文章编号: 1001-0920(2003)03-0267-05

Markov 控制过程在紧致行动集上的迭代优化算法

唐 昊, 奚宏生, 殷保群
(中国科学技术大学 自动化系, 安徽 合肥 230026)

摘 要: 研究一类连续时间 Markov 控制过程(CTMCP)在紧致行动集上关于平均代价性能准则的优化算法。根据CTMCP的性能势公式和平均代价最优性方程,导出了求解最优或次最优平稳控制策略的策略迭代算法和数值迭代算法,在无需假设迭代算子是sp-压缩的条件下,给出了这两种算法的收敛性证明。最后通过分析一个受控排队网络的例子说明了这种方法的优越性。

关键词: Markov 控制过程; 紧致行动集; 性能势; 策略迭代; 数值迭代
中图分类号: TP202 文献标识码: A

Iteration optimization algorithms for Markov control processes with compact action set

TANG Hao, XI Hong-sheng, YIN Bao-qun
(Department of Automation, University of Science and Technology of China, Hefei 230026, China)

Abstract: Optimization algorithms are studied for a class of continuous-time Markov control processes (CTMCPs) with infinite horizon average-cost criteria and compact action set. By using the formula of performance potentials and an average-cost optimality equation for CTMCPs, a policy iteration algorithm and a value iteration algorithm are derived, which can lead to an optimal or suboptimal stationary policy in a finite number of iterations. The convergence of these algorithms is established, without the assumption of the corresponding iteration operator being an sp-contraction. A numerical example of queuing networks shows advantages of the proposed value iteration method.

Key words: Markov control processes; Compact action set; Performance potentials; Policy iteration; Value iteration

1 引 言

Markov 控制过程(MCP)可分为离散时间 MCP(DTMCP)和连续时间 MCP(CTMCP)。由于它在通讯网络、生产调度和军事运筹等实际系统中具有重要应用,其性能优化问题已成为国内外控制界的研究热点。目前主要工作集中在具有有限行动集和可数行动集的 DTMCP 分析上[1~3]。对于一些实际系统,如受控排队网络,需要考虑连续时间和紧致行动集的情况。文献[4]将 Markov 性能势理论引入 Markov 过程的灵敏度分析,揭示了性能势与 MCP 的密切联系。在此基础上,性能势被运用到 MCP 的优化研究中,取得了满意的结果[5,6]。

本文根据CTMCP的性能势公式和平均代价最优性方程,研究其关于紧致行动集的优化问题,给出了一种策略迭代算法和数值迭代算法,在无需假设迭代算子是 sp-压缩的条件下,证明了这两种算法的收敛性。

收稿日期: 2002-01-15; 修回日期: 2002-04-08。
基金项目: 国家自然科学基金资助项目(69974037); 国家高性能计算基金资助项目(00208)。
作者简介: 唐昊(1972—),男,安徽庐江人,博士生,从事离散事件动态系统、控制理论及应用等研究; 奚宏生(1950—),男,上海人,教授,博士生导师,从事离散事件动态系统和现代控制理论等研究。

2 问题描述

考虑连续时间Markov过程$\{X(t), t \geq 0\}$, 具有有限状态空间 $\Phi = \{1, 2, \cdots, M\}$ 和紧致集 $D = D(1) \times D(2) \times \cdots \times D(M)$, 这里$D(i)$ 是状态i的容许行动集。一个平稳策略记为 $v = (v(1), \cdots, v(M))$, 且$v(i) = d_i \in D(i)$。令Ω_s是全体平稳策略集, 假设在任意策略$v \in \Omega_s$驱动下, $\{X(t), t \geq 0\}$ 是不可约正常返的, 其状态转移矩阵为 $P^v(t) = [p_{ij}(t, v(i))]$, 无穷小生产元为$A^v = [a_{ij}(v(i))]$。若$\{X(t), t \geq 0\}$ 的嵌入Markov链的转移矩阵记为 $P^v = [p_{ij}(v(i))]$, 则有

$$a_{ij}(v(i)) = \begin{cases} \lambda(i, v(i))[p_{ii}(v(i)) - 1], & i = j \\ \lambda(i, v(i)) p_{ij}(v(i)), & i \neq j \end{cases}$$

