CONSTRAINED MARKOV DECISION PROCESSES UNDER UNCERTAIN RUNNING COSTS AND TRANSITION PROBABILITIES

V VARAGAPRIYA

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
APRIL 2024

© Indian Institute of Technology Delhi (IITD), New Delhi, 2024

CONSTRAINED MARKOV DECISION PROCESSES UNDER UNCERTAIN RUNNING COSTS AND TRANSITION PROBABILITIES

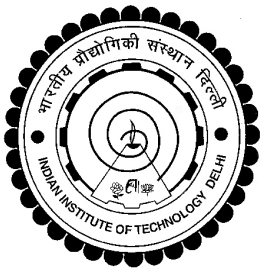
by

V Varagapriya

Department of Mathematics

Submitted
in fulfillment of the requirements of the degree of
Doctor of Philosophy

to the



Indian Institute of Technology Delhi

April 2024

Certificate

This is to certify that the thesis entitled **CONSTRAINED MARKOV DECISION PROCESSES UNDER UNCERTAIN RUNNING COSTS AND TRANSITION PROBABILITIES** submitted by **Ms V Varagapriya** to the **Indian Institute of Technology Delhi**, for the award of the degree of **Doctor of Philosophy**, is a record of the original bonafide research work carried out by her under my supervision and guidance. The thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

New Delhi
18/04/2024

Dr. Vikas Vikram Singh
Assistant Professor
Department of Mathematics
Indian Institute of Technology Delhi
New Delhi 110016

Acknowledgments

I would like to convey my profound gratitude to my supervisor, Prof. Vikas Vikram Singh, for his continuous support and guidance. His distinct way of assistance enabled me to carve my niche and grow as an independent researcher. He was constantly available, both offline and online, to address all my queries and motivated me to pursue the problems I proposed. I sincerely appreciate his critical insights, significant feedback, and teaching me the nuances of dedicated research. I will always hold his mentorship and professionalism in high regard.

I wish to extend my thanks to Prof. Abdel Lisser for his valuable discussions during my research visit to Université Paris-Saclay and for his academic guidance throughout.

I would like to express my thanks to my Student Research Committee (SRC) members, Prof. Aparna Mehra, Prof. S. Dharmaraja, and Prof. Arpan Chattopadhyay, for devoting their valuable time and providing insightful suggestions that helped me improve my research work. I am deeply indebted to Prof. Aparna Mehra for guiding me on multiple occasions and keeping the door open for me. I will continually find inspiration in her as an academic and as a person.

I acknowledge the financial assistance provided by CSIR for my PhD. I thank my seniors, Navnit, Sakshi, and Tanvi, for all the help, and I appreciate the support provided by other colleagues, staff, and faculty members of the Department of Mathematics, Indian Institute of Technology Delhi.

My heartfelt thanks go out to my sweet niece, Vishi, for keeping me entertained and bringing fresh air to the family, and my mum and dad for giving me this life and for believing in me;

they are the only two people waiting for my call at the end of the day, for which I will always feel blessed.

New Delhi V Varagapriya

Abstract

In this thesis, we consider a constrained Markov decision process (CMDP) where running costs and transition probabilities are uncertain. We model the uncertain running costs and transition probabilities using robust optimization, chance-constrained optimization, and distributionally robust chance-constrained optimization frameworks.

We start with a robust CMDP problem with known transition probabilities, and uncertain running cost vectors that are known to belong to an uncertainty set. We consider polytopic, ellipsoidal, and semidefinite cone uncertainty sets and equivalently reformulate the robust CMDP problem as a linear programming (LP), second-order cone programming (SOCP), and semidefinite programming (SDP) problem, respectively. As an application, we propose a variant of a machine replacement problem and perform numerical experiments on randomly generated instances of different sizes. Furthermore, we study a robust CMDP problem under uncertain transition probabilities and known running cost vectors. Under the assumption that the uncertainty is driven by a single parameter belonging to an interval, we equivalently reformulate the problem into a bilinear programming (BP) problem. By exploiting the structure of the BP problem, we construct an LP-based algorithm to find its global optimal solution. We propose a sufficient condition under which an optimal policy of the robust CMDP problem is unaffected by uncertainty. The numerical experiments are performed on a machine replacement problem and on randomly generated CMDP problems of various sizes using LP-based algorithm as well as Gurobi solver. We observe that in some cases, the LP-based algorithm outperforms the Gurobi solver. We extend this work to the case where the uncertainty in the transition probabilities is driven by multiple parameters belonging to a polytopic uncertainty set. Using the duality theory of LP problem, we equivalently reformulate the problem into a BP problem and perform numerical experiments on randomly generated CMDP problems of various sizes, using the Gurobi solver.

