

**DECISION-THEORETIC PLANNING  
UNDER RISK-SENSITIVE PLANNING OBJECTIVES**

A Dissertation  
Presented to  
The Academic Faculty

by

**Yaxin Liu**  
**刘亚新**

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

College of Computing  
Georgia Institute of Technology  
April 2005

Copyright © 2005 by Yaxin Liu

# DECISION-THEORETIC PLANNING UNDER RISK-SENSITIVE PLANNING OBJECTIVES

Approved by:

Dr. Sven Koenig, Co-advisor  
College of Computing  
*Georgia Institute of Technology*

Dr. Frank Dellaert  
College of Computing  
*Georgia Institute of Technology*

Dr. Craig Tovey, Co-advisor  
College of Computing  
*Georgia Institute of Technology*

Dr. Ashok Goel  
College of Computing  
*Georgia Institute of Technology*

Dr. Anton Kleywegt  
School of Industrial and Systems Engineering  
*Georgia Institute of Technology*

Dr. Richard Goodwin  
*IBM T.J. Watson Research Center*

Date Approved: April 11, 2005

*To Zhijie and Emily.*

*Hofstadter's Law: It always takes longer than you think, even when you take into account  
Hofstadter's Law.*

— *Douglas R. Hofstadter*

## ACKNOWLEDGEMENTS

Many thanks to my advisor, Sven Koenig, for helping me through the process. Thanks to my co-advisor, Craig Tovey, for many encouraging conversations. Thanks to Richard Goodwin, for his guidance of my summer internships at IBM and for shaping the direction of this work. Thanks to Anton Kleywegt, for his encouragement and support from early on. Thanks to Frank Dellaert, for his advice on writing. And thanks to Ashok Goel, for helping in various ways.

Thanks to my fellow students, David Furcy, for long-term encouragement and collaboration; to Jimmy Davies, for moral support; to Chip Mappus, for much help with English; and to Patrick Yaner, for his bookshelf.

Thanks to Peter Stone, for allowing me to finish the thesis at UT Austin.

Special thanks to my friends Yuan Chen and Xiaoling Cao, for helping taking care of Emily.

I am forever in debt to Zhijie for the many years of enduring, and to Emily for not-so-many years. Lastly and not leastly, I appreciate the help and support from my parents and parents-in-law.

This work is partially supported by an IBM Fellowship.

# TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> . . . . .	<b>v</b>
<b>LIST OF TABLES</b> . . . . .	<b>x</b>
<b>LIST OF FIGURES</b> . . . . .	<b>xi</b>
<b>LIST OF ALGORITHMS</b> . . . . .	<b>xiv</b>
<b>SUMMARY</b> . . . . .	<b>xv</b>
<b>CHAPTER I INTRODUCTION</b> . . . . .	<b>1</b>
1.1 Risk Attitudes . . . . .	2
1.2 AI Planning and Two Examples . . . . .	5
1.3 Decision-Theoretical Planning . . . . .	9
1.4 Risk-Sensitive Planning . . . . .	14
1.5 Scope of Research . . . . .	17
1.6 Contributions . . . . .	18
1.6.1 Exponential Utility Functions . . . . .	19
1.6.2 General Risk-Sensitive Utility Functions . . . . .	21
1.6.3 Problems with Arbitrary Rewards . . . . .	22
1.6.4 Summary . . . . .	23
<b>CHAPTER II BACKGROUND</b> . . . . .	<b>25</b>
2.1 Utility Theory . . . . .	25
2.1.1 Principle of Maximization of Expected Utility . . . . .	26
2.1.2 Utility Functions on the Real Line and Risk Attitudes . . . . .	30
2.2 Markov Decision Process Models . . . . .	33
2.2.1 Definitions . . . . .	33
2.2.2 Planning Objectives . . . . .	38
2.2.3 Basic Problems Concerning Planning with MDPs . . . . .	46
2.3 Risk-Neutral Planning Objectives and Results . . . . .	48
2.3.1 Finite Horizon . . . . .	49
2.3.2 Infinite Horizon with Discounting . . . . .	51
2.3.3 Infinite Horizon without Discounting . . . . .	53

