Constrained Continuous-Time Markov Decision Processes in Polish Spaces

Xianping Guo
(Zhongshan University, Guangzhou)

Beijing Normal University, 22 July, 2010

Outline

1. Model of constrained MDPs
2. Optimality problem
3. Conditions for regularity and finiteness
4. Existence of optimality policies
5. Algorithms for optimal policies
6. Numerable examples
7. Remarks

1. Model of constrained MDPs

$$\{S, (A(x) \subset A, x \in S), q(\cdot|x,a), r(x,a), (c_n(x,a), d_n)\},$$

- S : the state space, a Borel space;
- $A(x) \subseteq A$: the admissible action sets;
- $q(\cdot|x,a)$: the transition rates, $a \in A(x), x \in S$;
- $r(x,a)$: the reward, $a \in A(x), x \in S$.
- $c_n(x,a)$: the costs, $a \in A(x), x \in S$.
- d_n : constrained constants, $1 \le n \le N$

2. **The optimality problem**

Notation:

- $\Omega^0 := (S \times R_+)^\infty$, with $R_+ := (0, \infty), x_\infty \notin S$,
- $\Omega := \Omega^0 \cup \{(x_0, \theta_1, x_1, \ldots, \theta_{k-1}, x_{k-1}, \infty, x_\infty, \ldots)| \quad \theta_l > 0, x_l \in S \text{ for each } 0 \leq l \leq k-1 \text{ and } k \geq 1\}$.
- $X_k(e) := x_k$, $T_k(e) := \theta_0 + \ldots + \theta_k, k = 0, \ldots (\theta_0 := 0)$
- $\xi_t(e) := \sum_{k \geq 0} x_k I_{\{T_k \leq t < T_{k+1}\}}(e) + x_\infty I_{\{T_\infty \leq t\}}(e), t \in [0, \infty)$, where $e := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \ldots) \in \Omega$.

- Introduce the integer-valued random measure μ^*

$$\mu^*(e, dt, dx) = \sum_{k \geq 0} I_{\{T_k < \infty\}}(e) \delta_{(T_k(e), X_k(e))}(dt, dx), \quad (1)$$

where $\delta_y(\cdot)$ is the Dirac measure at point y.

- Define the *predictable* σ-algebra:

$$\mathcal{P} := \sigma(B \times \{0\}, C \times (s, \infty) | \; B \in \mathcal{F}_0, C \in \mathcal{F}_{s-}, s > 0),$$

where $\mathcal{F}_t := \sigma\{\mu^*([0, s] \times D), s \in [0, t], D \in \mathcal{B}(S)\}$.

Definition 1. (Policies)

- Randomized history-dependent policy π: Transition probability π from $(\Omega \times R_+^0, \mathcal{P})$ onto $(A_\infty, \mathcal{B}(A_\infty))$ such that $\pi(A(\xi_{t-}(e))|e,t) \equiv 1$.
- Randomized stationary policy ϕ: Transition probability ϕ from $(S, \mathcal{B}(S))$ onto $(A, \mathcal{B}(A))$ such that $\phi(A(x)|x) \equiv 1$.
- Stationary policy f: A measurable function f from $(S, \mathcal{B}(S))$ onto $(A, \mathcal{B}(A))$ such that $\phi(\{f(x)\}|x) \equiv 1$.
- Π, Π_s, F: Corresponding classes of three kinds of policies.

Definition 2. (Policy measure)

Fix any each $\pi \in \Pi$ and initial distribution on S:

- The existence of a unique probability measure P_γ^π on $(\Omega, \mathcal{F})$, such that $P_\gamma^\pi\{x_0 \in dx\} = \gamma(dx)$, is ensured.
- Policy measure: P_γ^π depending on π.
- E_γ^π: Expectation operator with respect to P_γ^π
- E_x^π and P_x^π denote E_γ^π and P_γ^π respectively, when γ is the Dirac measure at point x.

Basic assumptions:

Regularity of the process: $\{\xi_t, t \geq 0\}$: $P_x^\pi(\xi_t \in S) \equiv 1$.