其中$\lambda(i, v(i))$ 是Markov过程在状态i的转移率。记过程在策略v下的性能函数为$f^v = (f(1, v(1)), \cdots, f(M, v(M)))^\tau$, 状态的稳态概率为 $\pi^v = (\pi(1, v(1)), \cdots, \pi(M, v(M)))$, 则

$$\pi^v e = 1, \quad A^v e = 0, \quad \pi^v A^v = 0 \tag{1}$$

其中$e = (1, 1, \cdots, 1)^\tau$是分量均为1的$M$维列向量。

假设 **1** 对任意$i, j \in \Phi, t \geq 0, p_{ij}(t, v(i))$ 是定义在$D(i)$ 上的连续函数。

假设 **2** 对任意$i \in \Phi, f(i, v(i))$ 是定义在$D(i)$ 上的连续实值函数。

假设 **3** 存在正常数λ, 满足

$$\sup_{i \in \Phi, v \in \Omega_s} \{\lambda(i, v(i))\} = \lambda < +\infty$$

记$X = (X(t), \Phi, D, P^v(t), f^v)$ 为约束在Ω_s上的CTMCP。X 的无穷水平平均代价期望值准则为

$$\eta = \lim_{T \to \infty} \frac{1}{T} E\left\{\int_0^T f(X(t), v(X(t)))\, dt\right\}, \quad v \in \Omega_s \tag{2}$$

由遍历性, 系统的稳态性能测度为

$$\eta = \sum_{i=1}^{M} \pi(i, v(i)) f(i, v(i)) = \pi^v f^v \tag{3}$$

在CTMCP问题中, 优化的目的是要选择一控制决策方案, 使系统稳态性能测度达到最小。

3 性能势和最优性方程

对于α-折扣代价问题, 本文定义折扣Poisson方程为

$$(\alpha I - A^v + \lambda e \pi^v) g_\alpha^v = f^v \tag{4}$$

这里g_α^v为一列向量。令 $\bar{P}^v = I + A^v / \lambda, \beta = \lambda/(\lambda + \alpha)$。显然$\bar{P}_v$是一随机矩阵, $0 < \beta < 1$, 且$(\alpha I - A^v + \lambda e \pi^v) = (\lambda + \alpha)(I - \beta \bar{P}^v + \beta e \pi^v)$。因$(\bar{P}^v - e\pi^v)$的全体特征值均在单位圆内[3], 故$(\alpha I - A^v + \lambda e \pi^v)$非奇异, 式(4) 存在唯一非零解

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

称$g_\alpha^v = (g_\alpha^v(1), \cdots, g_\alpha^v(M))^\tau$为$\alpha$-势向量, $g_\alpha^v(i)$ 为α-势。令$\alpha = 0$, 则有$g^v = (-A^v + \lambda e \pi^v)^{-1} f^v$, 即

$$(-A^v + \lambda e \pi^v) g^v = f^v \tag{5}$$

称式(5) 为CTMCP的平均代价Poisson方程, $g^v = (g^v(1), \cdots, g^v(M))^\tau$为性能势向量, $g^v(i)$ 为性能势。特别地, 当$\lambda(i, v(i)) \equiv 1$即$\lambda = 1$时, 有$g^v = (-A^v + e\pi^v)^{-1} f^v$, 这与文献[4] 中取$A = P - I$情况下的性能势是一致的。

对于约束在紧致行动集上的CTMCP优化问题, 已证明有下面的定理和推论:

定理 **1** $v^* \in \Omega_s$是使式(2) 达到最小的平稳策略的充分必要条件为: 对任意$v \in \Omega_s$, 有

$$f^{v^*} + A^{v^*} g^{v^*} \leq f^v + A^v g^{v^*} \tag{6}$$

其中$\leq$表示对应分量等于或小于关系成立。

式(5) 可写为$\lambda e \pi^v g^v = f^v + A^v g^v$, 两边同乘以$\pi^v$且根据式(1) 得$\lambda \pi^v g^v = \pi^v f^v = \eta$。则

$$e\eta^v = f^v + A^v g^v \tag{7}$$

故最优性定理可等价地表示成如下推论:

