

"A man is  
great by  
deeds, not by  
birth"

-Chanakya

Welcome to IIMK



INDIAN INSTITUTE OF MANAGEMENT KOZHIKODE



**Working Paper**

**IIMK/WPS/556/ITS/2023/02**

**March 2023**

**A Note on Generalized Second Order Value Iteration in Markov  
Decision Processes**

V Anthony Vijesh<sup>1</sup>

Shreyas S R<sup>2</sup>

Mohammed Shahid Abdulla<sup>3</sup>

©

**All rights belong to their respective authors.**

**Please contact the corresponding authors if you would like to access the full case.**

<sup>1</sup>Assistant Professor, Department Of Mathematics, IIT Indore

<sup>2</sup>PhD student, Department Of Mathematics, IIT Indore

<sup>3</sup>Associate Professor, Information Systems Management Area, Indian Institute of Management Kozhikode, IIMK Campus PO Kunnamangalam, Kozhikode, Kerala 673 570, India; Email - shahid@iimk.ac.in, Phone Number - 0495-2809254,

## **A Note on Generalized Second Order Value Iteration in Markov Decision Processes**

### **Abstract:**

Value iteration is one of the first-order algorithms to approximate the solution of the Bellman equation arising from the Markov Decision Process (MDP). In recent literature, by approximating the max operator in the Bellman equation by a smooth function, an interesting second-order iterative method was discussed to solve the new Bellman equation. During the numerical simulation, it was observed that this second-order method is computationally expensive for a reasonable size of state and action. This second-order iterative method also faces difficulty in numerical implementation due to the calculation of an exponential function for larger values. In this manuscript, a few first-order iterative schemes have been derived from the second-order method to overcome the above practical problems. All the proposed iterative schemes possess the global convergence property. The proposed iterative schemes take less time to converge to the solution of the Bellman equation than the secondorder method in many cases. These algorithms are efficient and easy to implement. An interesting theoretical comparison is provided between the algorithms. Numerical simulation supports our theoretical results.

**Keywords:** Markov decision processes, Q-learning, reinforcement learning, value iteration

Research Office  
Indian Institute of Management Kozhikode  
IIMK Campus P. O.,  
Kozhikode, Kerala, India,  
PIN - 673 570  
Phone: +91-495-2809238  
Email: [research@iimk.ac.in](mailto:research@iimk.ac.in)  
Web: <https://iimk.ac.in/publications>