Assumption A. There exist a continuous function $w \geq 1$ on S and constants $\rho, b \geq 0$ and a sequence of nondecreasing subsets $\{S_k\}$ of S, such that

(1) $\int_S w(y)q(dy|x,a) \leq \rho w(x) + b \quad$ for all $(x,a) \in K$;

(2) $\inf_{x \notin S_k} w(x) \uparrow +\infty$ as $k \to \infty$, with $\inf \emptyset := \infty$;

(3) $S_k \uparrow S$, and $\sup_{a \in A(x), x \in S_k} |q(\{x\}|x,a)| < \infty$ for $k \geq 1$

Assumption A ensures the regularity of $\{\xi_t, t \geq 0\}$!

Optimality criteria:

Let $c_0(x,a) := r(x,a)$;

The discounted criteria: for $0 \le n \le N$,

$$V_n(x,\pi) := \int_0^\infty e^{-\alpha t} \int_A E_x^\pi \left[c_n(\xi_{t-},a)\pi(da|e,t)\right] dt,$$

$$V_n(\pi) := \int_S V_n(x,\pi)\gamma(dx)$$

Denote by

$$U := \{\pi |\ V_n(\pi) \le d_n, n = 1, \ldots, N\}.$$

the set of all constrained policies.

Definition 3.

- A policy π^* in U is called **constrained-optimal** if $V_0(\pi^*) \geq V_0(\pi)$ for all $\pi \in U$.
- A policy π^* in Π is said to be **optimal policy** if $V_0(\pi^*) \geq V_0(\pi)$ for all $\pi \in \Pi$

Main goal:

(a) Find conditions ensuring the existence of (constrained) optimal policies;

(b) Give algorithms for solving a (constrained) optimal policies.

3. Conditions for regularity and finiteness

Theorem 1. Under Assumption A, we have

(a) $P_x^\pi(T_\infty = \infty) = 1$, and $P_x^\pi(\xi_t \in S) = 1$.

(b) $E_x^\pi[w(\xi_t)] \le e^{\rho t} w(x) + \frac{b}{\rho}(e^{\rho t} - 1)$

(c) The analog of the forward Kolmogorov equation holds:

$$P_x^\pi(\xi_t(\omega) \in D) = I_D(x) + E_x^\pi[\int_0^t \int_A \pi(da|e,s) q(D|\xi_{s-}(e), a) ds]$$

Remark 1: Theorem 1 generalizes the corresponding results for Markov chains.

Assumption B. (Finiteness conditions).

(1) There exists a constant $M > 0$ such that, $|c_n(x, a)| \leq Mw(x)$ for every $(x, a) \in K$ and $n = 0, 1, \ldots, N$.

(2) The discount factor α verifies that $\alpha > \rho$, with ρ as in Assumption A.

(3) $\int_S w(x)\gamma(dx) < \infty$.

(4) $q^*(x) \leq Lw(x)$ for all $x \in S$, with some constant $L > 0$.

Theorem 2. Under Assumptions A and B, we have

(a) $E_x^{\pi}[|c_n(\xi_t, a)|\pi(da|e,t)] \leq M E_x^{\pi}[w(\xi_t)]$ for all $t \geq 0$

(b) $|V_n(x,\pi)| \leq M[\alpha w(x) + b]/[\alpha(\alpha - \rho)]$,

(c) $|V_n(\pi)| \leq M M_1^*, M_1^* := [\alpha \int_S w(x)\gamma(dx) + b]/[\alpha(\alpha - \rho)]$.

4. Existence of optimal policies

Definition 3. Fix policies $\pi, \pi_1, \pi_2 \in \Pi$.

(i) Occupation measure of π: η^π, which is defined by

$$\eta^\pi(D \times \Gamma) := \alpha \int_0^\infty e^{-\alpha t} E_\gamma^\pi \left[I_{\{\xi_t \in D\}}(e) \pi(\Gamma | e, t) \right] dt,$$

for $D \in \mathcal{B}(S)$ and $\Gamma \in \mathcal{B}(A)$.

(ii) π^1 and π^2 are called equivalent if $\eta^{\pi^1} = \eta^{\pi^2}$.