推论 **1** $v^* \in \Omega_s$是最优平稳控制策略的充分必要条件为: v^* 满足$e\eta^{v^*} = \min_{v \in \Omega_s}\{f^v + A^v g^{v^*}\}$或

$$0 = \min_{v \in \Omega_s}\{f^v + A^v g^{v^*} - e\eta^{v^*}\} \tag{8}$$

式(8) 称为CTMCP基于性能势的平均代价最优性方程。

定理 **2** 在假设1～3下, 存在对应一个最优平稳策略的数据对(η, g), 满足

$$0 = \min_{v \in \Omega_s}\{f^v + A^v g - e\eta\} \tag{9}$$

若(η', g')是满足式(9) 的任意一个解, 则$\eta' = \eta$。

4 迭代算法和收敛性证明

4.1 策略迭代算法

Step1: 令$k = 0, \epsilon > 0$, 选择初始策略v_k;

Step2: 利用式(1) 和式(5) 计算π^{v_k}和g^{v_k};

Step3: 选择v_{k+1}, 对每一状态$i \in \Phi$, 满足

$$v_{k+1}(i) = \arg \min_{d_i \in D(i)}\{f(i, d_i) + \sum_{j=1}^{M} a_{ij}(d_i) g^{v_k}(j)\} \tag{10}$$

Step4: 如果$\mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) < \epsilon$, 则算法停止, 否则令$k := k + 1$, 转Step2。

这里定义$\mathrm{sp}(h) = \max_i\{h(i)\} - \min_i\{h(i)\}$, 即$\mathrm{sp}(\cdot)$为一半范数[2]。由假设1～3可知, 对每一状态

$i \in \Phi, f(i,d_i) + \sum_{j=1}^{M} a_{ij}(d_i)\, g(j)$ 在紧致集 $D(i)$ 上连续, 因此存在 v_{k+1} 满足式(10), 即

$$f^{v_{k+1}} + A^{v_{k+1}} g^{v_k} \leq f^{v} + A^{v} g^{v_k}, \quad \forall v \in \Omega \tag{11}$$

如果 $f^{v_{k+1}} + A^{v_{k+1}} g^{v_k} = f^{v_k} + A^{v_k} g^{v_k}$, 则有

$$f^{v_k} + A^{v_k} g^{v_k} \leq f^{v} + A^{v} g^{v_k}, \quad \forall v \in \Omega$$

根据定理 1 可知 v_k 是平均代价最优策略。

不失一般性, 现假设 $f^{v_{k+1}} + A^{v_{k+1}} g^{v_k} \prec f^{v_k} + A^{v_k} g^{v_k}$, 其中 $\prec$ 表示至少存在一对对应分量小于关系成立, 而其他对应分量均相等。因为 $\pi(i, v(i)) > 0, \forall i \in \Phi, v \in \Omega$, 由(1)和(7)两式得到

$$\eta^{v_{k+1}} = \pi^{v_{k+1}} f^{v_{k+1}} = \pi^{v_{k+1}} (f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) < \pi^{v_{k+1}} (f^{v_k} + A^{v_k} g^{v_k}) = \pi^{v_{k+1}} e\eta^{k} = \eta^{v_k}$$

即 v_{k+1} 是改进策略。

定理 3　在假设 1 ～ 3 下, 若 $\{v_k, k = 0, 1, \cdots\}$ 是改进策略序列, 则有

1) $\lim_{k\to\infty} \mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) = 0$;

2) 存在一最优平稳策略 v^* 满足 $\lim_{k\to\infty} v_k = v^*$, 且 $\lim_{k\to\infty} \eta^{k+1} = \eta^*$。

证明　相应于改进策略序列 $\{v_k, k = 0, 1, \cdots\}$, $\{\eta^{v_k}\}$ 是关于 k 的单调递减序列, 且有下界 η^*。根据连续性条件, 存在 $\delta = (\delta_1, \delta_2, \cdots, \delta_M) \in D$, 满足 $\lim_{k\to\infty} \eta^{v_k} = \eta^{\delta}$, 且 $\lim_{k\to\infty} v_k = \delta$。因此

$$\lim_{k\to\infty} (f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) = f^{\delta} + A^{\delta} g^{\delta} = e\eta^{\delta}$$