As an alternative approach, we use a chance-constrained optimization framework to address uncertainties in the running costs and transition probabilities. This leads to the formulation of a joint chance-constrained Markov decision process (JCCMDP). We first consider a JCCMDP problem where running cost vectors are defined using random vectors and transition probabilities are known. We assume that the random cost vectors are elliptically distributed and the dependence among the random constraints is driven by a Gumbel–Hougaard copula. Under these assumptions, we present two SOCP approximations such that their optimal values provide upper and lower bounds to the optimal cost of the JCCMDP problem. The numerical experiments are performed on a queueing control problem, a budget optimization problem in advertising platforms, and randomly generated CMDP problems of various sizes. We also consider the case when the distributions of running cost vectors are not known. In this case, we use classical probability inequalities, namely, one-sided Chebyshev, Bernstein, and Hoeffding inequalities, to construct convex programming problems, each of whose optimal value provides an upper bound to the optimal cost of the associated JCCMDP problem. We propose an LP problem whose optimal value provides a lower bound to the optimal cost of the associated JCCMDP problem. In addition, we study a JCCMDP problem where running costs are known and transition probabilities have random perturbations. Under certain conditions on the random perturbation vector, we construct convex programming problems using Bernstein and Hoeffding inequalities, each of whose optimal value gives an upper bound to the optimal cost of the associated JCCMDP problem. Similar to the case of random costs, we propose an LP problem whose optimal value provides a lower bound to the optimal cost of the associated JCCMDP problem. The numerical experiments are performed on a queueing control problem and randomly generated CMDP problems of various sizes.

At the end, we consider a distributionally robust chance-constrained optimization framework to address uncertainties in the running costs and transition probabilities. This approach is used when the exact probability distribution is not known and the optimization model considers the worst-case scenario of the underlying distribution. This leads to the formulation of a distributionally robust joint chance-constrained Markov decision process (DRJCCMDP). We first consider a DRJCCMDP problem under known transition probabilities and random running cost vectors whose distribution is only partially known. The only information we have about the distribution is that it belongs to an uncertainty set which is constructed using the full or partial information available on the first two moments. We consider three different moments-based uncertainty sets, and for each case, we present convex approximations whose optimal values provide upper and lower bounds to the optimal cost of the original problem. Furthermore, we

study a DRJCCMDP problem under random transition probabilities and known running costs. We assume that the transition probability vector follows a discrete distribution and has a finite support. Using the estimates of the first two moments, we construct two moments-based uncertainty sets similar to the case of random costs. We show that the DRJCCMDP problem can be equivalently reformulated either as a mixed integer bilinear programming (MIBP) problem or a mixed integer semidefinite programming (MISDP) problem with bilinear constraints. In both cases, we perform numerical experiments on randomly generated CMDP problems of various sizes.

सार

इस थीसिस में, हम एक सीमित मार्कोव निर्णय प्रक्रिया (CMDP) का विचार करते हैं जहां चलते लागतें और परिवर्तन संभावनाएं अनिश्चित हैं। हम अनिश्चित चलते लागतों और परिवर्तन संभावनाओं को स्थिर अनुकुलन, संयोजन-सीमित अनुकुलन, और वितरणात्मक रूप से स्थिर अनुकुलन के फ्रेमवर्क का उपयोग करके मॉडल करते हैं।

