Robust adaptive dynamic programming for linear and nonlinear systems: An overview ☆

Zhong-Ping Jiang *, Yu Jiang

Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY 11201, USA

ARTICLE INFO

ABSTRACT

The field of adaptive dynamic programming with diverse applications in control engineering has undergone rapid progress over the past few years. A new theory called "Robust Adaptive Dynamic Programming" (for short, RADP) is developed for the design of robust optimal controllers for linear and nonlinear systems subject to both parametric and dynamic uncertainties. A central objective of this paper is to give a brief overview of our recent contributions to the development of the theory of RADP and to outline its potential applications in engineering and biology.

© 2013 European Control Association. Published by Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Background

Approximate/adaptive dynamic programming (for short, ADP) is a biologically-inspired, non-model-based, computational method that has been used to compute optimal control laws; see, e.g., [43,49,62,64,66] and numerous references therein. It is well-known that conventional dynamic programming [3] requires the perfect knowledge of system dynamics and suffers from the curse of dimensionality. To avoid these difficulties, Werbos first pointed out in [61] that adaptive approximation to the Hamilton–Jacobi–Bellman (HJB) equation [37] can be achieved by designing appropriate reinforcement learning systems (see, [53] for an excellent introduction to the theory of reinforcement learning). In his seminal work [63,65,66], Werbos further proposed two basic approaches for implementing ADP: heuristic dynamic programming (HDP) and dual dynamic programming. They can be used to approximate the optimal cost function or its gradient, and their generalized versions can be found in [66] in which the approximation of the optimal control policy is considered. Similar problems were also studied by Bertsekas and Tsitsiklis [5] under the name of neuro-dynamic programming and were restricted exclusively to discrete-time systems. A rigorous development of

the mathematical principles behind neuro-dynamic programming is provided, along with numerous methods and applications.

The development of ADP theory consists of three phases. In the first phase, ADP was extensively investigated within the communities of computer science and operations research. Two basic algorithms, policy iteration [17] and value iteration [3], are usually employed. In [52], Sutton introduced the temporal difference method. In 1989, Watkins proposed the well-known Q-learning method in his PhD thesis [60]. Q-learning shares similar features with the action-dependent HDP scheme proposed by Werbos in [64]. Other related research work under a discrete time and discrete state-space Markov decision process framework can be found in [4,5,7,9,43,44,45,53,54] and reference therein. In the second phase, stability is brought into the context of ADP while real-time control problems are studied for dynamic systems. To the best of the authors' knowledge, Lewis is the first who has contributed to the integration of stability theory and ADP theory [38]. An essential advantage of ADP theory is that an optimal control policy can be obtained via a recursive numerical algorithm using online information without solving the HJB equation (for nonlinear systems) and the algebraic Riccati equation (ARE) (for linear systems), even when the system dynamics are not precisely known. Optimal feedback control designs for linear and nonlinear dynamic systems have been proposed by several researchers over the past few years; see, e.g., [6,8,11,16,42,56,58,59,68,69]. While most of the previous work on ADP theory was devoted to discrete-time systems (see [36] and references therein), there has been relatively less research for the continuous-time counterpart. This is mainly because ADP is considerably more difficult for continuous-time systems than for discrete-time systems. Indeed, many results

developed for discrete-time systems [39] cannot be extended straightforwardly to continuous-time systems. Nonetheless, early attempts were made to apply Q-learning for continuous-time systems via discretization technique [2,12]. However, the convergence and stability analysis of these schemes are challenging. In [42], Murray et al. proposed an implementation method which requires the measurements of the derivatives of the state variables. As said previously, Lewis and his co-worker proposed the first solution to stability analysis and convergence proofs for ADP-based control systems by means of LQR theory [58]. A synchronous policy iteration scheme was also presented in [55]. For continuous-time linear systems, the partial knowledge of the system dynamics (i.e., the input matrix) must be precisely known. This restriction has been completely removed in [21]. A nonlinear variant of this method can be found in [27].

