第 29 卷 第 2 期 自 动 化 学 报 Vol. 29, No. 2
2003 年 3 月 ACTA AUTOMATICA SINICA Mar., 2003

Optimal Policies for a Continuous Time MCP with Compact Action Set[1)]

XI Hong-Sheng TANG Hao YIN Bao-Qun
(*Department of Automation, University of Science and Technology of China, Hefei* 230026)
(E-mail: xihs@ustc.edu.cn)

Abstract In this paper, we study optimal policies for a class of continuous-time Markov control processes (CTMCPs) with infinite horizon average-cost criteria. Using the basic properties of infinitesimal generators and performance potentials, we give directly the optimality equation and establish the existence of solutions to this equation for the average-cost model on a compact action set. A fast value iteration algorithm, which leads to an ε-optimal stationary policy, is proposed and the convergence of this algorithm is studied. Finally, we provide one numerical example to show applications of the proposed method.

Key words Performance potentials, average-cost criteria, compact action set, value iteration

1 Introduction

Markov control processes (MCPs) are a class of Markov chains driven by the control policy, and evolve according to the rule of state transition and action selection, which can be classified in two groups, i. e., discrete-time MCP (DTMCP) and continuous-time MCP (CTMCP). In the literature[1～3], under some strong assumptions, the authors discussed the average-cost optimality equation for DTMCPs and the existence of the solutions, and presented some convergent iteration algorithms. Recently, Cao introduced the theory of Markov performance potentials to the performance analysis of MCPs[4,5], which can relax the restrictions. The results have been extended to the study of queuing networks[6～8]. In this paper, based on work of [4] and [5], we deal with the optimization problems of a class of CTMCPs with finite state space and compact action set under infinite horizon average-cost criteria. With some weak assumptions, we deduce an optimality equation and an existence theorem of the solutions using the basic properties of infinitesimal generators and performance potentials for CTMCPs. In addition, we propose a value iteration algorithm that can lead to an ε-optimal policy. All the results provide a foundation for further study of performance optimization of many real systems such as queuing networks.

2 CTMCP

Consider a continuous-time Markov process $\{X(t), t\geqslant 0\}$ with finite state space $\Phi=\{1,2,\cdots,M\}$ and compact action set $D=D(1)\times D(2)\times\cdots\times D(M)$, where $D(i)$ is the feasible action set at state i. A stationary policy is denoted as $v=(v(1),\cdots,v(M))$, and $v(i)=d_i\in D(i)$, $i=1,2,\cdots.M$. Let Ω_s be the set of all such possible stationary policies. Assume that under any policy $v\in\Omega_s$, $\{X(t), t\geqslant 0\}$ is irreducible and positively recurrent, and $P^v(t)=[p_{ij}(t, v(i))]$ is the transition matrix . Furthermore, a general form of the infinitesimal generator is $A^v=[a_{ij}(v(i))]$. Let $\{X_n, n=0,1,2,\cdots\}$ be one of the embedded Markov chain, and $P^v=[p_{ij}(v(i))]$ be its transition matrix such that

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

1) Supported by National Natural Science Foundation of P. R. China (69974037) and National High Performance Computing Foundation of P. R. China(00208)
Received October 8, 2001; in revised form May 22, 2002
收稿日期 2001-10-08; 收修改稿日期 2002-05-22

where $\lambda(i,v(i))>0$, denoting the state transition rate of the process at state i under policy v. Equation (1) implies $A^v=\mathrm{diag}(\lambda(1,v(1)),\cdots,\lambda(M,v(M)))\cdot(P^v-I)$. Let the stationary distribution of the process under v be denoted by $\pi^v=(\pi(1,v(1)),\cdots,\pi(M,v(M)))$, we have

$$\pi^v e=1, A^v e=0, \pi^v A^v=0 \tag{2}$$

Here, $e=(1,1\cdots,1)^\tau$ is a M-dimensional column vector with all components equal to 1. Suppose f is a real performance function depending on policy v, and denote $f^v=(f(1,v(1)),\cdots,f(M,v(M)))^\tau$. We make the following assumptions.

Assumption 1. For any $i,j\in\Phi$, $p_{ij}(t,v(i))$ is a continuous function on $D(i)$.

Assumption 2. For any $i\in\Phi$, $f(i,v(i))$ is a continuous function on $D(i)$.

