Engineering Applications of Artificial Intelligence 74 (2018) 294–302

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Neural approximations in discounted infinite-horizon stochastic optimal control problems

Giorgio Gnecco [a], Marcello Sanguineti [b,*]

[a] AXES Research Unit - IMT School for Advanced Studies, Piazza S. Francesco, 19 - 55100 Lucca, Italy
[b] DIBRIS - University of Genoa, Via all'Opera Pia, 13 - 16145 Genoa, Italy

ARTICLE INFO

Keywords:
Stochastic optimal control
Infinite horizon with discount
Optimal stationary closed-loop control
Neural networks
Approximation

ABSTRACT

Neural approximations of the optimal stationary closed-loop control strategies for discounted infinite-horizon stochastic optimal control problems are investigated. It is shown that for a family of such problems, the minimal number of network parameters needed to achieve a desired accuracy of the approximate solution does not grow exponentially with the number of state variables. In such a way, neural-network approximation mitigates the so-called "curse of dimensionality". The obtained theoretical results point out the potentialities of neural-network based approximation in the framework of sequential decision problems with continuous state, control, and disturbance spaces.

1. Introduction

This paper focuses on discounted infinite-horizon stochastic optimization problems, in which decisions (or controls) have to be chosen at each time stage, in such a way to maximize the expected value, with respect to the uncertainties, of a reward (or, equivalently, minimize an expected cost), expressed as a summation over an infinite number of stages. The decisions taken at each stage depend on *state variables*, which capture the "history" of the optimization process. Among the kinds of uncertainties that arise in problems from applied sciences and engineering, we mention outcomes of market analysis in production planning, rain inflows in water reservoirs systems, stock prices in financial applications, traveling times in traffic management, lengths of message queues in telecommunication networks, etc. We include in the model a discount factor, in such a way that the decision maker can weight the current reward (or cost) more than future ones. The presence of the discount factor also guarantees, under mild conditions, the existence of stationary (i.e., stage-independent) optimal closed-loop control functions (Bertsekas, 2012), which are more attractive from a computational point of view than stage-dependent ones. Typically, discounted infinite-horizon stochastic optimal control problems cannot be solved in closed form. Hence, one has to search for suboptimal solutions.

The approach that we follow in this work to derive suboptimal solutions is based on constraining the control functions to be made up of relatively few computational units with a simple structure, since such a choice has already proved effective in high-dimensional optimization tasks (see, e.g., Giulini and Sanguineti, 2009; Gnecco and Sanguineti, 2011; Juditsky et al., 1995; Kůrková and Sanguineti, 2008a, b and Smith, 1999 and the references therein, and, more specifically, Bertsekas, 1996, 2012; Gnecco and Sanguineti, 2012; Gnecco et al., 2012; Lewis and Liu, 2013; Lewis and Vrabie, 2009; Liu et al., 2013, 2015; Narendra and Mukhopadhyay, 1997; Powell, 2007; Sutton and Barto, 1998; Zhang et al., 2013; Zoppoli et al., 2002 for the case of neural-network-based approximation schemes applied to suboptimal control). More precisely, the closed-loop control functions that we consider take on the form of linear combinations of input–output maps computed by units belonging to some *dictionary* (Gnecco et al., 2011a, b; Gribonval and Vandergheynst, 2006). Well-known dictionaries are those made by ridge, radial, or kernel units, Hermite functions, trigonometric polynomials, and splines. When the elements of the dictionaries are nonlinearly parametrized, one has nonlinear approximation schemes; a well-known example is represented by sigmoidal neural networks (Kůrková and Sanguineti, 2002), where the parameters are the neural weights and the biases.