हम एक स्थिर CMDP समस्या के साथ शुरू करते हैं जिसमें ज्ञात परिवर्तन संभावनाएं होती हैं, और अनिश्चित चलते लागत वेक्टर होते हैं जो एक अनिश्चितता सेट में शामिल होने के लिए जाने जाते हैं। हम पॉलीटोपिक, अलिप्सोइडल, और सेमीडिफिनेट कोन अनिश्चितता सेट को ध्यान में रखते हुए और समरूपता के साथ स्थिर CMDP समस्या को बारीकी से पुनः रचित करते हैं, जैसा कि लीनियर प्रोग्रामिंग (LP), द्वितीयक कोन प्रोग्रामिंग (SOCP), और सेमीडिफिनेट प्रोग्रामिंग (SDP) समस्या के रूप में, क्रमशः। एक अनुप्रयोग के रूप में, हम एक मशीन प्रतिस्थापन समस्या का एक विविधांत प्रस्ताव प्रस्तुत करते हैं और विभिन्न आकारों के यादृच्छिक उत्पन्न मामलों पर संख्यात्मक प्रयोग आयोजित करते हैं। इसके अतिरिक्त, हम अनिश्चित परिवर्तन संभावनाओं के तहत एक स्थिर CMDP समस्या का अध्ययन करते हैं और ज्ञात चलते लागत वेक्टरों के तहत। एक ही पैरामीटर के अनिश्चितता से प्रेरित होने की परिगणना के तहत, हम समस्या को एक द्विचारी प्रोग्रामिंग (BP) समस्या में पुनः रचना करते हैं। BP समस्या की संरचना का उपयोग करके, हम इसका वैश्विक सर्वोत्तम समाधान ढूंढने के लिए एक LP-आधारित एल्गोरिदम निर्मित करते हैं। हम एक पर्याप्त स्थिति का प्रस्ताव करते हैं जिसके तहत स्थिर CMDP समस्या का एक श्रेष्ठ नीति अनिश्चितता से प्रभावित नहीं होती। संख्यात्मक प्रयोग एक मशीन प्रतिस्थापन समस्या पर और LP-आधारित एल्गोरिदम का उपयोग करके विभिन्न आकारों के यादृच्छिक CMDP समस्याओं पर होते हैं और Gurobi सॉल्वर का उपयोग करके। हम देखते हैं कि कुछ मामलों में, LP-आधारित एल्गोरिदम Gurobi सॉल्वर को परिपूर्ण करता है। हम इस काम को बढ़ाते हैं जहां संक्रमण संभावनाओं में अनिश्चितता कई पैरामीटर्स से प्रेरित है जो पॉलिटॉपिक अनिश्चितता सेट का हिस्सा है। LP समस्या के द्वारा द्वितीयता के सिद्धांत का उपयोग करते हुए, हम समस्या को बीपी समस्या में समान रूप से पुनः रचित करते हैं और विभिन्न आकारों की यादृच्छिक उत्पन्न CMDP समस्याओं पर संख्यात्मक प्रयोग करते हैं, Gurobi सॉल्वर का उपयोग करते हुए।

एक वैकल्पिक दृष्टिकोण के रूप में, हम चांस-संबंधित अनिश्चितताओं को संचालन लागतों और संक्रमण संभावनाओं में संबोधित करने के लिए एक अवसर-संयंत्रित अनुकूलन ढांचा का उपयोग करते हैं। इससे संयुक्त अवसर-संयंत्रित मार्कव निर्णय प्रक्रिया (JCCMDP) का निर्माण होता है। हम पहले एक JCCMDP समस्या का विचार करते हैं जहां संचालन

लागत वेक्टर यादृच्छिक वेक्टरों का उपयोग करके परिभाषित किए जाते हैं और संक्रमण संभावनाएं ज्ञात होती हैं। हम मानते हैं कि यादृच्छिक लागत वेक्टर अवृत्तियाँ हैं और यादृच्छिक प्रतिबंधों के बीच संबंध गंभीरदृष्टि–होगर्ड कोपुला के द्वारा प्रेरित हैं। इन मान्यताओं के अनुसार, हम दो SOCP अनुमापनों का प्रस्ताव प्रस्तुत करते हैं जिनके अद्वितीय मूल्य जोड़ी और निचले सीमा JCCMDP समस्या की श्रेणी की उत्कृष्ट लागत प्रदान करते हैं। संख्यात्मक प्रयोग एक क्यूइंग नियंत्रण समस्या पर, विज्ञापन प्लेटफार्म में बजट अनुकूलन समस्या पर, और विभिन्न आकारों की यादृच्छिक उत्पन्न CMDP समस्याओं पर किए जाते हैं। हम यह भी विचार करते हैं जब चलन लागत वेक्टरों का वितरण नहीं जाना जाता है। इस मामले में, हम क्लासिकल प्रायिकता असम्मिश्रणों, यानी, एक-तरफीय चेबीशेव, बर्नस्टीन, और होफ्डिंग असम्मिश्रणों का उपयोग करते हैं, ताकि प्रत्येक कॉनभेक्स प्रोग्रामिंग समस्या का निर्माण किया जा सके, जिसका प्रत्येक अद्वितीय मूल्य अपने उच्चतम मूल्य का आदान-प्रदान करता है। जुड़ी हुई JCCMDP समस्या की श्रेणी की उत्कृष्ट लागत के लिए। हम एक LP समस्या का प्रस्ताव करते हैं जिसका अद्वितीय मूल्य उस संबद्ध JCCMDP समस्या की न्यूनतम लागत का आदान-प्रदान करता है। साथ ही, हम एक JCCMDP समस्या का अध्ययन करते हैं जहां चलन लागतें ज्ञात होती हैं और संक्रमण संभावनाएं यादृच्छिक परिवर्तनों को होती हैं। यादृच्छिक परिवर्तन वेक्टर पर निश्चित शर्तों के तहत, हम बर्नस्टीन और होफ्डिंग असम्मिश्रणों का उपयोग करके कॉनभेक्स प्रोग्रामिंग समस्याओं का निर्माण करते हैं, जिसका प्रत्येक अद्वितीय मूल्य उस संबंधित JCCMDP समस्या की उच्चतम लागत का आदान-प्रदान करता है। यादृच्छिक लागतों के मामले की तरह, हम एक LP समस्या का प्रस्ताव करते हैं जिसका अद्वितीय मूल्य उस संबंधित JCCMDP समस्या की न्यूनतम लागत का आदान-प्रदान करता है। संख्यात्मक प्रयोग एक क्यूइंग नियंत्रण समस्या पर और विभिन्न आकारों की यादृच्छिक उत्पन्न CMDP समस्याओं पर किए जाते हैं।

