999,49-54


## 本文作者简介


**周亚平** 1963年生.在中国科技大学控制理论与控制工程专业获学士、硕士、博士学位.现为中国科技大学管理科学系副教授,从事经济管理系统,排队网络性能灵敏度仿真估计及优化等方面的研究.

Email: zhouyp@ustc.edu.cn

**奚宏生** 见本刊2002年第2期第312页.

**殷保群** 见本刊2002年第2期第312页.

**孙德敏** 见本刊2002年第2期第312页.