2.3.4	Semi-MDPs . . . . .	60
2.3.5	Summary . . . . .	61
2.4	Related Work: Risk-Sensitive Planning . . . . .	62
2.4.1	MDPs with Exponential Utility Functions . . . . .	62
2.4.2	MDPs with Target-Level Utility Functions . . . . .	67
2.4.3	MDPs with General Risk-Sensitive Utility Functions . . . . .	69
2.4.4	MDPs with Even More General Utility Functions . . . . .	71
2.5	Related Work: Large-Scale MDPs under Risk-Neutral Objectives . . . . .	73
2.5.1	Symbolic Strategies . . . . .	73
2.5.2	Numerical Strategies . . . . .	77
<b>CHAPTER III EXPONENTIAL UTILITY FUNCTIONS . . . . .</b>		<b>79</b>
3.1	Exponential Utility Functions . . . . .	80
3.2	An Overview of Basic Properties . . . . .	82
3.2.1	Problems without Explicit Goal States . . . . .	83
3.2.2	Problems with Explicit Goal States . . . . .	86
3.2.3	Online Testing of Finiteness Conditions . . . . .	87
3.3	The Representation Transformation . . . . .	101
3.3.1	Definition . . . . .	101
3.3.2	Planning Problems with Proper Policies Only . . . . .	106
3.3.3	Planning Problems with Improper Policies . . . . .	107
3.3.4	Advantages and Limitations of the Representation Transformation . . . . .	108
3.4	Transformation of Algorithms . . . . .	108
3.4.1	Pseudo-Probability Transformation . . . . .	109
3.4.2	Pseudo-Discount Factor Transformation . . . . .	110
3.4.3	Short Summary . . . . .	111
3.5	Explicit Probabilities: The LAO* Example . . . . .	112
3.5.1	LAO* Search . . . . .	113
3.5.2	Risk-Neutral LAO* . . . . .	119
3.5.3	Risk-Sensitive LAO* . . . . .	120
3.6	Pseudo-Probability Transformation and Temporally Extended Probabilities . . . . .	123
3.6.1	Pseudo-Probabilities from Action Sequences . . . . .	124

3.6.2	Pseudo-Probabilities from Simple Options . . . . .	142
3.6.3	Options as Temporally Extended Actions . . . . .	156
3.6.4	Pseudo-Discount Factor Transformation? . . . . .	161
3.7	Pseudo-Discount Factor Transformation and Factored Probabilities . . . . .	162
3.7.1	Factored Representation of MDPs . . . . .	162
3.7.2	SPUDD . . . . .	175
3.7.3	Risk-Sensitive Symbolic LAO* . . . . .	188
3.8	Summary . . . . .	193
<b>CHAPTER IV GENERAL RISK-SENSITIVE UTILITY FUNCTIONS</b>		<b>195</b>
4.1	Introduction . . . . .	196
4.2	Finite Horizon . . . . .	197
4.2.1	Basic Properties . . . . .	197
4.2.2	Augmenting the State Space . . . . .	199
4.2.3	Basic Properties Revisited . . . . .	206
4.2.4	Functions as Values . . . . .	208
4.2.5	Approximations . . . . .	209
4.2.6	Piecewise Linear Approximations . . . . .	213
4.3	Infinite Horizon . . . . .	217
4.3.1	The Augmented Model . . . . .	217
4.3.2	Existence and Finiteness of Optimal Values . . . . .	219
4.3.3	Existence of Optimal Policies . . . . .	222
4.3.4	Value Iteration and Approximation . . . . .	223
4.3.5	Example: Deadline Utility Functions . . . . .	238
4.4	One-Switch Utility Functions . . . . .	252
4.4.1	Finite Horizon . . . . .	254
4.4.2	Infinite Horizon: Limiting Cases . . . . .	256
4.4.3	Infinite Horizon: Negative Models . . . . .	257
4.4.4	Infinite Horizon: Positive Models . . . . .	266
4.5	Summary . . . . .	273

