CHAPTER VI

CONCLUSION AND FUTURE WORK

6.1 Summary

In this thesis, I studied risk-sensitive planning objectives in decision-theoretical planning. Following the MEU principle, risk-sensitive planning objectives incorporate the risk attitudes of human decision makers into planning, and thus overcome some of the shortcomings of risk-neutral planning objectives, which are the focus of current decision-theoretic planning research. Instead of starting from scratch, ideas from decision-theoretic planning under the MER objective can be adapted to decision-theoretic planning under risk-sensitive planning objectives. Using Markov decision processes as the formal model, I established theoretical properties and developed efficient algorithms that make this possible. I demonstrated that optimal plans under a risk-sensitive objective are different from those under a risk-neutral objective, and that such plans can be constructed efficiently.

Chapter 3 studied risk-sensitive planning with exponential utility functions, which model constant risk attitudes. I showed that existing decision-theoretic planners can be transformed to take into account constant risk attitudes. The transformed algorithms bear visual resemblance to the original algorithms but special conditions are needed to ensure their validity. Moreover, different versions of the transformation are needed if the transition probabilities are implicitly given: the pseudo-probability transformation for temporally extended probabilities and the pseudo-discount factor transformation for probabilities given in a factored form. I showed that the main symbolic strategies in decision-theoretic planning for solving large-scale planning problems can still be used for risk-sensitive planning with constant risk attitudes, and the risk-sensitive versions of algorithms using such strategies can be obtained using the transformation and its variants to transform their risk-neutral counterparts.

Chapter 4 studied risk-sensitive planning with more general utility functions, which model variable risk attitudes. Using a state-augmentation approach, I showed that a functional interpretation of value functions and piecewise linear approximation methods can be used to solve planning problems efficiently, based on backward induction (for finite planning horizons) and value iteration (for infinite planning horizons and asymptotically constant, linear, or exponential utility functions). For one-switch utility functions, I also obtained an exact method similar to backward induction, based on results for planning with general, exponential, and linear utility functions.

Chapter 5 studied risk-sensitive planning with arbitrary rewards, while Chapter 3 and Chapter 4 considered risk-sensitive planning with negative and positive models. For problems with arbitrary rewards, the theoretical foundation is incomplete. In this chapter, I proposed different sets of conditions that form a spectrum and showed that under these conditions, the values of stationary policies exist and are finite. I also conjectured that under these conditions, the values of all policies and thus the optimal values exist and are finite.

6.2 Future Work: Short Term

This thesis developed theories and algorithms for risk-sensitive planning. The treatment is, however, far from complete. Here we list some possibilities for future research.

6.2.1 Theoretical Results

The convergence rates of the algorithms we presented in this thesis, especially those based on value iteration, are unknown. Although we know that these methods converge to the optimal values or an optimal policy, and even that methods for exponential utility functions converge at a geometric rate, it is not clear how close the results are to the optimal ones. Recent results for the MER objective (Bonet, 2002) suggest that the convergence rate can be estimated by examining how the values change over time.

For planning with arbitrary rewards, we need to investigate the conjectures for the existence and finiteness of the optimal values. The concepts of recurrent and transient states need to be extended for the induced random processes under a general HR-policy.

6.2.2 Solution Methods

In Chapter 3, we applied the transformation approach to methods using symbolic strategies such as heuristic search, temporal abstraction, and state abstraction. Another class of methods for large-scale problems have numerical strategies such as function approximators and direct policy search. It is worthwhile considering methods that use numerical strategies, since they can be more efficient for some types of problems than symbolic strategies. The main challenge is to show under what conditions such methods converge.

In Chapter 4, we considered value iteration using functional value functions. I anticipate that functional value functions can also be used in methods for large-scale problems with a symbolic strategy, therefore extending the applicability of the state-augmentation approach. The main difficulty is how to manage the complexity of functional value functions.

6.2.3 Extensions

Utility functions are elicited from human decision makers to be used by autonomous agents acting on their behalf. It is often the case that we can determine the form of utility functions, but only with imprecise parameters. In this case, a research topic is to study the sensitivity of optimal plans with respect to these imprecise parameters, and to develop methods for planning with imprecise utility functions.