Assumption 3. There exists a constant λ satisfying $\sup\limits_{i\in\Phi,v\in\Omega_s}\{\lambda(i,v(i))\}=\lambda<+\infty$.

We denote $X=(X(t),\Phi,D,P^v(t),f^v)$ to be a CTMCP constrained on the stationary policy set Ω_s. The infinite horizon discounted-cost expectation criteria of X is

$$\eta_\alpha^v(i)=E\left\{\int_0^{+\infty}e^{-\alpha t}f(X(t),v(X(t)))\mathrm{d}t\,\middle|\,X(0)=i\right\}, v\in\Omega_s, i\in\Phi \tag{3}$$

where $\alpha>0$ is a discount factor; the average-cost expectation criteria is

$$\eta^v=\lim_{T\to\infty}\frac{1}{T}E\left\{\int_0^T f(X(t),v(X(t)))\mathrm{d}t\right\}, v\in\Omega_s \tag{4}$$

Since X is ergodic, $\eta^v=\sum\limits_{i=1}^{M}\pi(i,v(i))f(i,v(i))=\pi^v f^v$. In CTMCPs, the objective of optimization is to select a policy so that the right-hand of Equation (3) or (4) attains the minimum.

3 Performance potentials

For any $v\in\Omega_s,\alpha>0$, let $R_\alpha^v=\int_0^{+\infty}e^{-\alpha t}P^v(t)\mathrm{d}t$. It is easy to prove that $(\alpha I-A^v)R_\alpha^v=R_\alpha^v(\alpha I-A^v)=I$, this is, $\int_0^{+\infty}e^{-\alpha t}P^v(t)\mathrm{d}t=(\alpha I-A^v)^{-1}$. From Equation (3),

$$\eta_\alpha^v(i)=E\left\{\int_0^{+\infty}e^{-\alpha t}f(X(t),v(X(t)))\mathrm{d}t\,\middle|\,X(0)=i\right\}=$$
$$\int_0^{+\infty}e^{-\alpha t}\sum_{j=1}^{M}p_{ij}(t,v(i))f(j,v(j))\mathrm{d}t.$$

Letting $\eta_\alpha^v=(\eta_\alpha^v(1),\eta_\alpha^v(2),\cdots,\eta_\alpha^v(M))^\tau$, we have

$$\eta_\alpha^v=\int_0^{+\infty}e^{-\alpha t}P^v(t)f^v\mathrm{d}t=(\alpha I-A^v)^{-1}f^v \tag{5}$$

We define the discount Poisson equation of a CTMCP as

$$(\alpha I-A^v+\lambda e\pi^v)g_\alpha^v=f^v \tag{6}$$

Here, g_α^v is a column vector. Let $\widetilde{P}^v=A^v/\lambda+I$, and $\beta=\lambda/(\lambda+\alpha)$. Then $\widetilde{P}^v$ is a stochastic matrix, and $0<\beta<1$. Since $(I-\beta\widetilde{P}^v+\beta e\pi^v)$ is nonsingular[4], $(\alpha I-A^v+\lambda e\pi^v)$ is also nonsingular, and there exists uniquely one nonzero solution to Equation (6), i. e., $g_\alpha^v=(\alpha I-A^v+\lambda e\pi^v)^{-1}f^v$. If $\alpha=0$, then

$$g^v=(-A^v+\lambda e\pi^v)^{-1}f^v \tag{7}$$

We call Equation (7) an average-cost Poisson equation of a CTMCP. $g^v=(g^v(1),\cdots,g^v(M))^\tau$ is a performance potential vector, and $g^v(i)$ is a performance potential. Especially, when $\lambda(i,v(i))=1$, $\forall i\in\Phi,v\in\Omega_s$, we have $g^v=(-A^v+e\pi^v)^{-1}f^v$, which is equal to the definition of performance potential in [4] and [9]. From Equation (2), it is easy to prove the following lemma, and we omit the details.

Lemma 1. Under Assumption 3, for any $v\in\Omega_s$ and $\alpha>0$, we have

a) $\pi^v(\alpha I-A^v+\lambda e\pi^v)^{-1}=\pi^v/(\lambda+\alpha)$,

b) $(\alpha I-A^v+\lambda e\pi^v)^{-1}e=e/(\lambda+\alpha)$,

c) $(\alpha I-A^v)^{-1}e=e/\alpha$.

