Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Reinforcement Learning with Function Approximation: From Linear to Nonlinear

Jihao Long * [1] and Jiequn Han † [2]

[1]Program of Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA
[2]Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, USA

Abstract. Function approximation has been an indispensable component in modern reinforcement learning algorithms designed to tackle problems with large state spaces in high dimensions. This paper reviews recent results on error analysis for these reinforcement learning algorithms in linear or nonlinear approximation settings, emphasizing approximation error and estimation error/sample complexity. We discuss various properties related to approximation error and present concrete conditions on transition probability and reward function under which these properties hold true. Sample complexity analysis in reinforcement learning is more complicated than in supervised learning, primarily due to the distribution mismatch phenomenon. With assumptions on the linear structure of the problem, numerous algorithms in the literature achieve polynomial sample complexity with respect to the number of features, episode length, and accuracy, although the minimax rate has not been achieved yet. These results rely on the L^∞ and UCB estimation of estimation error, which can handle the distribution mismatch phenomenon. The problem and analysis become substantially more challenging in the setting of nonlinear function approximation, as both L^∞ and UCB estimation are inadequate for bounding the error with a favorable rate in high dimensions. We discuss additional assumptions necessary to address the distribution mismatch and derive meaningful results for nonlinear RL problems.

Keywords:
Reinforcement Learning,
Function Approximation,
High-Dimensionality Analysis,
Distribution Mismatch.

Article Info.:
Volume: 2
Number: 3
Pages: 161 - 193
Date: September/2023
doi.org/10.4208/jml.230105

Article History:
Received: 05/01/2023
Accepted: 17/05/2023

Communicated by:
Weinan E

1 Introduction

Reinforcement learning (RL) studies how an agent can learn, through interaction with the environment, an optimal policy that maximizes his/her long-term reward [52]. When the problem involves a finite set of states and actions of moderate size, the corresponding value or policy functions can be represented precisely as a table, which is called the tabular setting. However, when the problem contains an enormous number of states or continuous states, often in high dimensions, function approximation must be introduced to approximate the involved value or policy functions. With the rapid development of machine learning techniques for function approximation, modern reinforcement learning (RL) algorithms increasingly rely on function approximation tools to tackle problems with growing complexity, including video games [41], Go [50], and robotics [32].

Despite the astonishing practical success of RL with function approximation when applied to challenging high-dimensional problems, the theoretical understanding of RL al-

*Corresponding author. jihaol@princeton.edu
†jiequnhan@gmail.com

gorithms with function approximation remains relatively limited, particularly when compared to the theoretical results in the tabular setting. In the tabular setting, roughly speaking, we can achieve the minimax sample complexity up to the logarithm term: we need samples of the order of $H^3|\mathcal{S}||\mathcal{A}|/\epsilon^2$ to obtain an ϵ-optimal policy, where H denotes the episode length, $|\mathcal{S}|$ and $|\mathcal{A}|$ denote the size of the state space and action space (see [8,13,25] for detailed discussions). Apparently, these kinds of results become vacuous when $|\mathcal{S}|$ (and/or $|\mathcal{A}|$) is extremely large or infinite. Therefore, the study of sample complexity in the presence of function approximation has received considerable attention in recent years in the RL community. Relatively simple function approximation methods, such as the linear model in [29, 57] or generalized linear model in [37, 56] have been examined in the context of RL algorithms. Meanwhile, nonlinear forms like kernel approximation [15, 39, 40, 58, 59] have also been studied in RL problems to further bridge the gap between theoretical results under restrictive assumptions and practice.

In this paper, we review recent theoretical results in RL with function approximation, from linear setting to nonlinear setting. We mainly focus on the results regarding approximation error and estimation error/sample complexity, which are errors introduced by function approximation and finite datasets, respectively. We first review the basic concepts of RL in Section 2 and introduce two categories of RL algorithms: value-based methods and policy-based methods in Section 3. We are interested in these algorithms when combined with function approximation. In Section 4, we give a general framework for the theoretical analysis of RL with function approximation. We adopt the concepts of approximation error, estimation error, and optimization error from supervised learning to RL and discuss the crucial challenges of analyzing these errors in RL. In Section 5, we introduce RL algorithms with linear function approximation, as it is the simplest function approximation. We introduce the basic linear MDP assumption [29], which assumes that both reward function and transition probability are linear with respect to d known features. Under this or similar assumptions, the Q-value function can be represented as a linear function with respect to the features, and numerous algorithms in the literature can achieve polynomial sample complexity with respect to the number of features d, episode length H, and accuracy ϵ. However, the minimax sample complexity has not been achieved yet.

In Section 6, we further discuss RL with nonlinear function approximation. We first introduce the theoretical results of supervised learning on reproducing kernel Hilbert space, neural tangent kernel, and Barron space, and then discuss how to analyze the approximation error in RL problems with nonlinear function approximation. We then focus on the distribution mismatch phenomenon, which is a crucial challenge of RL compared to supervised learning when analyzing the estimation error in the presence of function approximation. In tabular and linear settings, the distribution mismatch is handled by the L^∞ and UCB estimation. However, as we will point out, both L^∞ and UCB estimation suffer from the curse of dimensionality for various function spaces, including neural tangent kernel, Barron space, and many common reproducing kernel Hilbert spaces. This challenge reveals an essential difficulty of RL problems with nonlinear function approximation, and thus additional assumptions are needed to derive meaningful results for nonlinear RL in the literature, including assumptions on the fast eigenvalue decay of the kernel and assumptions on the finite concentration coefficient. We finally introduce the perturbational