Reinforcement learning is the problem for an agent to learn how to act by acting in the environment and receiving feedbacks. Using MDPs as the formal model, reinforcement learning methods can be viewed as online versions of planning methods, and do not assume that the dynamics of the agent's interaction with its environment is known. On the other hand, the current state is often only partially known in real-world planning problems. Such problems are captured by partially observable MDP (POMDP) models. Extensions for reinforcement learning and POMDPs under MEU objectives therefore constitute one more step towards solving real-world planning problems.

6.3 Future Work: Long Term

In the long term, I am interested in building autonomous agents that are able to act intelligently and provide valuable services to people in a complex environment involving uncertainty while taking into account the preference structures of their human users. Besides risk attitudes, there are other aspects that affect the preference structures of human users, such as the tradeoff among multiple types of rewards, and the requirement of achieving goals.

In real applications, the decision maker often trades off different resources, such as product quality and cost in business situations, or energy and time for robot navigation on Mars. Executing actions then results in reward tuples. In this case, preference structures can usually be captured by multi-attribute utility functions, and the planning objective is to maximize the expected multi-attribute utility.

Many planning problems require the agent to achieve some desired goal, which can be predefined goal states or temporally extended goals. Goals are very important in classical AI planning but not in decision-theoretic planning, since the latter uses rewards exclusively in its planning objectives. Although in some problems, the rewards can be assigned so that a plan maximizing the expected reward (or utility) also achieves the goal. But this is not the case in general. On the other hand, people often prefer a plan that guarantees goal achievement if only a small amount of the expected reward (or utility) is sacrificed. Therefore we need a more thorough study of planning tasks that require goal achievements.

All these aspects, and likely more, need to be taken into consideration when building usable planning systems including decision support systems for real-world applications such as e-commerce or planetary rovers. This thesis takes a first step towards building such planning systems by integrating ideas from artificial intelligence planning, operations research, and utility theory. Our results showed that it is promising that complex human preference structures, such as risk attitudes, can be incorporated into decision-theoretic planning without compromising too much of their efficiency.

REFERENCES

Ávila-Godoy, Guadalupe; Brau, Agustin; and Fernández-Gaucherand, Emmanuel, 1997. Controlled Markov chains with discounted risk-sensitive criteria: Applications to machine replacement. In *Proceedings of the 36th Conference on Decision and Control (CDC-97)*.

Ávila-Godoy, Micaela Guadalupe, 1999. *Controlled Markov Chains with Exponential Risk-Sensitive Criteria: Modularity, Structured Policies and Applications*. Ph.D. thesis, Department of Mathematics, University of Arizona.

Bahar, R. Iris; Frohm, Erica A.; Gaona, Charles M.; Hachtel, Gary D.; Macii, Enrico; Pardo, Abelardo; and Somenzi, Fabio, 1993. Algebraic decision diagrams and their applications. In *Proceedings of the 1993 IEEE/ACM International Conference on Computer-Aided Design (ICCAD-93)*, pages 188–191.

Barberá, Salvador; Hammond, Peter J.; and Seidl, Christian, eds., 1998. *Handbook of Utility Theory, Volume 1: Principles*. Kluwer Academic Publishers.

Barto, Andrew G.; Bradtke, Steven J.; and Singh, Satinder P., 1995. Learning to act using real-time dynamic programming. *Artificial Intelligence*, 72:81–138.

Baxter, Jonathan and Bartlett, Peter L., 2001. Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:319–350.

Baxter, Jonathan; Bartlett, Peter L.; and Weaver, Lex, 2001. Experiments with infinite-horizon, policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:351–381.

Bell, David E., 1988. One-switch utility functions and a measure of risk. *Management Science*, 34(12):1416–1424.

Bell, David E. and Fishburn, Peter C., 2001. Strong one-switch utility. *Management Science*, 47(4):601–604.

Bell, David E.; Raiffa, Howard; and Tversky, Amos, eds., 1988. *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press.

Bertsekas, Dimitri P., 2001. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, 2nd edition.

Bertsekas, Dimitri P. and Tsitsiklis, John N., 1991. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595.