Using a) and c), we have $(\alpha I-A^v)^{-1}=(\alpha I-A^v+\lambda e\pi^v)^{-1}+\lambda e\pi^v/[\alpha(\lambda+\alpha)]$. Right-multiplying both sides of the equation by f^v, and combining with Equations (5) and (6) leads to

$$\eta_\alpha^v = g_\alpha^v + \lambda e\eta^v/[\alpha(\lambda+\alpha)] \tag{8}$$

4 Average-cost optimality equation

For α-discounted problems, we summarize Puterman's results as the following theorem[10].

Theorem 1. Under Assumptions 1, 2 and 3, for any $\alpha>0$, there exist uniquely a stationary policy $v^* \in \Omega_s$ and a corresponding bounded real vector $\eta_\alpha^{v^*}$ satisfying

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

If let $(v(1),\cdots,v(M))=d\in D$, $(v^*(1),\cdots,v^*(M))=\delta^\infty\in D$, and

$$\delta = \arg\min_{d\in D}\{f^d - (\alpha I - A^d)\eta_\alpha^{\delta^\infty}\} \tag{9}$$

then, for any $v\in\Omega_s$, we have $0=f^\delta-(\alpha I-A^\delta)\eta_\alpha^{\delta^\infty}\leqslant f^v-(\alpha I-A^v)\eta_\alpha^{\delta^\infty}$.

In this section, we mainly discuss the average-cost optimality equation and the existence of the solutions on a compact action set. By Equations (2), (4) and (7), it is easy to prove the following lemma and theorem.

Lemma 2. Under Assumption 3, for any $v',v\in\Omega_s$, we have

$$\eta^{v'} - \eta^v = \pi^v[(f^{v'} + A^{v'}g^{v'}) - (f^v + A^v g^{v'})].$$

Theorem 2(Optimality theorem). $v^*\in\Omega_s$ is average-cost optimal if and only if

$$f^{v^*} + A^{v^*}g^{v^*} \leqslant f^v + A^v g^{v^*}, \ \forall v\in\Omega_s.$$

From Equation (7), we have $(-A^{v^*}+\lambda e\pi^{v^*})g^{v^*}=f^{v^*}$, and $\lambda e\pi^{v^*}g^{v^*}=f^{v^*}+A^{v^*}g^{v^*}$. Using Equation (2), we have $\lambda\pi^{v^*}g^{v^*}=\pi^{v^*}f^{v^*}$. Therefore $e\eta^{v^*}=f^{v^*}+A^{v^*}g^{v^*}$. We have the following corollary.

Corollary 1. $v^*\in\Omega_s$ is average-cost optimal if and only if

$$e\eta^{v^*} = \min_{v\in\Omega_s}\{f^v + A^v g^{v^*}\} \text{ or } 0 = \min_{v\in\Omega_s}\{f^v + A^v g^{v^*} - e\eta^{v^*}\} \tag{10}$$

Equation (10) is called an average-cost optimality equation based on performance potentials of a CTMCP.

Theorem 3. Under Assumptions 1, 2 and 3, there exists a two-tuple (η, g) corresponding to one optimal policy satisfying

$$0 = \min_{v\in\Omega_s}\{f^v + A^v g - e\eta\} \tag{11}$$

In addition, if (η', g') corresponding to one policy is another solution to Equation (11), then $\eta'=\eta$.

Proof. Choose a discount factor sequence $\alpha_k \downarrow 0$. From Theorem 1, for each α_k, there exist an action $\delta_k\in D$ and an action $\delta_k^\infty\in D$, corresponding to the unique optimal policy v_k^*, satisfying Equation (9). Since D is compact, we can choose a subsequence $\{\delta_{k_j}\}$ of $\{\delta_k\}$ which converges to an action $\delta\in D$. Then the corresponding subsequence $\{\delta_{k_j}^\infty\}$ of $\{\delta_k^\infty\}$ also converges to an action $\delta^\infty\in D$. Furthermore δ^∞ and δ satisfy Equation (9). To simplify subsequence notation, denote $\{k_j\}$ by $\{k\}$. Then

$$0 = f^{\delta_k} - (\alpha_k I - A^{\delta_k})\eta_{\alpha_k}^{\delta_k^\infty} \leqslant f^v - (\alpha_k I - A^v)\eta_{\alpha_k}^{\delta_k^\infty}, \ v\in\Omega_s \tag{12}$$

