J. Math. Anal. Appl. 443 (2016) 1323–1361

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures ☆

Jonatha Anselmi [a], François Dufour [b], Tomás Prieto-Rumeau [c,*]

[a] *INRIA Bordeaux Sud Ouest, France*
[b] *Institut Polytechnique de Bordeaux; INRIA Bordeaux Sud Ouest, Team: CQFD and IMB, Institut de Mathématiques de Bordeaux, Université de Bordeaux, France*
[c] *UNED, Madrid, Spain*

A R T I C L E   I N F O

Article history:
Received 30 September 2015
Available online 8 June 2016
Submitted by J.A. Filar

Keywords:
Continuous-time Markov decision processes
Piecewise Lipschitz continuous control models
Approximation of the optimal value function
ϵ-optimal policy

A B S T R A C T

In this paper, we propose an approach for approximating the value function and an ϵ-optimal policy of continuous-time Markov decision processes with Borel state and action spaces, with possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion. Under adequate assumptions, which in particular include that the transition rate has a density function with respect to a reference measure, together with piecewise Lipschitz continuity of the elements of the control model, we approximate the original controlled process by a model with finite state and action spaces. The approximation error is related to the 1-Wasserstein distance between suitably defined probability measures and approximating measures with finite support. We also study the case when the reference measure is approximated with empirical distributions and we show that convergence of the approximations takes place at an exponential rate in probability.

1. Introduction

1.1. Motivation and contribution

This paper concerns approximating numerically the value function and computing an ϵ-optimal policy of a continuous-time Markov decision processes (CTMDP) with Borel state and action spaces, and possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion. From a theoretical point of view, such models have been extensively studied (using the techniques, e.g., of dynamic and linear programming), but except for some specific models, it is not possible to determine explicitly an

☆ This research was supported by grant MTM2012-31393 from the Spanish Ministerio de Economía y Competitividad.
* Corresponding author.
E-mail addresses: jonatha.anselmi@inria.fr (J. Anselmi), dufour@math.u-bordeaux1.fr (F. Dufour), tprieto@ccia.uned.es (T. Prieto-Rumeau).

http://dx.doi.org/10.1016/j.jmaa.2016.05.055
0022-247X/

optimal policy and the value function of the CTMDP. This shows the need for numerical methods to get quasi-optimal solutions for these problems.

In this paper we will deal with a CTMDP model $\mathcal{M}$ with Borel state space $\mathbf{X}$ and action space $\mathbf{A}$, and action sets $\mathbf{A}(x) \subseteq \mathbf{A}$ for $x \in \mathbf{X}$. Our main assumptions on $\mathcal{M}$ consist in supposing the existence of a strictly unbounded function w (bounding the transition and cost rates of the $\mathcal{M}$) satisfying suitable Lyapunov type conditions, and assuming that the elements of the control model and related functions (cost rate, densities, action sets multifunction, etc.) are piecewise Lipschitz continuous. This allows dealing with, e.g., discontinuous transition and cost rates, which is an important departure point from previous works; see, e.g., [8–10]. Our approximation technique proceeds in three steps, each one with its own approximation error (we try to give a flavor of our methods in this paper, and the steps described next are simplified versions of the techniques we will develop later).

1. The first step is to approximate $\mathcal{M}$ with control models $\{\mathcal{M}_k\}_{k\geq 1}$ with bounded transition and cost rates. The dynamic of $\mathcal{M}_k$ is similar to that of $\mathcal{M}$ as long as the original process remains in a specific subset $\mathbf{S}_k$ of the state space (with the property that $\mathbf{S}_k \uparrow \mathbf{X}$ as $k \to \infty$), and it is "killed" upon leaving the set $\mathbf{S}_k$. The error when approximating $\mathcal{M}$ with $\mathcal{M}_k$ is related to the strictly unbounded function w.
2. The second step consists in discretizing the state space of $\mathcal{M}_k$. To do so, it is assumed that the positive part $q_k^+(dy|x,a)$ of the transition rates of $\mathcal{M}_k$ is absolutely continuous with respect to a so-called reference probability measure μ_k, i.e., $q_k^+(dy|x,a) = p_k(y|x,a)\mu(dy)$ for some density function p_k. The idea is to approximate μ_k with a probability measure $\hat{\mu}_k$ with finite support $\hat{\Gamma}_k$, and then consider the finite state space $\hat{\Gamma}_k$. Typically, the error made with this approximation is shown to depend on the 1-Wasserstein distance between μ_k and $\hat{\mu}_k$, denoted by $\mathcal{W}_1(\mu_k, \hat{\mu}_k)$.
3. Finally, the action sets $\mathbf{A}(x)$ for $x \in \hat{\Gamma}_k$ are replaced with finite sets $\mathbf{A}_\delta(x) \subseteq \mathbf{A}(x)$. The error of this approximation is measured in terms of the Hausdorff distance between $\mathbf{A}(x)$ and $\mathbf{A}_\delta(x)$, which is assumed to be of (small) order $\delta > 0$.

We note that, in fact, steps 2 and 3 will be performed somehow simultaneously. Following this procedure we have therefore approximated $\mathcal{M}$ with a control model with finite state space $\hat{\Gamma}_k$ and finite action sets $\mathbf{A}_\delta(x)$. Choosing k large enough, and small enough $\delta > 0$ and $\mathcal{W}_1(\mu_k, \hat{\mu}_k)$, we can thus approximate the optimal discounted value and obtain an ϵ-optimal policy of $\mathcal{M}$.

Regarding the construction of a probability measure $\hat{\mu}_k$ with finite support that approximates μ_k in the 1-Wasserstein metric, several approaches are possible. One consists in deriving $\hat{\mu}_k$ starting from a covering of $\mathbf{S}_k$ with small radius. This "deterministic" approach allows controlling the distance $\mathcal{W}_1(\mu_k, \hat{\mu}_k)$ but it may pose additional computational challenge. Another possibility is to use a "random" approximation by considering the empirical probability measure $\hat{\mu}_k^n$ obtained from n i.i.d. draws from μ_k. The approximation error $\mathcal{W}_1(\mu_k, \hat{\mu}_k^n)$ (which is a random variable) is measured with a concentration inequality for the non-asymptotic deviation. In this case, the approximation errors converge in probability to zero at an exponential speed in the sample size n.

The approximation of discrete-time Markov decision processes (DTMDP) can be traced back to the middle of the 70's with the work of T.L. Morin [17, Section 2.3]. The techniques of approximation are typically based on the discretization of the state and actions spaces, and on the *convergence* properties of the operator associated to the dynamic programming optimality equation. Bounds on the error or convergence rates for the approximating control model can be derived depending on the regularity hypotheses made on the parameters of the model. There exists a huge literature related to that approach: see, among others, [1,2,7,8,12,14,16,24,27] and the references therein. To some extent, our approach here (for CTMDPs) is related to the references [1, Chapter 17] and [7] on DTMDPs. Indeed, a common hypothesis is to assume that the stochastic kernel governing the dynamics of the control model (either the transition rates or the transition probabilities) has a density function with respect to a reference measure. However, our approxi-