Bertsekas, Dimitri P. and Tsitsiklis, John N., 1996. *Neuro-Dynamic Programming*. Athena Scientific.

Blythe, Jim, 1997. *Planning under Uncertainty in Dynamic Domains*. Ph.D. thesis, School of Computer Science, Carnegie Mellon University.

Blythe, Jim, 1999. Decision-theoretic planning. *AI Magazine*, 20(2):37–54.

Bonet, Blai, 2002. Stochastic shortest-path problems and real-time DP: New theoretical results. Available from the Web.

Bonet, Blai and Geffner, Héctor, 2001. Heuristic search planner: 2.0. *AI Magazine*, 22(3):77–80.

Bonet, Blai and Geffner, Héctor, 2002. Solving stochastic shortest-path problems with RTDP. Technical report, Department of Computer Science, University of California at Los Angeles.

Bonet, Blai and Geffner, Héctor, 2003a. Faster heuristic search algorithms for planning with uncertainty and full feedback. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1233–1238.

Bonet, Blai and Geffner, Héctor, 2003b. Labeled RTDP: Improving the convergence of real-time dynamic programming. In *Proceedings of the Thirteenth International Conference on Automated Planning and Scheduling (ICAPS-03)*, pages 12–21.

Bouakiz, Mokrane, 1985. *Risk-Sensitivity in Stochastic Optimization with Applications*. Ph.D. thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology.

Bouakiz, Mokrane and Kebir, Youcef, 1995. Target-level criterion in Markov decision processes. *Journal of Optimization Theory and Applications*, 86(1):1–15.

Boutilier, Craig, 1997. Correlated action effects in decision theoretic regression. In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 30–37.

Boutilier, Craig; Dean, Thomas; and Hanks, Steve, 1999. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94.

Boutilier, Craig and Dearden, Richard, 1996. Approximating value trees in structured dynamic programming. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96)*.

Boutilier, Craig; Dearden, Richard; and Goldszmidt, Moisés, 1995. Exploiting structure in policy construction. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*.

Boutilier, Craig; Dearden, Richard; and Goldszmidt, Moisés, 2000. Stochastic dynamic programming with factored representations. *Artificial Intelligence*, 121:49–107.

Brau-Rojas, Agustin, 1999. *Controlled Markov Chains with Risk-Sensitive Average Cost Criterion*. Ph.D. thesis, Department of Mathematics, University of Arizona.

Cavazos-Cadena, Rolando and Fernández-Gaucherand, Emmanuel, 1999. Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions. *Mathematics Methods of Operations Research*, 49:299–324.

Cavazos-Cadena, Rolando and Montes-de-Oca, Raúl, 2000a. Nearly optimal policies in risk-sensitive positive dynamic programming on discrete spaces. *Mathematics Methods of Operations Research*, 52:133–167.

Cavazos-Cadena, Rolando and Montes-de-Oca, Raúl, 2000b. Optimal stationary policies in risk-sensitive dynamic programs with finite state space and nonnegative rewards. *Applicationes Mathematicae*, 27(2):167–185.

Chawla, Jay P., 2000. *Optimal Risk Sensitive Control of Semi-Markov Decision Processes*. Ph.D. thesis, Institute of Systems Research, University of Maryland at College Park.

Chung, Kun-Jen and Sobel, Matthew J., 1987. Discounted MDP's: Distribution functions and exponential utility maximization. *SIAM Journal of Control and Optimization*, 35(1):49–62.

Cohn, Paul R.; Greenberg, Michael L.; Hart, David M.; and Howe, Adele E., 1989. Trial by fire: Understanding the design requirements for agents in complex environments. *AI Magazine*, 10(3):32–48.

Coraluppi, Stefano P. and Marcus, Steven I., 1996. Risk-sensitive control of Markov decision processes. In *Proceedings of the 30th Annual Conference on Information Sciences and Systems (CISS-96)*, pages 934–939.

Corman, Thomas H.; Leiserson, Charles E.; and Rivest, Ronald L., 1990. *Introduction to Algorithms*. The MIT Press.

Corner, James L. and Corner, Patricia D., 1995. Characteristics of decisions in decision analysis practice. *The Journal of Operational Research Society*, 46:304–314.