即对任意 $i \in \Phi$, $\lim_{k\to\infty} (f^{v_{k+1}} + A^{v_{k+1}} g^{v_k})(i) = \eta^{\delta}$。因为 $\mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) = \max_{i\in\Phi} \{(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k})(i)\} - \min_{i\in\Phi} \{(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k})(i)\}$, 两边同取极限得

$$\lim_{k\to\infty} \mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) = \eta^{\delta} - \eta^{\delta} = 0$$

于是对任意给定常数 $\epsilon > 0$, 存在自然数 K, 当 $k > K$ 时, $\mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) < \epsilon$ 成立。

现令 $k > K$, 则由式(1)得

$$\eta^{k+1} = \pi^{v_{k+1}} f^{v_{k+1}} = \pi^{v_{k+1}} (f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) \leq \max_{i\in\Phi} \{(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k})(i)\}$$

由式(11)得

$$\eta^{*} = \pi^{v^*} f^{v^*} = \pi^{v^*} (f^{v^*} + A^{v^*} g^{v_k}) \geq \pi^{v^*} (f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) \geq \min_{i\in\Phi} \{(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k})(i)\}$$

上面两式相减得

$$\eta_{k+1} - \eta^{*} \leq \mathrm{sp}(f^{v_{k+1}} + A^{v_{k+1}} g^{v_k}) < \epsilon$$

则 v_{k+1} 是 ϵ-最优策略。根据 ϵ 的任意性, 有 $\lim_{k\to\infty} \eta^{k+1} = \eta^{*}$。由极限唯一性可得 $\eta^{*} = \eta^{\delta}$, 即策略序列 $\{v_k, k = 1, 2, \cdots\}$ 必然收敛到一最优平稳策略 $v^{*} = \delta$, 所以有 $\lim_{k\to\infty} v_k = v^{*}$。□

4.2　数值迭代算法

策略迭代方法需要反复求解改进策略的稳态概率和性能势向量, 在状态空间很大而存在“维数灾”的系统中, 这将耗费大量计算时间。本文给出一种数值迭代寻优算法, 并且在不需假设迭代算子是 sp-压缩的情况下, 证明其收敛性。首先易证明

$$\arg\min_{d\in D} \{f^{d} + A^{d} g - e\eta\} = \arg\min_{d\in D} \{\tilde{f}^{d} + \tilde{P}^{d} g\}$$

这里 $\tilde{f}^{d} = f^{d}/\lambda$。因此可构造下列数值迭代算法:

Step1: 令 $k = 0, \epsilon > 0$, 选择一个初始策略 v_k, 计算 g^{v_k}, 且令 $h^{k} = g^{v_k}$;

Step2: 选择 v_{k+1}, 对每一状态 $i \in \Phi$, 满足

$$v_{k+1}(i) = \arg\min_{d_i \in D(i)} \{\tilde{f}(i, d_i) + \sum_{j=1}^{M} \tilde{p}_{ij}(d_i) h^{k}(j)\} \tag{12}$$

Step3: 计算

$$h^{k+1} = \tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^{k} \tag{13}$$

Step4: 如果 $\mathrm{sp}(h^{k+1} - h^{k}) < \epsilon/\lambda$, 则记 $v_{\epsilon} = v_{k+1}$, 算法停止, 否则令 $k := k + 1$, 转 Step2。

定理 4　在假设 1 ～ 3 下, 上面的数值迭代算法在有限步内停止, 且得到一个 ϵ-最优策略 v_{ϵ}。

证明　由假设 1 ～ 3 知, 对任意 k, h^{k} 满足

$$h^{k} = \tilde{f}^{v_k} + \tilde{P}^{v_k} h^{k-1} \leq \tilde{f}^{v^*} + \tilde{P}^{v^*} h^{k-1}$$

