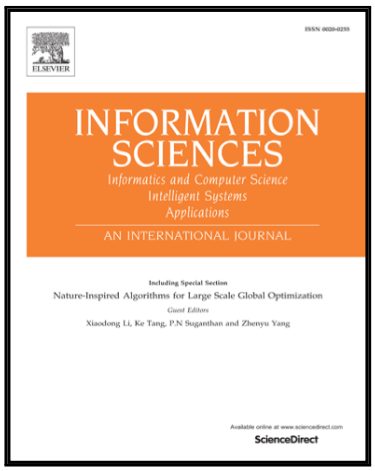

Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems

Biao Luo, Derong Liu, Tingwen Huang, Xiong Yang, Hongwen Ma

PII: S0020-0255(16)31285-3
DOI: 10.1016/j.ins.2017.05.005
Reference: INS 12875

To appear in: *Information Sciences*

Received date: 13 October 2016
Revised date: 1 March 2017
Accepted date: 6 May 2017

Please cite this article as: Biao Luo, Derong Liu, Tingwen Huang, Xiong Yang, Hongwen Ma, Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.05.005

Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems

Biao Luo*[a,1], Derong Liu[b], Tingwen Huang[d], Xiong Yang[c], Hongwen Ma[a]

[a] *The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[b] *The School of Automation, Guangdong University of Technology, Guangzhou 510006, China*
[c] *The School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China*
[d] *Texas A&M University at Qatar, PO Box 23874, Doha, Qatar*

Abstract

Policy iteration and value iteration are two main iterative adaptive dynamic programming frameworks for solving optimal control problems. Policy iteration converges fast while requiring an initial stabilizing control policy, which is a strict constraint in practice. Value iteration avoids the requirement of initial admissible control policy while converging much slowly. This paper tries to utilize the advantages of policy iteration and value iteration, and avoids their drawbacks at the same time. Therefore, a multi-step heuristic dynamic programming (MsHDP) method is developed for solving the optimal control problem of nonlinear discrete-time systems. MsHDP speeds up value iteration and avoids the requirement of initial admissible control policy in policy iteration at the same time. The convergence theory of MsHDP is established by proving that it converges to the solution of the Bellman equation. For implementation purpose, the actor-critic neural network (NN) structure is developed. The critic NN is employed to estimate the value function and its NN weight vector is computed with a least-square scheme. The actor NN is used to estimate the control policy and a gradient descent method is proposed for updating its NN weight vector. According to the comparative simulation studies on two examples, the effectiveness and advantages of MsHDP are verified.

Keywords: Optimal control, multi-step heuristic dynamic programming, adaptive dynamic programming, nonlinear systems, discrete-time, neural networks.