Dean, Thomas; Kaelbling, Leslie Pack; Kirman, Jak; and Nicholson, Ann, 1995. Planning under time constraints in stochastic domains. *Artificial Intelligence*, 76(1-2):35–74.

Dean, Thomas and Kanazawa, Keiji, 1989. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150.

Dean, Thomas and Lin, Shieu-Hong, 1995. Decomposition techniques for planning in stochastic domains. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*.

Dechter, Rina, 1996. Bucket elimination: A unifying framework for probabilistic inference. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI-96)*, pages 211–219.

Denardo, Eric V. and Rothblum, Uriel G., 1979. Optimal stopping, exponential utility, and linear programming. *Mathematical Programming*, 16:228–244.

di Masi, Giovanni B. and Stettner, Lukasz, 1999. Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM Journal of Control and Optimization*, 38(1):61–78.

Dietterich, Thomas G., 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13:227–303.

Dolgov, Dmitri and Durfee, Edmund, 2004. Approximate probabilistic constraints and risk-sensitive optimization criteria in Markov decision processes. In *Proceedings of the Eighth International Symposium on Artificial Intelligence and Mathematics*. Fort Lauderdale, FL.

Eagle, James Norfleet, II, 1975. *A Utility Criterion for the Markov Decision Process*. Ph.D. thesis, Department of Engineering-Economic Systems,Stanford University.

Fainberg [sic], Eugene A., 1982. Controlled Markov processed with arbitrary numerical criteria. *Theory of Probability and its Applications*, 27(3):486–503.

Farquhar, Peter H. and Nakamura, Yutaka, 1987. Constant exchange risk properties. *Operations Research*, 35(2):206–214.

Feinberg, Eugene A., 2002. Total reward criteria. In Eugene A. Feinberg and Adam Shwartz, eds., *Handbook of Markov Decision Processes: Methods and Applications*, chapter 6, pages 173–208. Kluwer.

Feinberg, Eugene A. and Shwartz, Adam, eds., 2002. *Handbook of Markov Decision Processes: Methods and Applications*. Kluwer.

Feng, Zhengzhu and Hansen, Eric A., 2001. Approximate planning for factored POMDPs. In *Proceedings of the Sixth European Conference on Planning (ECP-01)*.

Feng, Zhengzhu and Hansen, Eric A., 2002. Symbolic heuristic search for factored markov decision processes. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*.

Feng, Zhengzhu; Hansen, Eric A.; and Zilberstein, Shlomo, 2003. Symbolic generalization for on-line planning. In *Proceedings of the Eighteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, pages 209–216.

Fern, Alan; Yoon, SungWook; and Givan, Robert, 2003. Approximate policy iteration with a policy language bias. In *Advances in Neural Information Processing Systems 16 (NIPS-03)*.

Fikes, Richard E. and Nilsson, Nils J., 1971. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208.

Fishburn, Peter C., 1967. Bounded expected utility. *The Annals of Mathematical Statistics*, 38(4):1054–1060.

Fishburn, Peter C., 1970. *Utility Theory for Decision Making*. John Wiley & Sons.

Fishburn, Peter C., 1975. Unbounded expected utility. *The Annals of Statistics*, 3(4):884–896.

French, Simon, 1986. *Decision Theory: An Introduction to the Mathematics of Rationality*. Prentice Hall.

Friedman, Milton and Savage, Leonard J., 1948. The utility analysis of choices involving risk. *The Journal of Political Economy*, 56(4):279–304.

Givan, Robert; Dean, Thomas; and Greig, Matthew, 2003. Equivalence notions and model minimization in Markov decision processes. *Artificial Intelligence*, 147:163–223.

Givan, Robert; Leach, Sonia; and Dean, Thomas, 2000. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 122(1-2):71–109.

Goodwin, Richard T.; Akkiraju, Rama; and Wu, Frederick, 2002. A decision-support system for quote-generation. In *Proceedings of the Fourteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-02)*, pages 830–837.

Guestrin, Carlos; Koller, Daphne; and Parr, Ronald, 2001. Max-norm projections for factored MDPs. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, pages 673–680.