根据递推关系有

$$h^{k} \leq \sum_{n=0}^{k-1} P_n \tilde{f}^{v^*} + P_k h^{0} \tag{14}$$

这里定义 $P_n = [\tilde{P}^{v^*}]^{n}$, $n \geq 1$, 且 $[\tilde{P}^{v^*}]^{0} = I$, 可见 P^{n} 也是一随机矩阵, 有 $P_n e = e$。由式(7)得

$$e\eta^{*} = f^{v^*} + A^{v^*} g^{v^*} = f^{v^*} + \lambda(\tilde{P}^{v^*} - I) g^{v^*}$$

两边同除以 λ, 整理得

$$\tilde{f}^{v^*} = -(\tilde{P}^{v^*} - I) g^{v^*} + e\eta^{*}/\lambda$$

代入式(14), 则有

$$h^{k} \leq \sum_{n=0}^{k-1} P_n(-\tilde{P}^{v^*} g^{v^*} + g^{v^*} + e\eta^{*}/\lambda) + P_k h^{0} \leq ke\eta^{*}/\lambda + g^{v^*} + \max_{i\in\Phi} \{(h^{0} - g^{v^*})(i)\} e \tag{15}$$

对应算法的迭代过程, 定义一确定性 Markov 策略 $\Pi = (v_1, v_2, \cdots)$, 根据式(13)通过迭代得

$$h^{k} = \tilde{f}^{v_k} + \tilde{P}^{v_k} h^{k-1} = \tilde{f}^{v_k} + \tilde{P}^{v_k} (\tilde{f}^{v_{k-1}} + \tilde{P}^{v_{k-1}} h^{k-2}) = \sum_{n=1}^{k} P_{\Pi}^{k-n} \tilde{f}^{v_n} + P_{\Pi}^{k} h^{0} \tag{16}$$

这里定义 $P_{\Pi}^{0}=I, P_{\Pi}^{n}=\tilde{P}^{v_k}\tilde{P}^{v_{k-1}}\cdots\tilde{P}^{v_{k-n+1}}, n\geq 1$。可见 P_{Π}^{n} 也是一随机矩阵。根据式(8) 有

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

$$\min_{v\in\Omega_s}\{\tilde{f}^v+\tilde{P}^v g^{v^*}-g^{v^*}-e\eta^*/\lambda\}$$

因此对任意 n, 有

$$\tilde{f}^{v_n}+\tilde{P}^{v_n}g^{v^*}-g^{v^*}-e\eta^*/\lambda\geq 0$$

即 $$\tilde{f}^{v_n}\geq g^{v^*}-\tilde{P}^{v_n}g^{v^*}+e\eta^*/\lambda$$

代入式(16) 得

$$h^k\geq\sum_{i=1}^{k}P_{\Pi}^{k-n}(g^{v^*}-\tilde{P}^{v_n}g^{v^*}+e\eta^*/\lambda)+P_{\Pi}^{k}h^0\geq$$

$$ke\eta^*/\lambda+g^{v^*}+\min_{i\in\Phi}\{(h^0-g^{v^*})(i)\}e \qquad (17)$$

由式(15) 和式(17) 得

$$\min_{i\in\Phi}\{(h^0-g^{v^*})(i)\}e\leq$$

$$h^k-ke\eta^*/\lambda-g^{v^*}\leq$$

$$\max_{i\in\Phi}\{(h^0-g^{v^*})(i)\}e$$

对该不等式同除以 k, 且取 $k\to\infty$ 时的极限, 得 $\lim_{k\to\infty}(h^k/k-e\eta^*/\lambda)=0$。记 $h^k/k-e\eta^*/\lambda=\delta$, 这里 δ 为无穷小量, 即 $\lim_{k\to\infty}\delta=0$。则 $h^k=k(e\eta^*/\lambda+\delta)$, 所以有

$$\lim_{k\to\infty}(h^{k+1}-h^k)=e\eta^*/\lambda+\lim_{k\to\infty}\delta=e\eta^*/\lambda$$

可见 $\lim_{k\to\infty}\mathrm{sp}(h^{k+1}-h^k)=0$, 所以对任意给定的常数 ϵ, 存在 K, 当 $k>K$ 时, 恒有 $\mathrm{sp}(h^{k+1}-h^k)<\epsilon/\lambda$ 成立, 即算法在有限步内保证停止。