<b>CHAPTER V</b>	<b>PROBLEMS WITH ARBITRARY REWARDS . . . . .</b>	<b>275</b>
5.1	Results for the MER Objective . . . . .	275
5.1.1	Structural Implications . . . . .	278
5.2	A Template for the Proofs of Existence Results . . . . .	281
5.3	Exponential Utility Functions . . . . .	285
5.3.1	Concave Exponential Utility Functions . . . . .	295
5.3.2	Convex Exponential Utility Functions . . . . .	300
5.4	General Risk-Sensitive Utility Functions . . . . .	302
5.4.1	Bounded Utility Functions . . . . .	302
5.4.2	Linearly Bounded Utility Functions . . . . .	304
5.4.3	Exponentially Bounded Utility Functions . . . . .	308
5.4.4	Bounded Total Rewards and Arbitrary Utility Functions . . . . .	312
5.5	Summary . . . . .	315
<b>CHAPTER VI</b>	<b>CONCLUSION AND FUTURE WORK . . . . .</b>	<b>316</b>
6.1	Summary . . . . .	316
6.2	Future Work: Short Term . . . . .	317
6.2.1	Theoretical Results . . . . .	317
6.2.2	Solution Methods . . . . .	318
6.2.3	Extensions . . . . .	318
6.3	Future Work: Long Term . . . . .	319
<b>REFERENCES</b>	<b>. . . . .</b>	<b>320</b>

## LIST OF TABLES

1.1	An illustrative example for risk attitudes . . . . .	2
1.2	Summary of existing research and contributions . . . . .	19
2.1	Notation for planning objectives, values, and policies . . . . .	46
2.2	Results for finite models with risk-neutral planning objectives . . . . .	61
3.1	Trace of LAO* for the painted-blocks problem . . . . .	118
3.2	Optimal risk-sensitive sensing policies . . . . .	140
3.3	Planning complexity . . . . .	141
3.4	Transition probabilities and rewards under the MER objective . . . . .	146
3.5	Pseudo-probabilities and rewards under the $MEU_{exp}$ objective . . . . .	151
3.6	CPTs for <code>move_X_from_Y_to_Z</code> with the <code>Outcome</code> node . . . . .	169
3.7	CPTs for <code>move_X_from_Y_to_Z</code> without the <code>Outcome</code> node . . . . .	173
4.1	Optimal values (probabilities of success) with hard deadlines . . . . .	240
4.2	Breakpoints of the value functions from Figure 4.12 . . . . .	243
4.3	Breakpoints of the value functions from Figure 4.17 . . . . .	246
4.4	Breakpoints of the value functions from Figure 4.22 . . . . .	252
4.5	Segments of the value functions from Figure 4.25 . . . . .	265

## LIST OF FIGURES

1.1	Interpretation of the lottery example using a concave exponential utility function	4
1.2	A grid-world robot navigation example . . . . .	5
1.3	Action models . . . . .	6
1.4	Deterministic and contingent plans . . . . .	7
1.5	Two possible trajectories under the contingent plan . . . . .	8
1.6	A painted blocks problem . . . . .	9
1.7	A probabilistic action model . . . . .	9
1.8	Simulated frequencies of state visits for the contingent plan . . . . .	10
1.9	Simulated frequencies of total rewards for the contingent plan . . . . .	10
1.10	An optimal plan under the MER objective . . . . .	13
1.11	Optimal plans under the MEU objective with exponential utility functions . . .	15
2.1	The relationships among different classes of policies . . . . .	37
2.2	An example for the (non)-existence of values . . . . .	41
2.3	An example for the finiteness of values . . . . .	42
3.1	The “nominal transformation of algorithms” approach . . . . .	80
3.2	The “representation transformation” approach . . . . .	80
3.3	Exponential utility functions . . . . .	82
3.4	An example MDP with infinite values . . . . .	90
3.5	Backward and forward searches in LAO* . . . . .	114
3.6	An optimal plan obtained by LAO* under the MER objective . . . . .	119
3.7	Optimal risk-sensitive plans for the painted-blocks problem . . . . .	123
3.8	Segments of two possible sensing plans . . . . .	128
3.9	State-visit frequencies of sensing plans . . . . .	140
3.10	A simple option for traversing the top horizontal path . . . . .	143
3.11	Available options . . . . .	147
3.12	An “optimal” policy under the MER objective using simple options only . . .	148
3.13	“Optimal” policies under the $MEU_{exp}$ objective using simple options only . . .	152
3.14	Options obtained from pre-options using hierarchical dynamic programming . .	155

