Ann. of Appl. Math.
35:4(2019), 449-464

CONVERGENCE OF CONTROLLED MODELS FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH CONSTRAINED AVERAGE CRITERIA*†

Wenzhao Zhang, Xianzhu Xiong‡

(*College of Math. and Computer Science, Fuzhou University, Fuzhou 350108, Fujian, PR China*)

Abstract

This paper attempts to study the convergence of optimal values and optimal policies of continuous-time Markov decision processes (CTMDP for short) under the constrained average criteria. For a given original model $\mathcal{M}_\infty$ of CTMDP with denumerable states and a sequence $\{\mathcal{M}_n\}$ of CTMDP with finite states, we give a new convergence condition to ensure that the optimal values and optimal policies of $\{\mathcal{M}_n\}$ converge to the optimal value and optimal policy of $\mathcal{M}_\infty$ as the state space S_n of $\mathcal{M}_n$ converges to the state space S_∞ of $\mathcal{M}_\infty$, respectively. The transition rates and cost/reward functions of $\mathcal{M}_\infty$ are allowed to be unbounded. Our approach can be viewed as a combination method of linear program and Lagrange multipliers.

Keywords continuous-time Markov decision processes; optimal value; optimal policies; constrained average criteria; occupation measures

2000 Mathematics Subject Classification 90C40; 60J27

1 Introduction

Markov decision processes have wide application in queueing system, telecommunications systems, etc.; see, for instance, [2, 11, 13, 16, 18] and the reference therein. The existence and computation of optimal value and optimal policies form a hot research area in Markov decision processes. The basic method to study the existence of optimal policies include the dynamic programming approach, the linear programming and duality programming method. Based on above methods, the value iteration algorithms, policy iteration algorithms, linear programming algorithms for unconstrained optimality problems and linear programming algorithms for constrained optimality problems have been proposed; see, for instance, [4, 11, 13, 15]. However,

*The first author is supported by the NNSF of China (No.11801080).
†Manuscript received June 3, 2019; Revised October 10, 2019
‡Corresponding author. E-mail: xiongxianzhu2001@sina.com

these algorithms are only adapt to tackle the optimality problems with finite states. It is natural to use finite-state models to approximate the original model with denumerable state space or general Borel state space. Hence, from the theoretical and practical point of view, the convergence of optimal values and optimal policies are important and interesting issues in Markov decision processes.

In the discrete-time context, [3] considered the convergence of optimal value and optimal policies of Markov decision processes with denumerable states under the constrained expected discounted cost criteria. [5, 6] developed the approximation method of optimal value and optimal policies of Markov decision processes with Borel state and action spaces under the constrained expected discounted cost criteria.

In the continuous-time formulation, [16] studied the convergence of optimal value and optimal policies of Markov decision processes with denumerable states under the expected discounted cost and average cost criteria. [17, 18] developed an approximation procedure for CTMDP with denumerable state space under the finite-horizon expected total cost criterion and risk-sensitive finite-horizon cost criterion, respectively. For constrained optimal problem, [12] proposed an approach based on occupation measures to study the convergence problem of optimal value and optimal policies, and gave condition imposed on the original model with denumerable states to ensure the original model can be approximated by a sequence of CTMDP with finite states.

In this paper, we consider the similar convergence problem as in [12] with denumerable states but under the constrained expected average criteria. More precisely, the original controlled model has the following features: 1) The state space is denumerable and the action space is a Polish space; 2) the transition rates, cost and reward functions may be unbounded from above and from below. Firstly, by introducing the average occupation measures and Lagrange multipliers, we prove that the constrained optimality problem of each model $\mathcal{M}_n$ of CTMDP equals to a unconstrained optimality problem, and deduce the optimality equation which includes some Lagrange multipliers. These results are extension of the results in [16] for constrained optimality problem with one constraint. Then, we derive the bound of the Lagrange multipliers in each model $\mathcal{M}_n$. Secondly, according to the optimality equations, we give the exact bound of of the optimal values between the finite-state model $\mathcal{M}_n$ and the original model $\mathcal{M}_\infty$. Finally, using some approximation properties of expected average reward/cost, we obtain the asymptotic convergence of optimal policies of finite-state models to the optimal policy of the original model.

The rest of the paper is organized as follows. In Section 2, we introduce the constrained average model we are concerned with. In Section 3, we deduce the optimality equation of each constrained model $\mathcal{M}_n$ and give the error bounds of the