现假设 $k>K$。注意到对任意 $v\in\Omega$, $\eta^v=\pi^v f^v=\lambda\pi^v\tilde{f}^v$, $\pi^v\tilde{P}^v=\pi^v$, 故有

$$\eta^{v_\epsilon}=\eta^{v_{k+1}}=$$

$$\lambda\pi^{v_{k+1}}(\tilde{f}^{v_{k+1}}+\tilde{P}^{v_{k+1}}h^k-h^k)=$$

$$\lambda\pi^{v_{k+1}}(h^{k+1}-h^k)\leq$$

$$\lambda\max_{i\in\Phi}\{(h^{k+1}-h^k)(i)\} \qquad (18)$$

由式(12) 得 $\tilde{f}^{v^*}+\tilde{P}^{v^*}h^k\geq\tilde{f}^{v_{k+1}}+\tilde{P}^{v_{k+1}}h^k$, 则

$$\eta^{v^*}=\pi^{v^*}f^{v^*}=$$

$$\lambda\pi^{v^*}(\tilde{f}^{v^*}+\tilde{P}^{v^*}h^k-h^k)\geq$$

$$\lambda\pi^{v^*}(\tilde{f}^{v_{k+1}}+\tilde{P}^{v_{k+1}}h^k-h^k)=$$

$$\lambda\pi^{v^*}(h^{k+1}-h^k)\geq$$

$$\lambda\min_{i\in\Phi}\{(h^{k+1}-h^k)(i)\} \qquad (19)$$

(18) 和(19) 两式相减得

$$\eta_\epsilon-\eta^*\leq\lambda\mathrm{sp}(h^{k+1}-h^k)<\epsilon$$

即 v_ϵ 是 ϵ–最优平稳策略。□

5 数值例子

考虑一个具有 2 个单类服务节点和 3 个顾客的受控闭排队网络(CCQN)[6] 系统的可能状态数为 $M=C_4^3=4$。记状态为向量 $n=(n_1,n_2)$, 其中 n_i 表示系统在状态 n 时节点 i 的顾客数。状态空间为 $\Phi=\{(3,0),(2,1),(1,2),(0,3)\}$, 用该集合中元素的序号表示状态。假设路径转移概率满足 $q_{1,1}=q_{2,2}=0.3$, $q_{1,2}=q_{2,1}=0.7$, 服务率 $\mu_{i,n}$ 满足 $0.01\leq\mu_{i,n}\leq 10$, $n_i\neq 0;\ \mu_{i,n}=0, n_i=0$。系统在状态 n 的性能函数为

$$f(n,v(n))=\sum_{i=1}^{M}[c_i(n,v(n))+h_i(n,v(n))]$$

式中

$$c_i(n,v(n))=\begin{cases}0, & n_i=0\\ \ln(1+n_i/N)\mu_{i,n}, & \text{otherwise}\end{cases}$$

$$h_i(n,v(n))=\begin{cases}0, & n_i=0\\ \dfrac{n_i}{2\mu_{i,n}}, & \text{otherwise}\end{cases}$$

对这样一个 CTMCP 问题, 显然假设 1～3 都满足。因为在计算中始终有 $\mu_{1,1}=\mu_{2,4}$, $\mu_{1,2}=\mu_{2,3}$, $\mu_{1,3}=\mu_{2,2}$, 且 $\mu_{2,1}=\mu_{1,4}=0$, 所以为简单起见, 仅记策略为 $v=(\mu_{1,1},\mu_{1,2},\mu_{1,3})$。

该优化问题可用基于梯度的直接寻优方法求解[6], 取初始策略 $v_0=(1,1,1)$, 解得最优策略为 $v^*=(0.583\,046\,86, 0.679\,274\,85, 1.125\,184\,58)$, 最优代价为 $\eta^*=0.990\,229\,62$, 运行时间 t_s 为 22.5 s。表 1 是采用文中提供的数值迭代算法, 在同样的初始策略 v_0 和不同的 ϵ 下的计算结果。

从表 1 中的数据可看出, 数值迭代算法具有较高的精度, 得到的策略是 ϵ–最优策略, 并且运算速度很快。对于状态维数很大的问题, 相对于梯度直接寻优方法, 其速度优势将更为明显。因为从(12) 和(13) 两式可看到, 数值迭代算法只是对每个分量分别进行低维变量寻优和简单的迭代计算, 而直接梯