Our target consists in estimating, for a class of discounted infinite-horizon stochastic optimal control problems, the accuracy of suboptimal stationary closed-loop control functions obtained via neural networks. This choice of the approximation scheme is motivated by the extensive use, reported in the literature, of neural networks for the approximate solution of optimal control problems via Adaptive Dynamic Programming (ADP) algorithms (see Bertsekas, 1996, 2012; Gaggero et al., 2013,

* Corresponding author.
E-mail addresses: giorgio.gnecco@imtlucca.it (G. Gnecco), marcello.sanguineti@unige.it (M. Sanguineti).

https://doi.org/10.1016/j.engappai.2018.07.004
Received 22 June 2016; Received in revised form 12 July 2018; Accepted 13 July 2018
0952-1976/

2014; Gnecco and Sanguineti, 2008b, 2010; Lewis and Liu, 2013; Lewis and Vrabie, 2009; Liu et al., 2013, 2015; Powell, 2007; Sutton and Barto, 1998; Zhang et al., 2013 and the recent survey papers Kiumarsi et al., 2018; Wang et al., 2017). With respect to existing results, in this work we provide additional insights into the motivations for the use in ADP of neural-network-based approximation schemes with sigmoidal computational units, possibly guiding also the choice of the number of such units. Our main contributions are as follows:

(a) as a first step of our analysis, we replace the optimal stationary closed-loop control function by a suitable approximation, and we derive an upper bound on the corresponding loss in performance (Theorem 3.1);
(b) vice-versa, we investigate a case for which an upper bound on such a loss can be translated into an upper bound on the error in the approximation of the optimal stationary closed-loop control function (Theorem 3.3 and the related Theorem 3.2);
(c) we construct an example for which all the assumptions needed to obtain the results summarized in items (a) and (b) hold;
(d) finally, we combine the obtained estimates with an error bound for the approximation of the optimal stationary closed-loop control function through neural networks with sigmoidal computational units (Theorems 5.2 and 5.3).

As to relationships with the available literature, we recall that results in the flavor of our Theorems 3.1 and 3.2 were derived in Gaggero et al. (2013, 2014) and Gnecco and Sanguineti (2008b, 2010). However, in the models adopted therein, the transition between pairs of states is described by a correspondence, rather than by a state equation as in the present paper, and the optimization horizon is finite. Moreover, the investigation of how an upper bound on the error in approximate optimization translates into an upper bound on the approximation error of the optimal stationary closed-loop control function, made in our Theorem 3.3, was not performed in the above-mentioned works.

A state-equation model was considered also in Gnecco and Sanguineti (2016), but its analysis was limited again to the finite-horizon case, and specialized to the class of bilinear stochastic dynamical systems. Moreover, another significant difference with respect to the analysis performed in Gnecco and Sanguineti (2016) is in the way in which the required degree of smoothness for the optimal (stationary) closed-loop control function and cost-to-go function – needed for the successive application of neural-network-based approximation error bounds – is obtained. This is a nontrivial issue, since, usually, deriving smoothness results for infinite-horizon optimization problems requires more technical conditions than obtaining corresponding results for the finite-horizon case (see Gnecco and Sanguineti, 2008b, Section 5.1 for a short survey of such results).

Global smoothness results for the infinite-horizon case were reported in Gnecco and Sanguineti (2008b, Section 5.2), but the degree of smoothness is too small for the application of neural-network-based approximation error bounds able to mitigate the curse of dimensionality, such as the one reported in Theorem 5.1. Higher-order smoothness results were also reported in Gnecco and Sanguineti (2008b, Remark 5.7), but their local nature (particularly, the fact that they hold on usually unspecified sets) seriously limits the applicability of such bounds. In contrast, the higher-order global smoothness result provided by our Theorem 3.2(i) (inspired by Blume et al., 1982, Theorem 3.1) makes such an application possible (see Theorems 5.2 and 5.3).

The paper is organized as follows. In Section 2, the model of discounted infinite-horizon stochastic optimal control problems that we address is presented. Section 3 investigates the relationships between estimates of the approximation error of the optimal stationary closed-loop control function and estimates of the loss in performance determined by the use of suboptimal control functions. Section 4 presents an example for which all the assumptions made in Section 3 are verified, and discusses how to construct other examples for which this happens. Then, the analysis is specialized in Section 5 to the case in which neural networks are used to approximate the optimal stationary closed-loop control function. In particular, we focus on the possibility of mitigating the curse of dimensionality: we provide conditions under which the minimal number of network parameters required to achieve a desired accuracy of the suboptimal solutions does not grow exponentially with the number of state variables. Section 6 discusses the results and possible extensions. All the proofs are contained in Appendix A. Other technical details are provided in Appendix B.