अंत में, हम चलन लागतों और संक्रमण संभावनाओं में अनिश्चितताओं को संभालने के लिए एक वितर्णीय रोबस्ट अवसर-संबंधित अनुकूलन ढांचा का विचार करते हैं। यह दृष्टिकोण उपयोग किया जाता है जब सटीक प्रायिकता वितरण नहीं जाना जाता है और अनुकूलन मॉडल मूल वितरण के अत्यंत मामूली स्थिति का अत्यधिक मामूली स्थिति का मूल्यांकन करता है। इससे एक वितर्णीय रोबस्ट संयुक्त अवसर-संयंत्रित मार्कव निर्णय प्रक्रिया $(\mathrm{DRJCCMDP})$ का निर्माण होता है। हम पहले जानते हैं संक्रमण संभावनाओं और यादृच्छिक चलन लागत वेक्टरों के तहत एक $\mathrm{DRJCCMDP}$ समस्या का विचार करते हैं जिसका वितरण केवल आंशिक रूप से ज्ञात है। वितरण के बारे में हमारा केवल जानकारी है कि यह एक अनिश्चितता सेट में शामिल होता है जो पहले दो क्षणों पर उपलब्ध पूर्ण या आंशिक जानकारी का उपयोग करके निर्मित किया गया है। हम तीन विभिन्न क्षणों पर आधारित अनिश्चितता सेट का विचार करते हैं, और प्रत्येक मामले के लिए, हम उनका उपयुक्त मूल्य प्रस्तुत करते हैं जो मूल समस्या की उच्चतम लागत के लिए उच्चतम और निम्न सीमा प्रदान करते हैं। इसके अतिरिक्त, हम एक $\mathrm{DRJCCMDP}$ समस्या का अध्ययन करते हैं जहां यादृच्छिक संक्रमण संभावनाएं और जानी गई चलन लागतें होती हैं। हम मानते हैं कि संक्रमण संभावना वेक्टर एक विकेन्द्रित वितरण का पालन करता है और एक सीमित समर्थन रखता है। पहले दो क्षणों के अनुमानों का उपयोग करके, हम यहां भी दो क्षणों पर आधारित अनिश्चितता सेट निर्मित करते हैं, जो यादृच्छिक लागतों के मामले के समान हैं। हम दिखाते हैं कि $\mathrm{DRJCCMDP}$ समस्या को समकक्षीय रूप से पुनः रचित किया जा सकता है, या तो एक मिश्रित पूर्णांक बायलिनियर प्रोग्रामिंग (MIBP) समस्या या एक मिश्रित पूर्णांक अर्द्धवृत्तीय प्रोग्रामिंग (MISDP) समस्या बायलिनियर बाधाओं के साथ। दोनों मामलों में, हम विभिन्न आकारों की यादृच्छिक उत्पन्न CMDP समस्याओं पर संख्यात्मक प्रयोग करते हैं।