(iii) For any p.m. η on $K := \{(x, a) | x \in S, a \in A(x)\}$, let

$$\eta(dx, da) =: \hat{\eta}(dx) \phi^\eta(da|x), \quad \text{where } \phi^\eta \in \Pi_s. \tag{2}$$

The original optimality problem is equivalent to

$$\text{maximize } \frac{1}{\alpha}\int_K c_0(x,a)\eta(dx,da) \tag{3}$$
$$\text{over } \eta \in \{\eta^\pi : \int_K c_n(x,a)\eta^\pi(dx,da) \le \alpha d_n, 1 \le n \le N\}.$$

To solve problem (3), we need

- to seek a certain compactness structure on the set of all occupation measures: $\{\eta^\pi : \pi \in \Pi\}$.
- to characterize an occupation measure.

Theorem 3. Under Assumption A, we ave

(a) η^π (for each fixed $\pi \in \Pi$) satisfies the following equation

$$\alpha\hat{\eta}^\pi(D) = \alpha\gamma(D) + \int_{S\times A} q(D|x,a)\eta^\pi(dx,da)$$

(b) Conversely, if a p.m. η on K satisfies

$$\alpha\hat{\eta}(D) = \alpha\gamma(D) + \int_{S\times A} q(D|x,a)\eta(dx,da)$$

and $\int_S |q(\{x\}|x,\phi^\eta)|\hat{\eta}(dx) < \infty$, then $\eta^{\phi^\eta} = \eta$, where ϕ^η is as in (2).

(c) If, in addition, Assumptions B(2)-B(4) are satisfied, then $\phi^{\eta^\phi} = \phi$ for all $\phi \in \Pi_s$.

To further analyze properties of occupation measures, let $\mathcal{P}_w(K)$ be the set of all p.m. on K.

$$\mathcal{M}_o := \{\eta^\pi | \int_S w(x)\hat{\eta}^\pi(dx) < \infty, \pi \in \Pi\} \subseteq \mathcal{P}_w(K),$$
$$\mathcal{M}_o^c := \{\eta \in \mathcal{M}_o | \int_{S\times A} c_n(x,a)\eta(dx,da) \leq \alpha d_n, 1 \leq n \leq N\}.$$

Definition 4. $\bar{w}$-weak topology on $\mathcal{P}_{\bar{w}}(S \times A)$ is defined by the $\bar{w}$-weak convergence as follows: A sequence $\{\eta_k, k \geq 1\} \subseteq \mathcal{P}_{\bar{w}}(S \times A)$ is called to $\bar{w}$-converge weakly to $\eta \in \mathcal{P}_{\bar{w}}(S \times A)$ (and written as $\eta_k \xrightarrow{\bar{w}} \eta$) if

$$\lim_{k \to \infty} \int_{S \times A} u(x,a) \eta_k(dx, da) = \int_{S \times A} u(x,a) \eta(dx, da),$$

for each continuous function $u(x,a)$ on $S \times A$ such that $|u(x,a)| \leq L_u w(x)$ for all $(x,a) \in K$, with some nonnegative constant L_u depending on u.

$\eta_k \xrightarrow{\bar{w}} \eta$ implies the standard weak convergence of p.m.

Theorem 4. Under Assumptions A, B(2)-B(4), we have

(a) $\mathcal{M}_o$ and $\mathcal{M}_o^c$ are convex.

(b) If, in addition, $\int_S g(y)q(dy|x,a)$ is continuous on K for each bounded continuous functions g, then $\mathcal{M}_o$ is closed (with respect to the w-weak topology).

For the solvability of (3), by Theorem 4 we introduce the following condition.

Assumption C. Let w be as in Assumption A.

(1) The functions $c_n(x,a)$ and $\int_S g(y)q(dy|x,a)$ are continuous on K for bounded continuous functions g;

(2) There exist a measurable function $w' \geq 1$ on S and an increasing sequence of compact sets $K_m \uparrow K$, such that $\lim_{m\to\infty} \inf_{(x,a)\notin K_m} \frac{w(x)}{w'(x)} = \infty$, where $\inf \emptyset := \infty$;

Remark 2. Assumption C is new.

Theorem 5. Under Assumptions A, B, and C, we have

(a) $\mathcal{M}_o$ and $\mathcal{M}_o^c$ are metrizable and compact in the w-weak topology;

(b) there exists a constrained optimal policy.