2. Discounted infinite-horizon stochastic optimal control

We consider a stochastic dynamical system described by the state equation

$$\underline{x}_{t+1} = \underline{f}(\underline{x}_t, \underline{u}_t, \underline{\xi}_t), \quad t = 0, 1, \ldots, \tag{1}$$

where $\underline{x}_t \in X \subseteq \mathbb{R}^d$ is a continuous state vector, $\underline{x}_0 = \hat{\underline{x}} \in X$ is a given initial state, $\underline{u}_t \in U \subseteq \mathbb{R}^m$ is a continuous control vector, $\underline{\xi}_t \in \Xi \subseteq \mathbb{R}^r$ are mutually independent identically distributed random vectors, and $\underline{f} : X \times U \times \Xi \to X$ is a state transition function. The set X satisfies the constraint $X \supseteq \{\underline{y} \in \mathbb{R}^d : \underline{y} = \underline{f}(\underline{x}, \underline{u}, \underline{\xi}), \underline{x} \in X, \underline{u} \in U, \underline{\xi} \in \Xi\}$. We denote by $\underline{g}_t : X \to U$ the admissible closed-loop control functions (or policies) at time t, assuming that $\underline{x}_t$ is known to the decision maker at time stage t. In the model presented in the paper, the admissible closed-loop control functions are bounded and continuous functions of the state. Finally, $\beta \in (0, 1)$ denotes a fixed discount factor, used to actualize costs at future time stages.

We state the following discounted infinite-horizon stochastic optimal control problem (Problem SOCP$_\infty$). Without loss of generality, we assume that, at each time stage t, the same admissible closed-loop control function is applied (hence, in the following, $\underline{g}_t$ is replaced by $\underline{g}$). This is justified by the existence, under mild conditions, of an optimal stationary closed-loop control function for this kind of problem (see, e.g., Bertsekas, 2012; Bhattacharya and Majumdar, 2007).

Problem SOCP$_\infty$. Find an optimal stationary closed-loop control function $\underline{g}^\circ$ that minimizes the cost functional

$$J := \mathbb{E}_{\underline{\xi}_0, \underline{\xi}_1, \ldots} \left\{ \sum_{t=0}^{\infty} \beta^t h(\underline{x}_t, \underline{g}(\underline{x}_t), \underline{\xi}_t) \right\} \tag{2}$$

subject to the constraints $\underline{x}_t \in X, \underline{u}_t \in U$, and (1).

Remark 2.1. The definition of Problem SOCP$_\infty$ can be extended to the limit case $\beta = 0$, for which the functional (2) is replaced by

$$J := \mathbb{E}_{\underline{\xi}_0} \left\{ h(\underline{x}_0, \underline{g}(\underline{x}_0), \underline{\xi}_0) \right\} . \tag{3}$$

This is a (typically simpler) static optimization problem.

Let us consider the optimal stationary cost-to-go function,[1] which is defined as

$$J^\circ(\underline{x}_t) := \inf_{\underline{g}} \mathbb{E}_{\underline{\xi}_t, \underline{\xi}_{t+1}, \ldots} \left\{ \sum_{k=t}^{\infty} \beta^{k-t} h(\underline{x}_t, \underline{g}(\underline{x}_t), \underline{\xi}_t) \right\} . \tag{4}$$

Then, the stationary Dynamic Programming (DP) recursive equation (which holds under mild conditions, see Assumption 3.1(i–iii,vi) later) is given by

$$J^\circ(\underline{x}_t) = \inf_{\underline{u}_t \in U} \mathbb{E}_{\underline{\xi}_t} \left\{ h(\underline{x}_t, \underline{u}_t, \underline{\xi}_t) + \beta J^\circ(\underline{x}_{t+1}) \right\} . \tag{5}$$

We denote by $\tilde{\underline{g}}$ an approximation of the optimal stationary closed-loop control function $\underline{g}^\circ$. Using such an approximate closed-loop control function, we now consider an approximation of the optimal stationary cost-to-go function of the form

[1] In the following, depending on the context, we use either $\underline{x}$, $\underline{x}_0$, or $\underline{x}_t$ to denote the argument of J°.