From Assumption 1, it is easy to verify that, for any $i\in\Phi$, $a_{ij}(v(i))$ and $\pi(i,v(i))$ are continuous functions defined on $D(i)$. So, $\lim_{k\to\infty}\eta^{\delta_k^\infty}=\lim_{k\to\infty}\pi^{\delta_k^\infty}f^{\delta_k^\infty}=\pi^{\delta^\infty}f^{\delta^\infty}=\eta^{\delta^\infty}$, and

$$\lim_{k\to\infty} g_{\alpha_k}^{\delta_k^\infty} = \lim_{k\to\infty}(\alpha_k I - A^{\delta_k^\infty} + \lambda e\pi^{\delta_k^\infty})^{-1}f^{\delta_k^\infty} = (-A^{\delta^\infty} + \lambda e\pi^{\delta^\infty})^{-1}f^{\delta^\infty} = g^{\delta^\infty}.$$

Substituting Equation (8) into Equation (12), and using Lemma 1 (c), we get

$$0 = f^{\delta_k} - (\alpha_k I - A^{\delta_k})g_{\alpha_k}^{\delta_k^\infty} - \frac{\lambda}{\lambda + \alpha_k} e\eta^{\delta_k^\infty} \leqslant f^v - (\alpha_k I - A^v)g_{\alpha_k}^{\delta_k^\infty} - \frac{\lambda}{\lambda + \alpha_k} e\eta^{\delta_k^\infty}.$$

Letting $k \to \infty$ $(\alpha_k \to 0)$, we obtain $0 = f^\delta + A^\delta g^{\delta^\infty} - e\eta^{\delta^\infty} \leqslant f^v + A^v g^{\delta^\infty} - e\eta^{\delta^\infty}$. That is, $(\eta^{\delta^\infty}, g^{\delta^\infty})$ satisfies Equation (11). If (η', g') is another solution to Equation (11), by Equation (10) we obtain $\eta' \leqslant \eta^{\delta^\infty}$ and $\eta^{\delta^\infty} \leqslant \eta'$. Therefore $\eta^{\delta^\infty} = \eta'$. □

5 Value iteration algorithm

Letting $\tilde{f}^d = f^d/\lambda$, we have the following value iteration algorithm.

Step1. Let $k=0, \varepsilon>0$; select an arbitrary M-dimensional vector h^0.

Step2. Choose a policy v_{k+1} such that, for each $i \in \Phi$,

$$v_{k+1}(i) \in \arg\min_{d_i \in D(i)} \left\{ \tilde{f}(i, d_i) + \sum_{j=1}^{M} \tilde{p}_{ij}(d_i) h^k(j) \right\} \tag{13}$$

Step3. Let $h^{k+1} = \tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^k$.

Step4. If $sp(h^{k+1} - h^k) < \varepsilon/\lambda$, let $v_\varepsilon = v_{k+1}$ and exit; otherwise, let $k := k+1$ and go to Step2.

Here, $sp(h^{k+1} - h^k) = (h^{k+1} - h^k)(i^*) - (h^{k+1} - h^k)(i_*)$, and $i^* = \arg\max_{i\in\Phi}\{(h^{k+1} - h^k)(i)\}$, $i_* = \arg\min_{i\in\Phi}\{(h^{k+1} - h^k)(i)\}$.

Assumption 4. Suppose $\mu = \inf_{i,s\in\Phi; d_i \in D(i); d_s \in D(s)} \sum_{j=1}^{M} \min[\tilde{p}_{ij}(d_i), \tilde{p}_{sj}(d_s)] > 0$.

Theorem 4 (Convergence theorem). Under Assumptions 1~4, the above described algorithm stops in a finite number of iterations, and leads to an ε-optimal policy v_ε.

Proof. For any k, $\tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^k \leqslant \tilde{f}^{v_k} + \tilde{P}^{v_k} h^k$, $\tilde{f}^{v_k} + \tilde{P}^{v_k} h^{k-1} \leqslant \tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^{k-1}$. Therefore, it is easy to prove that

$$sp(h^{k+1} - h^k) = (h^{k+1} - h^k)(i^*) - (h^{k+1} - h^k)(i_*) \leqslant$$
$$\{\tilde{P}^{v_k}(h^k - h^{k-1})\}(i^*) - \{\tilde{P}^{v_{k+1}}(h^k - h^{k-1})\}(i_*) \leqslant (1-\mu) \cdot sp(h^k - h^{k-1}).$$

Letting $\gamma = 1 - \mu$, by deduction we get $sp(h^{k+1} - h^k) \leqslant \gamma^k \cdot sp(h^1 - h^0)$. Since $\mu > 0$, we have $0 \leqslant \gamma < 1$. Thus for any given constant $\varepsilon > 0$, there exists an integer K such that $sp(h^{k+1} - h^k) < \varepsilon/\lambda$ holds as $k > K$.