Remark 3.

The conditions for Theorem 5(b) are weaker than those in the existing literature because some assumptions such as the nonnegativity of costs and the absolute integrability condition in the literature are not required here.

5. **Calculation of optimal policies**

First, by (3) and Theorem 3, it the original problem is equivalent to the following linear program (LP):

$$\mathrm{LP}:\ \sup_{\eta} \int_{S\times A} \frac{1}{\alpha} c_0(x,a)\eta(dx,da) \tag{4}$$

subject to

$$\begin{cases} \int_{S\times A} c_n(x,a)\eta(dx,da) \le \alpha d_n, n = 1,\ldots,N, \\ \alpha\hat{\eta}(D) = \alpha\gamma(D) + \int_{S\times A} q(D|x,a)\eta(dx,da) \\ \qquad\qquad \text{for all } D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty, \\ \int_S w(x)\hat{\eta}(dx) < \infty,\ \eta \in \mathcal{P}(K). \end{cases}$$

Thus, we obtain the following result on the solvability of constrained optimal policies.

Theorem 6. Under Assumptions A, B and C(3), the following assertions hold.

(a) If there exists a feasible solution to LP (4), then the set U of constrained policies is nonempty. Conversely, if U is nonempty, then there exists a feasible solution to LP (4).

(b) If there exists an optimal solution η^* to LP (4), then the randomized stationary policy ϕ^{η^*} is constrained optimal. Conversely, if π^* is constrained optimal, then η^{π^*} is an optimal solution to LP (4).

When S and $A(x)$ are finite, then LP (4) is the form of

$$\text{maximize} \sum_{x\in S} \sum_{a\in A(x)} \frac{1}{\alpha} c_0(x,a)\eta(x,a)$$

subject to

$$\begin{cases} \sum_{x\in S}\sum_{a\in A(x)} c_1(x,a)\eta(x,a) \le \alpha d_1 \\ \quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \sum_{x\in S}\sum_{a\in A(x)} c_n(x,a)\eta(x,a) \le \alpha d_N, \\ \alpha \sum_{a\in A(x)} \eta(x,a) = \alpha\gamma(x) \\ \qquad\qquad + \sum_{y\in S}\sum_{a\in A(y)} q(x|y,a)\eta(y,a) \quad \forall x \in S, \\ \eta(x,a) \ge 0, x \in S, a \in A(x), \end{cases} \tag{5}$$

which is an LP and can be solved by many methods such as the well-known simplex method.

6. Examples

Example 1.

- Let $S := (-\infty, \infty)$,
- $A(x) := [\beta_0, \beta(|x|+1)]$, with some constants $0 < \beta_0 < \beta$.
- Consider the transition rates $q(\cdot|x,a)$:

$$q(D|x,a) := (|x|+1)[\int_{D-\{x\}} f(y|x,a)dy - \delta_x(D)]$$

where $f(y|x,a) := \frac{1}{\sqrt{2\pi a}} e^{-\frac{(y-x)^2}{2a}}$.

Assumption D.

(1) $\alpha > \beta$, and $\int_S x^2 \gamma(dx) < \infty$. (Hence, there exists a constant ρ such that $\beta < \rho < \alpha$);

(2) $c_n(x,a)$ $(0 \le n \le N)$ are continuous on K and $|c_n(x,a)| \le L'(x^2+1)$ for all $(x,a) \in K$, with some constant $L' > 0$, where $c_0(x,a) := -r(x,a)$.

Proposition 1. Under Assumption D, Example 1 satisfies Assumptions A, B, and C. Therefore , there exists a constrained optimal policy for Example 1.

Example 2(on optimal policies). With the same data as in Example 1, we further suppose that $r(x,a)$ in Example 1 is given by

$$r(x,a) := px^2 - \delta a^2, \tag{6}$$

where $p, \delta > 0$ are fixed constants.

Assumption E.