Contents

Certificate **i**

Acknowledgments **iii**

Abstract **v**

List of Figures **xv**

List of Tables **xvii**

1 Introduction **1**

1.1 Motivation and Literature survey 1

1.2 Our Contributions 6

2 Constrained Markov decision processes **11**

2.1 The model 11

2.2 Uncertain MDP/CMDP problems 18

3 Robust constrained Markov decision processes **21**

3.1 Introduction 21

3.2 Robust CMDP under uncertain costs 22

3.2.1 Convex reformulations 24

3.2.2 Numerical experiments: Machine replacement problem 27

3.3 Robust CMDP under uncertain transition probabilities 30

3.3.1 Uncertainty driven by a single parameter 31

3.3.1.1 Bilinear programming formulation 35

3.3.1.2 LP-based algorithm 36

3.3.1.3 Sufficient condition for the uncertainty free CMDP problem . . 41

3.3.1.4 Numerical experiments 45

3.3.2 Uncertainty driven by multiple parameters 52

3.3.2.1 Reformulation under uncertainty structure (GS1) 56

3.3.2.2 Reformulation under uncertainty structure (GS2) 61

3.3.2.3 Numerical experiments 65

3.4 Conclusions 68

3.5 Example 69

4 Chance-constrained Markov decision processes 71

4.1 Introduction . . . 71

4.2 Chance-constrained CMDP under random costs with known distribution . . . 72

4.2.1 Preliminaries . . . 74

4.2.2 Second order cone approximations . . . 76

4.2.2.1 Lower bound approximation . . . 80

4.2.2.2 Upper bound approximation . . . 81

4.2.3 Numerical experiments . . . 83

4.2.3.1 Queueing control problem and randomly generated CMDPs . . . 83

4.2.3.2 Budget optimization of advertisement model . . . 88

4.3 Chance-constrained CMDP under random costs with unknown distribution . . . 93

4.3.1 Approximation of a linear chance constraint . . . 93

4.3.1.1 Inner approximations . . . 93

4.3.1.2 Outer approximation . . . 94

4.3.2 CMDP under random costs . . . 96

4.3.2.1 Upper bound approximations . . . 97

4.3.2.2 Lower bound approximation . . . 102

4.3.2.3 Numerical experiments: Queueing control problem . . . 104

4.4 Chance-constrained CMDP under random transition probabilities with unknown distribution . . . 106

4.4.1 Upper bound approximations . . . 111

4.4.1.1 Upper bound approximation under Bernstein inequality . . . 115

4.4.1.2 Upper bound approximation under Hoeffding inequality . . . 116

4.4.2 Lower bound approximation . . . 117

4.4.3 Numerical experiments: Randomly generated CMDPs . . . 120

4.5 Conclusions . . . 125

5 Distributionally robust chance-constrained Markov decision processes 127

5.1 Introduction 127

5.2 Distributionally robust chance-constrained MDP under random costs 128

5.2.1 Lower bound approximation 131

5.2.2 Upper bound approximation 133

5.3 Distributionally robust chance-constrained MDP under random transition probabilities 134

5.4 Numerical experiments: Randomly generated CMDPs 143

5.4.1 Random running costs 143

5.4.2 Random transition probabilities 144

6 Conclusions and Future research 147

6.1 Thesis contribution 147

6.2 Future research 149

Bibliography 151

List of Publications 159

Curriculum Vitae 161

List of Figures

3.1 Machine replacement problem 27

3.2 Optimal policies for machine replacement problem 30

3.3 Time analysis for machine replacement problem 31

3.4 Machine replacement problem 46

3.5 Uncertain transition probabilities. 69

4.1 Optimal policies for queueing control problem 85

List of Tables

3.1 Equivalent optimization problems for particular cases of $\boldsymbol{K}$. 26

3.2 Optimal policies of a machine replacement problem. 28

3.3 Entries of matrices C and D_1. 29

3.4 Optimal values in random instances of a machine replacement problem. 47

3.4 Optimal values in random instances of a machine replacement problem. 48

3.4 Optimal values in random instances of a machine replacement problem. 49

3.5 Optimal values and CPU(s) in randomly generated CMDPs. 51

3.6 Optimal values and CPU(s) in randomly generated CMDPs. 67

3.6 Optimal values and CPU(s) in random instances. 68

4.1 List of some spherical distributions. 79

4.2 Queueing control problem: Discounted cost criterion. 86

4.3 Queueing control problem: Average cost criterion. 87

4.4 Randomly generated CMDPs: Discounted cost criterion. 88

4.5 Randomly generated CMDPs: Average cost criterion. 89

4.6 Optimal values and CPU(s) in budget optimization problem. 91

4.6 Optimal values and CPU(s) in budget optimization problem. 92

4.7 Optimal values and CPU(s) in queueing control problem. 104

4.7 Optimal values and CPU(s) in queueing control problem. 105

4.7 Optimal values and CPU(s) in queueing control problem. 106

4.8 Optimal values and CPU(s) in randomly generated CMDPs. 121

4.8 Optimal values and CPU(s) in randomly generated CMDPs. 122

4.8 Optimal values and CPU(s) in randomly generated CMDPs. 123

4.8 Optimal values and CPU(s) in randomly generated CMDPs. 124

5.1 Optimal values and CPU(s) in randomly generated CMDPs. 143

5.1 Optimal values and CPU(s) in randomly generated CMDPs. 144

5.2 Optimal values and CPU(s) in randomly generated CMDPs. 145