Now, let $k > K$. For any $v \in \Omega_s$, $\eta^v = \pi^v f^v = \lambda \pi^v \tilde{f}^v$, and $\pi^v \tilde{P}^v = \pi^v$. Then

$$\eta^{v_\varepsilon} = \eta^{v_{k+1}} = \lambda\pi^{v_{k+1}}(\tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^k - h^k) = \lambda\pi^{v_{k+1}}(h^{k+1} - h^k) \leqslant \lambda \cdot \max_{i\in\Phi}\{(h^{k+1} - h^k)(i)\}.$$

Equation (13) yields $\tilde{f}^{v^*} + \tilde{P}^{v^*} h^k \geqslant \tilde{f}^{v_{k+1}} + \tilde{P}^{v_{k+1}} h^k = h^{k+1}$, so

$$\eta^{v^*} = \lambda\pi^{v^*}(\tilde{f}^{v^*} + \tilde{P}^{v^*} h^k - h^k) \geqslant \lambda\pi^{v^*}(h^{k+1} - h^k) \geqslant \lambda \cdot \min_{i\in\Phi}\{(h^{k+1} - h^k)(i)\}.$$

From the above equations, we obtain $\eta^{v_\varepsilon} - \eta^{v^*} \leqslant \lambda \cdot sp(h^{k+1} - h^k) < \varepsilon$, that is, v_ε is an ε-optimal policy. □

6 A numerical example

Consider a CTMCP with finite state space $\Phi = \{1, 2, 3\}$ and feasible compact action set $D(i) = [0.5, 20]$, $i = 1, 2, 3$. Under a policy $v = (v(1), v(2), v(3))$, elements of the transition matrix corresponding to an embedded Markov chain are of the following form

$$p_{11}(v(1)) = 1 - e^{-v(1)\ 2}, \qquad p_{12}(v(1)) = \frac{7}{8} e^{-v(1)/2};$$

$$p_{13}(v(1)) = \frac{1}{8} e^{-v(1)/2}, \qquad p_{21}(v(2)) = \frac{1 - e^{-v(2)/4}}{1 + e^{-v(2)\ 2}};$$

$$p_{23}(v(2)) = \frac{1}{3} e^{-v(2)/4}, \qquad p_{22}(v(2)) = 1 - p_{21}(v(2)) - p_{23}(v(2));$$

$$p_{31}(v(3)) = \frac{1 - e^{-v(3)\ 4}}{1 + e^{-v(3)/4}}, \qquad p_{33}(v(3)) = 1 - \frac{1 - e^{-v(3)/3}}{1 + e^{-v(3)/2}},$$

$$p_{32}(v(3)) = 1 - p_{31}(v(3)) - p_{33}(v(3)).$$

The performance function is $f(i,v(i))=\ln[(1+i)v(i)]+\sqrt{i}/2v(i)$, and let $\lambda(i)=1, i=1,2,3$.

This problem can be solved by using a direct gradient-based method[11]. With an initial policy $v_0=(1,1,5)$, we obtain a policy $v^*=(0.71026319, 0.87263717, 5.54840733)$ that is assumed to be optimal, the corresponding optimal cost is $\eta^*=1.88858599$. The whole computation time is 19. 4 seconds. If we use the value iteration method with the same initial policy v_0 and different ε, we obtain the results shown in Table 1.

Table 1 The results obtained by using the value iteration algorithm

ε	v_ε	η^{v_ε}	t_s(s)
0. 1	(0. 70191875,0. 86868196,5. 42041131)	1. 88861989	0. 05
0. 01	(0. 71083730,0. 87297918,5. 56273383)	1. 88858632	0. 11
0. 001	(0. 71038481,0. 87264556,5. 54579576)	1. 88858600	0. 17
0. 0001	(0. 71025866,0. 87263011,5. 54757805)	1. 88858599	0. 22

From Table 1, we see that the value iteration algorithm leads to an ε-optimal policy v_ε with high speed. For large-scale problems, the algorithm will have a notable advantage over the traditional gradient-based algorithms in computation speed. Observe that, in each iteration of the value iteration algorithm, we only need to search an improving action for every state respectively. But in a traditional gradient-based method, we have to search an improving policy in an M-dimensional gradient direction and calculate potentials.