(1) $d_n \geq L'[\alpha \int_S x^2 \gamma(dx) + \alpha + b]/[\alpha(\alpha - \beta)]$ for all $1 \leq n \leq N$), with $b := \beta(\frac{\rho+2\beta}{\rho-\beta} + 2)^2$;

(2) $2\alpha\beta_0 - \beta_0^2 \leq \frac{p}{\delta} \leq \min\{\alpha^2, 2\alpha\beta - \beta^2\}$, with p, δ as in (6).

Proposition 2. Under Assumptions D and E, we have

(a) The stationary policy f^* is optimal for Example 2, where

$$f^*(x) := (\alpha - \sqrt{\alpha^2 - \frac{p}{\delta}})(|x| + 1) \quad \forall x \in S.$$

(b) $V_0(f^*) = \int_S V_0(f^*, x)\gamma(dx)$, where

$$\begin{aligned} V(f^*, x) \; = \; & (2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta})x^2 \\ & +(4\delta\alpha - 4\sqrt{\delta^2\alpha^2 - p\delta} - \frac{2p}{\alpha})|x| \\ & +2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta} - \frac{p}{\alpha}. \end{aligned}$$

6. **Remarks**

(1) The existing works on continuous-time Markov decision processes can be classified into two groups:

- Group 1: Bounded transition rates, and history-dependent policies;
- Group 2: Unbounded transition rates, and Markov policies.
- Open problem: Unbounded transition rates, and history-dependent policies; see, for instance, Yushkevich, A.A.,

Theory Probab. Appl. **22**(1977), 215-235.

(2) In Examples 1 and 2, the transition rates and rewards are allowed to be unbounded, and policies may be history-dependent.

(3) From this talk, we can see some developments on the open problem by A. A. Yushkevich (1977).

Bibliography

[1] Altman, E. (1999). *Constrained Markov Decision Processes.* Chapman & Hall/CRC.

[2] Altman, E. and Shwartz, A. (1991). Markov decision problems and state-action frequencies. *SIAM J. Control Optim.* **29**, 786–809.

[3] Ash, R.B. (2000). *Probability and Measure Theory.* Academic Press.

[4] Bertsekas, D. P. and Shreve, A. (1996). *Stochastic Optimal Control: The Case of Discrete-Time Case.* Belmon, MA Athena Scientific.

[5] Chen, M.F. (2004). *From Markov Chains to Non-Equilibrium Particle Systems.* Second edition. World Scientific Publishing Co., Inc., River Edge, NJ.

[6] Chen, R.C and Feinberg, E.A. (2007). Non-randomized policies for constrained Markov decision processes. *Math. Methods Oper. Res.* **66**, 165–179.

[7] Feinberg, E.A. (2000). Constrained discounted Markov decision processes and Hamiltonian cycles. *Math. Oper. Res.* **25**, 130–140.

[8] Feinberg, E.A. and Shwartze, A. (1999). Constrained dynamic programming with two discount factors: applications and an algorithm. *IEEE Trans. Automat. Control* **44**, 628–631.

[9] Feinberg, E.A. and Shwartze, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.* **21**, 922–945.

[10] Feinberg, E.A. and Shwartze, A. (1995). Constrained Markov decision models with weighted discounted rewards. *Math. Oper. Res.* **20**, 302–320.

[11] Guo, X.P. (2007). Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces, *Math. Oper. Res.* **32**, 73-87.

[12] Guo, X.P. (2007). Constrained optimality for average cost continuous-time Markov decision processes. *IEEE Trans. Automat. Control* **52**, 1139-1143.

[13] Guo, X.P. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes*. Springer-Verlag, New York.

[14] Guo, X.P. and Hernández-Lerma, O. (2003). Continuous–time controlled Markov chains. *Ann. Appl. Probab.* **13**, 363-388.

[15] Guo, X.P. and Hernáandez-Lerma, O. (2003). Constrained continuous-time Markov controlled processes with discounted criteria. *Stochastic Anal. Appl.* **21**, 379-399.

[16] Guo, X.P., Hernández-Lerma, O. and Prieto-Rumeau T. (2006). A survey of resent results on continuous-time Markov decision processes. *TOP* **14**, 177-246.

[17] Guo, X.P. and Rieder, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces, *Ann. Appl. Probab.* **16**, 730-756.