表 1 采用数值迭代算法的结果

ϵ	v_ϵ	η^{v_ϵ}	t_s
0.01	(0.599 745 05, 0.679 274 85, 0.991 764 22)	0.991 764 22	0.11
0.001	(0.587 072 52, 0.679 274 85, 1.077 514 53)	0.990 342 71	2.28
0.000 1	(0.583 475 59, 0.679 274 85, 1.119 513 99)	0.990 231 09	7.30

度算法则需要同时对M维分量进行多变量寻优和性能势的求解计算。

6 结　论

本文在 Markov 性能势的基础上,提供了一种收敛的策略迭代算法和数值迭代算法,能保证得到ϵ-最优平稳控制策略,并且给出的假设条件在实际系统中容易满足。尤其是数值迭代算法,运算简单方便,相对于常规的梯度算法具有较大的优越性。另外,由于性能势能够根据单个样本轨道来估计[4,5],而且相应的优化算法易于并行计算,因此它为大规模实际 CTMCP 的性能优化提供了一条新途径。

参考文献(References):

[1] Arapostathis A, Borkar V S, Fernandez-Gaucher, et al. Discrete-time controlled Markov processes with average cost criterion: A survey[J]. *SIAM J Control Optimization*, 1993, 31(2): 282-344.

[2] Puterman M L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*[M]. New York: Wiley, 1994.

[3] Bertsekas D P, Tsitsiklis J N. *Neuro-dynamic Programming*[M]. Belmont: Athena Scientific, 1996.

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

[5] Cao X R. Single sample path-based optimization of Markov chains[J]. *J of Optimization Theory and Applications*, 1999, 100(3): 527-548.

[6] 周亚平,殷保群,奚宏生,等. 一类闭排队网络基于性能势的优化算法[J]. 中国科技大学学报, 2000, 30(2): 151-157.
(Zhou Y P, Yin B Q, Xi H S, et al. Algorithms of decentralized optimization for a class of closed queuing network by using performance potentials[J]. *J of University of Science and Technology of China*, 2000, 30(2): 151-157.)

(上接第 262 页)

[36] 王凌,李文峰,郑大钟. 基于一类混合策略的模型参数估计和控制器参数整定研究[J]. 控制与决策, 2001, 16(5): 530-534.
(Wang L, Li W F, Zheng D Z. Estimating model-parameter and tuning controller-parameter by a class of hybrid strategy[J]. *Control and Decision*, 2001, 16(5): 530-534.)

[37] Khoo L P, Chen C H. Integration of response surface methodology with genetic algorithms[J]. *Int J Advanced Manufacturing Technology*, 2001, 18(7): 483-489.

[38] Yucesan E, Luo Y C, Chen C H, et al. Distributed web-based simulation experiments for optimization[J]. *Simulation Practice and Theory*, 2001, 9(1/2): 73-90.

[39] Chen H C, Chen C H, Yucesan E. Computing efforts allocation for ordinal optimization and discrete event simulation[J]. *IEEE Trans Automatic Control*, 2000, 45(5): 960-964.

[40] Havlik I, Schorcht R. MOPSI/CM: A program system for simulation and optimization of dynamic systems with combined models[J]. *Systems Analysis Modelling Simulation*, 1990, 7(8): 637-647.

[41] Bengu G, Haddock J. An implementation of a simulation optimization system[J]. *Int J Computer Simulation*, 1994, 4(3): 305-325.

[42] Behe R, Bills R, Brachat P, et al. Simulation and optimization software for axisymmetrical radiating structures[J]. *Annals of Telecommunications*, 1994, 49(9/10): 575-588.

[43] Dote Y, Ovaska S J. Industrial applications of soft computing: A review[J]. *Proc of IEEE*, 2001, 89(9): 1243-1265.

[44] 王凌,郑大钟. 混合优化策略统一结构的探讨[J]. 控制与决策, 2002, 17(1): 33-36.
(Wang L, Zheng D Z. Study on unified framework of hybrid optimization strategies[J]. *Control and Decision*, 2002, 17(1): 33-36.)