7 Conclusions

We conclude that the optimization methods for CTMCPs based on Markov performance potentials have obvious advantages. First, we only need some weak assumptions, and the Assumptions 1～4 in this paper can be satisfied in many real systems; secondly, the proposed iteration algorithm will ensure an ε-optimal policy with high speed; in addition, the methods make parallel computing possible. Therefore, the results of this paper provide a new way for performance optimization of large-scale real CTMCPs, such as queuing networks.

References

1 Arapostathis A, Borkar V S, Fernandez-Gaucher E, Ghosh M K, Marcus S I. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM Journal of Control and Optimization*, 1993, **31**(2): 282～344

2 Raul Montes-de-Oca. The average cost optimality equation for Markov control processes on Borel spaces. *System & Control Letters*, 1994, **22**(5): 351～357

3 Sennot L I. Another set of conditions for average optimality in Markov control processes. *Systems & Control Letters*, 1995, **23**(2): 147～151

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

5 Cao X R. A unified approach to Markov decision problems and performance sensitivity analysis. *Automatica*, 2000, **36**(5): 771～774

6 Yin Bao-Qun, Zhou Ya-Ping, Yang Xiao-Xian, Xi Hong-Sheng, Sun De-Ming. Sensitivity formulas of performance in closed state-dependent queuing networks. *Control Theory and Applications*, 1999, **16**(2): 255～257

7 Yin Bao-Qun, Zhou Ya-Ping, Xi Hong-Sheng, Sun De-Ming. Sensitivity formulas of performance in two-server cyclic queuing networks with phase-type distributed service times. *International Transactions in Operation Research*, 1999, **6**(6): 649～663

8 Zou Chang-Chun, Xi Hong-Sheng, Yin Bao-Qun, Zhou Ya-Ping, Sun De-Ming. Derivative estimates parallel simulation algorithms based on performance potentials theory. In: Proceedings of IFAC 14th World Congress, Beijing: IFAC, 1999,**J**: 49～54

9 Cao X R, Chen H F. Perturbation realization, potentials and sensitivity analysis of Markov processes. *IEEE Transactions on Automatic Control*, 1997, **42**(10): 1382～1393

10 Puterman M L. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: Wiley, 1994

11 Zhou Ya-Ping, Yin Bao-Qun, Xi Hong-Sheng, Tan Xiao-Bin, Sun De-Ming. Algorithms of decentralized optimiza-

tion for a class of closed queuing network by using performance potentials. *Journal of Unıversıty of Scıence and Technology of Chına*, 2000, **30**(2): 151~157

XI Hong-Sheng Received his bachelor and master degrees from University of Scıence and Technolog of China (USTC) in 1977 and 1985, respectively. He is currently a professor of USTC and a vice director o the department of automation. His research interests include the optimization and applications of discrete event dynamic systems.

TANG Hao Received his bachelor degree from Anhui Institute of Technology of Chına in 1995 an master degree from Instıtute of Plasma Physics, Chinese Academy of Sciences, in 1998. He is currently Ph. D. candidate of University of Science and Technology of China. His research interests include the optimızation and applıcations of discrete event dynamic systems, and the methodology of neuro-dynamic programmıng.

连续时间 MCP 在紧致行动集上的最优策略

奚宏生 唐 昊 殷保群

(中国科学技术大学自动化系 合肥 230026)

(E-mail: xihs@ustc. edu. cn)

摘 要 文中研究了一类连续时间 Markov 控制过程(CTMCP)无穷水平平均代价性能的最优控制决策问题. 文章采用无穷小生成元和性能势的基本性质,直接导出了平均代价模型在紧致行动集上的最优性方程及其解的存在性定理,提出了求解 ε-最优平稳控制策略的数值迭代算法,并给出了这种算法的收敛性证明. 最后通过分析一个数值例子来说明这种方法的应用.

关键词 性能势,平均代价准则,紧致行动集,数值迭代

中图分类号 TP202