Approximate Dynamic Programming: theoretical guarantees and practical algorithms for a continuous space setting

DISS. ETH NO. 26165

Approximate Dynamic Programming

theoretical guarantees and practical algorithms for a continuous space setting

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Paul Nathaniel Beuchat

MSc ETH ME, ETH Zurich
born on 18. February 1986
citizen of Australia and Switzerland

accepted on the recommendation of

Prof. Dr. John Lygeros, examiner (ETH Zurich, Switzerland)
Prof. Dr. Anders Rantzer, co-examiner (Lund University, Sweden)
Prof. Dr. Angelos Georghiou, co-examiner (McGill University, Canada)

2019

Abstract

Many problems in science and engineering can be cast as discrete time, continuous space, infinite horizon stochastic optimal control problems. The so-called value function and Q-function both characterise the solution of such problems, but are both intractable to compute in all but a few special cases. This thesis focuses on methods that approximate the value function and Q-function.

We consider the linear programming approach to approximate dynamic programming, which computes approximate value functions and Q-functions that are point-wise under-estimators of the optimal by using the so-called Bellman inequality. For this approximation method we provide theoretical guarantees on the value function and Q-function approximation error, and also for the sub-optimality of a policy generated using such lower-bounding approximations. In particular, the online performance guarantee is obtained by analysing an iterated version of the greedy policy, and the fitting error guarantee by analysing an iterated version of the Bellman inequality. These guarantees complement the existing bounds that appear in the literature.

Given a collection of lower-bounding approximate value functions, an improved approximation can be constructed by taking the point-wise maximum over the collection, however, the challenge is how to compute the collection itself. To address this challenge, we introduce a novel formulation, referred to as the point-wise maximum approach to approximate dynamic programming, and use this to propose algorithms that iteratively construct a collection of lower-bounding value functions with the objective of maximising the point-wise maximum of the collection. We empirically demonstrate the advantages of the proposed algorithm through a range numerical examples that indicate classes of problems where the proposed algorithms improves upon state-of-the-art methods. A key result from the numerical studies is that the proposed algorithm can provide practically useful sub-optimality bounds for the online performance of any policy, even when the collection of approximate value functions is itself impractical to use for a policy.

Zusammenfassung

Viele Probleme in Wissenschaft und Technik können als zeitdiskrete stochastische optimale Steuerungsprobleme mit kontinuierlichem Zustandsraum und mit unendlichem Horizont betrachtet werden. Die sogenannte Wertefunktion und die Q-Funktion charakterisieren beide die Lösung derartiger Probleme, sind jedoch beide in allen außer einigen speziellen Fällen schwer zu berechnen. Die vorliegende Dissertation konzentriert sich auf Methoden, welche die Wertefunktion und die Q-Funktion approximieren.

Wir betrachten den linearen Programmieransatz für die annähernde Dynamische Programmierung, bei welchem Wertefunktionen und Q-Funktionen approximiert werden, welche punktweise unter den Schätzungen des Optimums liegen, indem die sogenannte Bellman'sche Ungleichung verwendet wird. Für diese Approximationsmethode geben wir theoretische Garantien für den Approximationsfehler der Wertefunktion und der Q-Funktion sowie für die Suboptimalität eines mit solchen Approximationen erzeugten Regelgesetzes. Genauer, wird die theoretische Garantie der Leistung des Regelgesetzes durch das Analysieren einer iterierten Version des Greedy-Regelgesetzes hergeleitet. Um die theoretische Garantie des Approximationsfehlers herzuleiten, wird eine iterierende Version der Bellman'schen Ungleichung verwendet. Diese Garantien ergänzen die bestehenden Garantien, welche in der Fachliteratur dokumentiert sind.

Bei einer Sammlung von approximierten Wertefunktionen welche jeweils eine untere Beschränkung der optimalen Wertefunktion darstellen, kann eine verbesserte Näherung konstruiert werden, indem das punktweise Maximum über die Sammlung genommen wird. Die Herausforderung besteht jedoch darin, eine Methode zu entwickeln welche in einer brauchbaren Sammlung resultiert. Um dieser Herausforderung zu begegnen, führen wir eine neuartige Formulierung ein, welche als punktweiser Maximalansatz für die annähernde Dynamische Programmierung bezeichnet wird. Weiter schlagen wir Algorithmen vor, welche iterativ eine Sammlung von Funktionen mit unterer Beschränkung aufbauen, um das punktweise Maximum der Sammlung zu maximieren. Wir demonstrieren die Vorteile des vorgeschlagenen Algorithmus empirisch anhand numerischer Beispiele, welche auf Problemklassen hinweisen, bei denen die vorgeschlagenen Algorithmen den Stand der Technik verbessern. Ein Schlüsselergebnis der numerischen Studien ist, dass der vorgeschlagene Algorithmus nützliche Suboptimalitätsbeschränkungen für die Leistung jedes beliebigen Regelgesetzes bereitstellen kann, selbst wenn das durch die Sammlung von approximierten Wertefunktionen definierte Regelgesetz nicht implementierbar ist.