[18] Haviv, M. and Puterman, M.L. (1998). Bias optimality in controlled queuing systems. *J. Appl. Probab.* **35**, 136-150.

[19] Hernández-Lerma, O. and Govindan, T.E. (2001). Nonstationary continuous-time Markov control processes with discounted costs on infinite horizon. *Acta Appl. Math.* **67**, 277-293.

[20] Hernáandez-Lerma, O. and González-Hernández, J. (2000). Constrained Markov controlled processes in Borel spaces: the discounted case, *Math. Meth. Oper. Res.* **52**, 271-285.

[21] Hernáandez-Lerma, O.; Gonzáaez-Hernáandez, J., and López-Martáinez, R. R. (2003). Constrained average cost Markov control processes in Borel spaces. *SIAM J. Control Optim.* **42**, no. 2, 442-468.

[22] Hernández-Lerma, O. and Lasserre, J.B. (1999). *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[23] Hernández-Lerma, O. and Lasserre, J.B. (1996). *Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[24] Hordijk, A. and Spieksma, F. (1989). Constrained admissible control to a queueing system. *Adv. in Appl. Probaba.* **21**, 409-431.

[25] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nicodym derivatives, representation of martingales, *Z.Wahrscheinlichkeitstheor. verw. Geb.* **31**, 235-253.

[26] Kadota, Y.; Kurano, M. and Yasuda, M. (2006). Discounted Markov decision processes with utility constraints. *Comput. Math. Appl.* **51**, 279-284.

[27] Kakumanu, P. (1971). Continuously discounted Markov decision models with countable state and action spaces. *Ann. Math. Statist.* **42**, 919-926.

[28] Kitaev, M.Y. (1985). Semi-Markov and Jump Markov controlled models: Average cost criterion. *Theory Probaba. Appl.* **30**, 272-288.

[29] Kitaev, M.Y. and Rykov, V.V. (1995). *Controlled Queueing Systems*, CRC Press.

[30] Kurano, M., Nakagami J.I., and Huang Y. (2002). Constrained Markov decision processes with compact state and action spaces: the average case. *Optim.* **48**, 255-269.

[31] Lewis, M.E. and Puterman, M. (2001). A probabilistic analysis of bias optimality in unichain Markov decision processes. *IEEE Trans. Automat. Control* **46**, 96-100.

[32] Lund, R.B., Meyn, S.P. and Tweedie, R.L. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Probab.* **6**, 218-237.

[33] Phelps, R.P. (2001). *Lectures on Choquet's Theorem*, Springer.

[34] Piunovskiy, A.B. (2005). Discounted continuous time Markov decision processes: the convex analytic approach, *16th Triennial IFAC World Congress*. Praha, Chekh. Republic.

[35] Piunovskiy, A.B. (1998). A controlled jump discounted model with constraints, *Theory Probab. Appl.* **42**, 51–72.

[36] Piunovskiy, A.B. (1997). *Optimal control of random sequences in problems with constraints*, Kluwer Academic Publishers, Dordrecht.

[37] Prieto-Rumeau, T. and Hernndez-Lerma, O. (2008). Ergodic control of continuous-time Markov chains with pathwise constraints. *SIAM J. Control Optim.* **47**, 1888-1908.

[38] Puterman, M.L. (1994). *Markov Decision Processes*. Wiley, New York.

[39] Rockafellar, R.T. (1989). *Conjugate Duality and Optimization*. SIAM Philadephia.

[40] Sennott, L.I. (1991). Constrained discounted Markov chains. *Probab. Eng. Info. Sci.* **5**, 463-476.

[41] Sennott, L.I. (1999). *Stochastic Dynamic Programming and the Control of Queueing System.* Wiley, New York.

[42] Yushkevich, A.A. (1977). Controlled Markov models with contable states and continous time. *Theory Probab. Appl.* **22**, 215-7235.

[43] Zadorojniy, A. and Shwartz, A. (2006). Robustness of policies in constrained Markov decision processes. *IEEE Trans. Automat. Control.* **51**, 635-638.

[44] Zhang, L.L. and Guo, X.P. (2008). Constrained continuous-time Markov decision processes with average criteria. *Math. Methods Oper. Res.* **67**, 323-340.

Many Thanks !!!