Guestrin, Carlos; Koller, Daphne; Parr, Ronald; and Venkataraman, Shobha, 2003. Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research*, 19:399–468.

Haddawy, Peter and Hanks, Steve, 1992. Representations for decision-theoretic planning: Utility functions for deadline goals. In *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR-92)*, pages 71–82.

Hansen, Eric A., 1994. Cost-effective sensing during plan execution. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, pages 1029–1035.

Hansen, Eric A., 1997. Markov decision processes with observation costs. Technical Report 97-01, Department of Computer Science, University of Massachusetts at Amherst.

Hansen, Eric A. and Feng, Zhengzhu, 2000. Dynamic programming for POMDPs using a factored state representation. In *Proceedings of the Fifth International Conference on Artificial Intelligence Planning and Scheduling (AIPS-00)*, pages 130–139.

Hansen, Eric A.; Zhou, Rong; and Feng, Zhengzhu, 2002. Symbolic heuristic search using decision diagrams. In *Proceedings of the Fifth International Symposium on Abstraction, Reformulation and Approximation (SARA-02)*.

Hansen, Eric A. and Zilberstein, Shlomo, 1998. Heuristic search in cyclic AND/OR graphs. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, pages 412–418.

Hansen, Eric A. and Zilberstein, Shlomo, 1999a. A heuristic search algorithm for Markov decision problems. In *Proceedings of Bar-Ilan Symposium on the Foundation of Artificial Intelligence*.

Hansen, Eric A. and Zilberstein, Shlomo, 1999b. Solving Markov decision problems using heuristic search. In *Proceedings of AAAI Spring Symposium on Search Techniques from Problem Solving under Uncertainty and Incomplete Information, March, Stanford, California*.

Hansen, Eric A. and Zilberstein, Shlomo, 2001. LAO*: A heuristic search algorithm that finds solutions with loops. *Artificial Intelligence*, 129:35–62.

Hart, Peter E.; Nilsson, Nils J.; and Raphael, Bertram, 1968. A formal basis for the heuristic determination of minimum cost paths in graphs. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107.

Hartfiel, Darald J., 2002. *Nonhomogeneous Matrix Products*. World Scientific.

Harvey, Charles, 1995. Proportional discounting of future costs and benefits. *Mathematics of Operations Research*, 20(2):381–399.

Hauskrecht, Milos; Meuleau, Nicolas; Kaelbling, Leslie Pack; Dean, Thomas; and Boutilier, Craig, 1998. Hierarchical solution of Markov decision processes using macro-actions. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-98)*.

Hernández-Hernández, Daniel and Marcus, Steven I., 1996. Risk-sensitive control of Markov processes in countable state space. *Systems and Control Letters*, 29:147–155.

Hernández-Hernández, Daniel and Marcus, Steven I., 1999. Existence of risk-sensitive optimal stationary policies for controlled Markov processes. *Applied Mathematics and Optimization*, 40(3):273–285.

Hill, Theodore P. and Pestien, Victor C., 1987. The existence of good Markov strategies for decision processes with general payoffs. *Stochastic Processes and their Applications*, 24:61–76.

Hinderer, Karl, 1970. *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*. Springer-Verlag.

Hoey, Jesse; St. Aubin, Robert; Hu, Alan J.; and Boutilier, Craig, 1999. SPUDD: Stochastic planning using decision diagrams. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 279–288.

Hoffman, Jörg and Nebel, Bernhard, 2001. The FF planning system: Fast plan generation through heuristic search. *Journal of Artificial Intelligence Research*, 14:253–302.

Howard, Ronald A. and Matheson, James E., 1972. Risk-sensitive markov decision processes. *Management Science*, 18(7):356–369.

Jaquette, Stratton C., 1973. Markov decision processes with a new optimality criterion: Discrete time. *The Annals of Statistics*, 1(3):496–505.

Jaquette, Stratton C., 1976. A utility criterion for Markov decision processes. *Management Science*, 23(1):43–49.

Jia, Zhongxiao, 1998. Generalized block Lanczos methods for large unsymmetric eigenproblems. *Numerische Mathematik*, 80:239–266.