3.15	An “optimal” policy under the MER objective using hierarchical dynamic programming . . . . .	156
3.16	Options obtained from pre-options for a risk-averse agent using hierarchical dynamic programming . . . . .	157
3.17	An “optimal” policy under the $MEU_{exp}$ objective for a risk-averse agent using hierarchical dynamic programming . . . . .	158
3.18	Options obtained from pre-options for a risk-seeking agent using hierarchical dynamic programming . . . . .	159
3.19	An “optimal” policy under the $MEU_{exp}$ objective for a risk-seeking agent using hierarchical dynamic programming . . . . .	160
3.20	2-time dynamic Bayesian networks . . . . .	163
3.21	The named version of the painted-blocks problem from Figure 1.6 . . . . .	167
3.22	Actions for the painted-blocks problem . . . . .	167
3.23	Convert <code>move_X_from_Y_to_Z</code> from a PSO representation to a 2TBN representation	169
3.24	2TBN representation for actions of the painted-blocks domain . . . . .	174
3.25	The ADD representation of factored CPTs from Figure 3.20(c) and the reward function . . . . .	176
3.26	The ADD representation of factored CPTs for action <code>move_A_from_B_to_C</code> . . .	177
3.27	Optimal plans for the factored version of painted-blocks problem . . . . .	192
4.1	The sandwich method . . . . .	215
4.2	Lower PWL approximation for asymptotically constant utility functions . . . .	226
4.3	Upper PWL approximation for asymptotically constant utility functions . . . .	228
4.4	Lower PWL approximation for asymptotically linear utility functions . . . . .	230
4.5	Upper PWL approximation for asymptotically linear utility functions . . . . .	233
4.6	A painted-blocks problem for general risk-sensitive utility functions . . . . .	239
4.7	A hard deadline utility function and its modification for value iteration . . . .	239
4.8	Optimal functional value functions of relevant states for the hard deadline utility function . . . . .	240
4.9	An SD-optimal policy for the augmented model with the hard deadline utility function . . . . .	241
4.10	The aSD-optimal policy for the original model with the hard deadline utility function . . . . .	241
4.11	Linearly soft deadline utility function . . . . .	242

4.12	Optimal functional value functions of relevant states for the linearly soft deadline utility function . . . . .	243
4.13	An SD-optimal policy for the augmented model with the linearly soft deadline utility function . . . . .	244
4.14	The aSD-optimal policy for the original model with the linearly soft deadline utility function . . . . .	244
4.15	The SD-optimal policy for the original model with a linear utility function . . .	245
4.16	Exponentially soft deadline utility function . . . . .	245
4.17	Optimal functional value functions of relevant states for the exponentially soft deadline utility function . . . . .	246
4.18	An SD-optimal policy for the augmented model with the exponentially soft deadline utility function . . . . .	247
4.19	The aSD-optimal policy for the original model with the exponentially soft deadline utility function . . . . .	247
4.20	The SD-optimal policy for the original model with an exponential utility function	248
4.21	Mixed soft deadline utility function . . . . .	249
4.22	Optimal functional value functions of relevant states for the mixed soft deadline utility function . . . . .	250
4.23	An SD-optimal policy for the augmented model with the mixed soft deadline utility function . . . . .	251
4.24	The aSD-optimal policy for the original model with the mixed soft deadline utility function . . . . .	251
4.25	Optimal functional value functions of relevant states for the one-switch utility function . . . . .	265
4.26	An aSD-optimal policy for the one-switch utility function . . . . .	266
5.1	The example MDP from Figure 2.2 and its positive and negative parts . . . . .	276
5.2	An example MDP with well-defined values but excluded by Condition 5.1 . . .	276
5.3	Example MDPs that illustrate Lemma 5.2 . . . . .	280
5.4	MDPs satisfying Condition 5.1 but without expected utilities . . . . .	287