The third phase in the development of ADP theory is related to extensions of previous ADP results to nonlinear uncertain systems. Neural networks and game theory are utilized to address the presence of uncertainty and nonlinearity in control systems. See, e.g., [56,57,69,36]. An implicit assumption in these papers is that the system order is known and that the uncertainty is static, not dynamic. The presence of dynamic uncertainty has not been systematically addressed in the literature of ADP. By dynamic uncertainty, we refer to the mismatch between the nominal model and the real plant when the order of the nominal model is lower than the order of the real system. A closely related topic of research is how to account for the effect of unseen variables [67]. It is quite common that the full-state information is often missing in many engineering and biological applications and only the output measurement or partial-state measurements are available. Adaptation of the existing ADP theory to this practical scenario is important yet non-trivial. Neural networks are sought for addressing the state estimation problem [13,32]. However, the stability analysis of the estimator/controller augmented system is by no means easy, because the total system is highly interconnected. The configuration of a standard ADP-based control system is shown in Fig. 1.

Our recent work [20,22–25] on the development of robust variants of ADP theory is exactly targeted at addressing these challenges.

1.2. What is RADP?

RADP is developed to address the presence of dynamic uncertainty in linear and nonlinear dynamical systems. See Fig. 2 for an
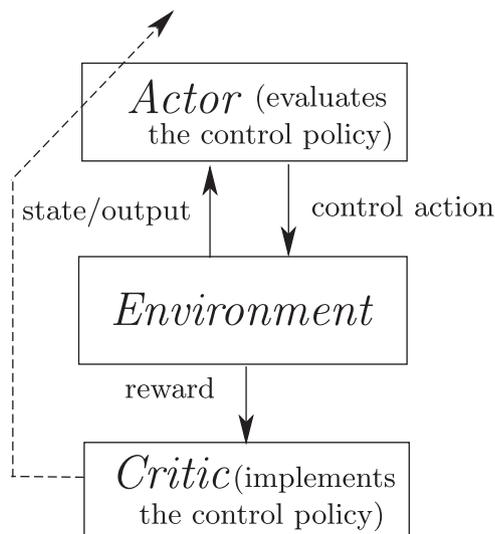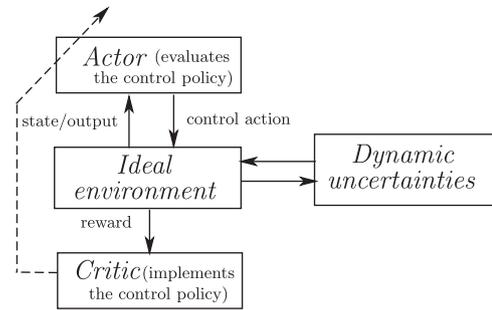

Fig. 2. RADP with dynamic uncertainty.

illustration. There are several reasons for which we pursue a new framework for RADP. First and foremost, it is well-known that building an exact mathematical model for physical systems often is a hard task. Also, even if the exact mathematical model can be obtained for some particular engineering and biological applications, simplified models are often more preferable for system analysis and control synthesis than the original complex system model. While we refer the mismatch between the simplified model and the original system to as dynamic uncertainty here, the engineering literature often uses the term of *unmodeled dynamics* instead. Secondly, the observation errors may often be captured by dynamic uncertainty. From the literature of modern nonlinear control [34,28,29], it is known that the presence of dynamic uncertainty makes the feedback control problem extremely challenging in the context of nonlinear systems. In order to broaden the application scope of ADP theory in the presence of dynamic uncertainty, our strategy is to integrate tools from nonlinear control theory, such as Lyapunov designs [34,19,31], input-to-state stability theory [51], and nonlinear small-gain techniques [30,28]. This way RADP becomes applicable to wide classes of uncertain dynamic systems with incomplete state information and unknown system order/dynamics.

Additionally, RADP can be applied to large-scale dynamic systems as shown in our recent paper [23]. By integrating a simple version of the cyclic-small-gain theorem [40], asymptotic stability can be achieved by assigning appropriate weighting matrices for each subsystem. Further, certain suboptimality property can be obtained. Because of several emerging applications of practical importance such as smart electric grid, intelligent transportation systems and groups of mobile autonomous agents, this topic deserves further investigations from a RADP point of view. The existence of unknown parameters and/or dynamic uncertainties, and the limited information of state variables, give rise to challenges for the decentralized or distributed controller design of large-scale systems.

2. ADP for completely unknown, continuous-time, linear systems

We begin with a linear time-invariant (LTI), continuous-time system

$$\dot{x} = Ax + Bu \tag{1}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the control input, $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ are unknown constant matrices, that are stabilizable. It is assumed that there is a known constant matrix K_0 such that $A - BK_0$ is a Hurwitz matrix.

The control objective is to find, if possible, an online, optimal control policy $u^* = -K^*x$ such that the following integral-quadratic

Fig. 1. Configuration of an ADP-based control system.