Kadota, Yoshinobu; Kurano, Masami; and Yasuda, Masami, 1994. Discounted Markov decision processes with general utility functions. In *Proceedings of the Third Conference of the Association of Asian-Pacific Operational Research Societies (APORS) within IFORS, July 26–29, 1994, Fukuoka, Japan*, pages 330–337.

Kadota, Yoshinobu; Kurano, Masami; and Yasuda, Masami, 1998a. On the general utility of discounted Markov decision processes. *The International Transactions in Operational Research*, 5(1):27–34.

Kadota, Yoshinobu; Kurano, Masami; and Yasuda, Masami, 1998b. Stopped decision processes in conjuction with general utility. In *Proceedings of the 7th International Conference on Stochastic Programming*.

Kaelbling, Lelie Pack; Littman, Michael L.; and Moore, Andrew W., 1996. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Kahneman, Daniel; Slovic, Paul; and Tversky, Amos, eds., 1982. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kearns, Michael; Mansour, Yishay; and Ng, Andrew Y., 2002. A sparse sampling algorithm for near-optimal planning in large Markov decision processes. *Machine Learning*, 49(2-3):193–208.

Keeney, Ralph L. and Raiffa, Howard, 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons.

Kemeny, John G. and Snell, James Laurie, 1960. *Finite Markov Chains*. D. Van Nostrand Company.

Kemeny, John G.; Snell, James Laurie; and Knapp, Anthony W., 1976. *Denumerable Markov Chains*. Springer-Verlag, 2nd edition.

Kennan, John, 1981. The existence of expected utility maximizing decisions when utility is unbounded. *Econometrica*, 49(1):215–218.

Kerr, Andrew L., 1999. Utility maximising stochastic dynamic programming: An overview. In *Proceedings of the 34th Annual Conference of the Operational Research Society of New Zealand (ORSNZ-99), December 10-11, Hamilton, New Zealand*.

Köbberling, Veronika, 2002. Preference foundations for difference representations. Technical report, METEOR, Maastricht University, The Netherlands.

Koenig, Sven, 1992. *Optimal Probabilistic and Decision-Theoretic Planning Using Markovian Decision Theory*. Master's thesis, Department of Electrical Engineering and Computer Science, University of California at Berkeley.

Koenig, Sven, 1997. *Goal-Directed Acting with Incomplete Information*. Ph.D. thesis, School of Computer Science, Carnegie Mellon University.

Koenig, Sven and Liu, Yaxin, 1999. Sensor planning with non-linear utility functions. In *Proceedings of the Fifth European Conference on Planning (ECP-99)*, pages 265–277.

Koenig, Sven and Liu, Yaxin, 2002. The interaction of representations and planning objectives for decision-theoretic planning tasks. *Journal of Experimental and Theoretical Artificial Intelligence*, 14(4):303–326.

Koenig, Sven and Simmons, Reid G., 1994a. How to make reactive planners risk-sensitive (without altering anything). In *Proceedings of the Second International Conference on Artificial Intelligence Planning Systems (AIPS-94)*, pages 293–298.

Koenig, Sven and Simmons, Reid G., 1994b. Risk-sensitive planning with probabilistic decision graphs. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR-94)*, pages 2301–2308.

Koller, Daphne and Parr, Ronald, 1999. Computing factored value functions for policies in structured MDPs. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*, pages 1332–1339.

Koller, Daphne and Parr, Ronald, 2000. Policy iteration for factored MDPs. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-00)*, pages 326–334.

Korf, Richard E., 1987. Planning as search: A quantitative approach. *Artificial Intelligence*, 33(1):65–88.

Korf, Richard E., 1990. Real-time heuristic search. *Artificial Intelligence*, 42(3):189–212.

Kreps, David M., 1977a. Decision problems with expected utility criteria, I: Upper and lower convergent utility. *Mathematics of Operations Research*, 2(1):45–53.

Kreps, David M., 1977b. Decision problems with expected utility criteria, II: Stationarity. *Mathematics of Operations Research*, 2(3):266–274.

Kreps, David M., 1978. Decision problems with expected utility criteria, III: Upper and lower transience. *SIAM Journal of Control and Optimization*, 16(3):420–428.