## LIST OF ALGORITHMS

2.1	Backward Induction for Finite Models under the $MER_T$ Objective . . . . .	50
2.2	Value Iteration Procedure for Finite Models under the $MER_\beta$ Objective . . . . .	52
2.3	Policy Iteration Procedure for Finite Models under the $MER_\beta$ Objective . . . . .	53
3.1	LAO* with Policy Iteration . . . . .	116
3.2	Forward and Backward Search Procedures in LAO* . . . . .	117
3.3	Policy Iteration with a Restricted Domain under the MER Objective . . . . .	119
3.4	Policy Iteration with a Restricted Domain under the $MEU_{exp}$ Objective . . . . .	121
3.5	Policy Iteration with a Restricted Domain under the $MER_\beta$ Objective . . . . .	121
3.6	Policy Improvement Using Heuristic Search under the MER Objective . . . . .	132
3.7	Policy Iteration for MDPwO under the MER Objective . . . . .	133
3.8	Policy Improvement Using Heuristic Search under the $MEU_{exp}$ Objective . . . . .	135
3.9	Policy Iteration for MDPwO under the $MEU_{exp}$ Objective . . . . .	136
3.10	SPUDD under the MER Objective . . . . .	182
3.11	SPUDD under the $MEU_{exp}$ Objective . . . . .	186
3.12	LAO* for Factored MDPs Using SPUDD . . . . .	189
3.13	Search in LAO* Using BDDs . . . . .	190
3.14	SPUDDExp with a Restricted Domain for LAO*Exp . . . . .	191
4.1	Backward Induction under the $MEU_T$ Objective . . . . .	207
4.2	Backward Induction under the $MEU_T$ Objective: Functional Form . . . . .	209
4.3	Value Iteration under the MEU Objective: Functional Form . . . . .	224
4.4	Backward Induction for Negative Models under the $MEU_{one}$ Objective . . . . .	262
4.5	Calculate and Insert the Switching Point for Negative Models . . . . .	263
4.6	Customized Insertion Operation for Algorithm 4.4 . . . . .	264
4.7	Backward Induction for Positive Models under the $MEU_{one}$ Objective . . . . .	271
4.8	Calculate and Insert the Switching Point for Positive Models . . . . .	272
4.9	Customized Insertion Operation for Algorithm 4.7 . . . . .	273

## SUMMARY

Risk attitudes crucially affect human decision making preferences, especially in scenarios where huge wins or losses are possible, as exemplified by planetary rover navigation, oilspill response, and business applications. Decision-theoretic planners therefore need to take risk aspects into account to serve their users better. However, most existing decision-theoretic planners use simplistic planning objectives that are risk-neutral. The thesis research is the first comprehensive study of how to incorporate risk attitudes into decision-theoretic planners and solve large-scale planning problems represented as Markov decision process models. The thesis consists of three parts.

The first part of the thesis studies risk-sensitive planning in the case where exponential utility functions are used to model risk attitudes. In this case, there exists an optimal plan that maps states to actions. I show that existing decision-theoretic planners can be transformed to take risk attitudes into account. My approach is more general than previous approaches, which transform the planning tasks rather than the planning algorithms. The transformed algorithms bear visual resemblance to the original algorithms but special treatment may be needed to ensure their validity. Moreover, different transformations are needed if the transition probabilities are implicitly given, namely, temporally extended probabilities and probabilities given in a factored form. I show how the transformations and their variants can be applied to various decision-theoretic planners.

The second part of the thesis studies risk-sensitive planning in the case where general nonlinear utility functions are used to model risk attitudes. In this case, there does not in general exist an optimal plan that maps states to actions, and an optimal plan must take into consideration the accumulated rewards starting from the initial state. I show that a state-augmentation approach can be used to reduce a risk-sensitive planning problem to a risk-neutral planning problem with an augmented state space. I further use a functional

interpretation of value functions and approximation methods to solve the planning problems efficiently with value iteration. I also develop an exact method for solving risk-sensitive planning problems where one-switch utility functions are used to model risk attitudes.

The third part of the thesis studies risk sensitive planning in case where arbitrary rewards are used. In this case, many of the basic properties are unknown, including the existence and finiteness of optimal expected utilities. I propose a spectrum of conditions that can be used to constrain the utility function and the planning problem so that the optimal expected utilities exist and are finite. These conditions are the basis for the further development of computational procedures. I prove that the existence and finiteness properties hold for stationary plans, where the action to perform in each state does not change over time, under different sets of conditions.

I use two running examples to demonstrate that risk-sensitive planners can be easily created from their risk-neutral counterparts, and the resulting optimal plans are qualitatively different from optimal plans under a risk-neutral planning objective.