Laibson, David I., 1994. *Hyperbolic Discounting and Consumption*. Ph.D. thesis, Department of Economics, Massachusetts Institute of Technology.

Liu, Yan-Qun; Teo, Kok-Lay; and Yang, Xiao-Qi, 1999. Approximation methods for non-convex curves. *European Journal of Operational Research*, 117:125–135.

Marcus, Steven I.; Fernández-Gaucherand, Emmanual; Hernández-Hernandez, Daniel; Coraluppi, Stefano; and Fard, Pedram, 1997. Risk sensitive Markov decision processes. In Christopher I. Byrnes; Biswa Nath Datta; Clyde F. Martin; and David S. Gilliam, eds., *Systems and Control in the Twenty-First Century*, pages 263–279. Birkhäuser.

McCarthy, John and Hayes, Patrick J., 1969. Some philosophical problems from the standpoint of artificial intelligence. In Bernald Meltzer and Donald Michie, eds., *Machine Intelligence*, pages 463–502. Edinburgh University Press.

Müller, Alfred, 2000. Expected utility maximization of optimal stopping problems. *European Journal of Operational Research*, 12:101–114.

Murthy, Sesh; Akkiraju, Rama; Goodwin, Richard; Keskinocak, Pinar; Rachlin, John; Wu, Frederick; Kumaran, Santhosh; Yeh, James; Fuhrer, Robert; Aggarwal, Alok; Sturzenbecker, Martin; Jayaraman, Ranga; and Daigle, Bob, 1999. Cooperative multi-objective decision-support for the paper industry. *Interface*, 29(5):5–30.

Nau, Dana S.; Cao, Yue; Lotem, Amnon; and Muñoz-Avila, Héctor, 1999. SHOP: Simple hierarchical ordered planner. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*, pages 968–973.

Ng, Andrew Y. and Jordan, Michael I., 2000. PEGASUS: A policy search method for large MDPs and POMDPs. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-00)*.

Nilsson, Nils J., 1980. *Principles of Artificial Intelligence*. Morgan Kaufman.

Ohtsubo, Yoshio and Toyonaga, Kenji, 2002. Optimal policy for minimizing risk models in Markov decision processes. *Journal of Mathematical Analysis and Applications*, 271:66–81.

Parr, Ronald, 1998. *Hierarchical Control and Learning for Markov Decision Processes*. Ph.D. thesis, Department of Electrical Engineering and Computer Science, University of California at Berkeley.

Parr, Ronald and Russell, Stuart, 1998. Reinforcement learning with hierarchies of machines. In *Advances in Neural Information Processing Systems 11 (NIPS-98)*, pages 1043–1049.

Patek, Stephen D., 2001. On terminating Markov decision processes with a risk averse objective function. *Automatica*, 37(9):1379–1386.

Patrascu, Relu; Poupart, Pascal; Schuurmans, Dale; Boutilier, Craig; and Guestrin, Carlos, 2002. Greedy linear value function approximation for factored Markov decision processes. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*.

Pearl, Judea, 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufmann.

Pednault, Edwin P.D., 1989. ADL: Exploring the middle ground between STRIPS and the situation calculus. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR-89)*, pages 324–332.

Pell, Barney; Bernard, Douglas; Chien, Steve; Gat, Erann; Muscettola, Nicola; Nayak, P. Pandurang; Wagner, Michael; and Williams, Brian, 1998. An autonomous spacecraft agent prototype. *Autonomous Robotics*, 5:1–27.

Porteus, Evan L., 1975. On the optimality of structured policies in countable stage decision processes. *Management Science*, 22(2):148–157.

Poupart, Pascal; Boutilier, Craig; Schuurmans, Dale; and Patrascu, Relu, 2002. Piecewise linear value function approximation for factored MDPs. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*.

Pratt, John W., 1964. Risk aversion in the small and in the large. *Econometrica*, 32(1-2):122–136.

Precup, Doina, 2000. *Temporal Abstraction in Reinforcement Learning*. Ph.D. thesis, Department of Computer Science, University of Massachusetts at Amherst.

Precup, Doina; Sutton, Richard S.; and Singh, Satinder, 1997. Planning with closed-loop macro actions. In *Working Notes of the AAAI Fall Symposium on Model-Directed Autonomous Systems*, pages 70–76. Cambridge, MA.

Puterman, Martin L., 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Rote, Günter, 1992. The convergence rate of the sandwich algorithm for approximating convex functions. *Computing*, 48:337–361.

Russell, Stuart J. and Novig, Peter, 2002. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition.

Schäl, Manfred, 1981. Utility functions and optimal policies in sequential decision problems. In Otto Moeschlin and Diethard Pallaschke, eds., *Game Theory and Mathematical Economics*, pages 357–365. North-Holland Publishing Company.

Seneta, Eugene, 1981. *Non-Negative Matrices and Markov Chains*. Springer-Verlag.

Simmons, Reid; Krotkov, Eric; Chrisman, Lonnie; Cozman, Fabio; Goodwin, Richard; Hebert, Martial; Katragadda, Lalitesh; Koenig, Sven; Krishnaswamy, Gita; Shinoda, Yoshikazu; Whittaker, William; and Klarer, Paul, 1995. Experience with rover navigation for lunar-like terrains. In *Proceedings of the 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-95)*, pages 441–446.

Simon, Herbert A., 1947. *Administrative Behavior*. MacMillan.

Somenzi, Fabio, 2004. CUDD: CU decision diagram package, Release 2.4.0. `http://vlsi.colorado.edu/~fabio/CUDD/cuddIntro.html`.

Sondik, Edward J., 1971. *The Optimal Control of Partially Observable Markov Processes*. Ph.D. thesis, Stanford University.

St. Aubin, Robert; Hoey, Jesse; and Boutilier, Craig, 2000. APRICODD: Approximate policy construction using decision diagrams. In *Advances in Neural Information Processing Systems 13 (NIPS-00)*.

Sutton, Richard S. and Barto, Andrew G., 1998. *Reinforcement Learning: An Introduction*. The MIT Press.

Sutton, Richard S.; McAllester, David; Singh, Satinder; and Mansour, Yishay, 1999a. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems 12 (NIPS-99)*, pages 1057–1063.

Sutton, Richard S.; Precup, Doina; and Singh, Satinder P., 1999b. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211.

Tate, Austin, 1977. Generating project networks. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)*, pages 888–893.

van de Kuilen, Gijs and Wakker, Peter P., 2004. Learning in the Allais paradox. Working paper, CREED, Department of Economics, University of Amsterdam, The Netherlands.

von Neumann, John and Morgenstern, Oskar, 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

White, Douglas John, 1987. Utility, probabilistic constraints, mean and variance of discounted rewards in Markov decision processes. *OR Spektrum*, 9:13–22.

White, Douglas John, 1993. Minimising a threshold probability in discounted Markov decision processes. *Journal of Mathematical Analysis and Applications*, 173(634-646):634–646.

Whittle, Peter, 1996. Why discount? The rationale of discounting in optimisation problems. In Chris C. Heyde; Yuri V. Prohorov; Ronald Pyke; and Svetlosar T. Rachev, eds., *Athens Conference on Applied Probability and Time Series, Volume 1: Applied Probability*, volume 114 of *Lecture Notes in Statistics*, pages 354–360. Springer-Verlag.

Wu, Congbin and Lin, Yuanlie, 1999. Minimizing risk models in Markov decision processes with policies depending on target values. *Journal of Mathematical Analysis and Applications*, 231:47–67.

Xie, Aiguo and Beerel, Peter A., 2000. Implicit enumeration of strongly connected components and an application to formal verification. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19:1225–1230.

Yu, Stella X.; Lin, Yuanlie; and Yan, Pingfan, 1998. Optimization models for the first arrival target distribution function in discrete time. *Journal of Mathematical Analysis and Applications*, 225:193–223.

Zilberstein, Shlomo; Washington, Richard; Bernstein, Daniel S.; and Mouaddib, Abdel-Illah, 2002. Decision-theoretic control of planetary rovers. In Michael Beetz; Joachim Hertzberg; Malik Ghallab; and Martha E. Pollack, eds., *Advances in Plan-Based Control of Robotic Agents*, volume 2466 of *Lecture Notes in Computer Science*, pages 270–289. Springer.