Bibliography

Abbeel, P. and A. Ng (2004). "Apprenticeship learning via inverse reinforcement learning". In: *Proc. 21st Int. Conf. Machine learning*.

Abraham, R., J. Marsden, and T. Ratiu (1988). *Manifolds, Tensor Analysis, and Applications*. Springer.

Abramson, N. (2009). "The ALOHAnet: Surfing for wireless data". In: *IEEE Communications Magazine* 47 (12), pages 21–25.

Alais, M. (1953). "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine". In: *Econometrica* 21 (4), pages 503–546.

Andrieu, C. et al. (2003). "An introduction to MCMC for machine learning". In: *Machine Learning* 50, pages 5–43.

Åström, K. (1965). "Optimal control of Markov processes with incomplete state information". In: *J. Mathematical Analysis and Applications* 10, pages 174–205.

Auer, P., N. Cesa-Bianchi, and P. Fisher (2002). "Finite-time analysis of the multiarmed bandit problem". In: *Machine Learning* 47, pages 235–256.

Baum, L. and T. Petrie (1966). "Statistical inference for probabilistic functions of finite state Markov chains". In: *Annals of Mathematical Statistics* 37 (6), pages 1554–1563.

Bellman, R. (1954). *The Theory of Dynamic Programming*. Technical report P-550. The RAND corporation.

Ben-David, S. and S. Shalev-Shwartz (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Benveniste, A., M. Métevier, and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag.

Bernoulli, D. (1738). "Specimen theoriae novae de mensura sortis". In: *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5, pages 175–192. "Exposition of a new theory on the measurement of risk". Translated by L. Sommers. In: *Econometrica* 22 (1), pages 23–36.

Bertsekas, D. (2015). *Convex Optimization Algorithms*. Athena Scientific.

Bertsekas, D. and D. Castañon (1999). "Rollout algorithms for stochastic scheduling problems". In: *J. Heuristics* 5, pages 89–108.

Bertsekas, D. and S. Shreve (1996). *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific.

Bertsekas, D. and J. Tsitsiklis (2008). *Introduction to Probability*. Athena Scientific.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.

Blackwell, D. (1965). "Discounted dynamic programming". In: *Annals of Mathematical Statistics* 36 (1), pages 226–235.

Boger, J. et al. (2005). "A decision-theoretic approach to task assistance for persons with dementia". In: *Proc. 19th Int. Joint Conf. Artificial Intelligence*, pages 1293–1299.

Borkar, V. (2008). *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press.

Boyan, J. and A. Moore (1995). "Generalization in reinforcement learning: Safely approximating the value function". In: *Adv. Neural Information Proc. Systems 7*, pages 369–376.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

Carothers, N.L. (2000). *Real Analysis*. Cambridge University Press.

Cassandra, A. (1998). "Exact and approximate algorithms for partially observable Markov decision processes". PhD thesis. Brown University.

Cassandra, A., L. Kaelbling, and J. Kurien (1996). "Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation". In: *Proc. 1996 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 963–972.

Cassandra, A., L. Kaelbling, and M. Littman (1994). "Acting optimally in partially observable stochastic domains". In: *Proc. 12th AAAI Int. Conf. Artificial Intelligence*, pages 1023–1028.

Cassandra, A., M. Littman, and N. Zhang (1997). "Incremental Prunning: A simple, fast, exact method for partially observable Markov decision processes". In: *Proc. 13th Conf. Uncertainty in Artificial Intelligence*, pages 54–61.

Chadès, I. et al. (2008). "When to stop managing or surveying cryptic threatened species". In: *Proc. National Academy of Sciences* 105 (37), pages 13936–13940.

Cheng, H. (Aug. 1988). "Algorithms for partially observable Markov decision processes". PhD thesis. Univ. British Columbia.

Choi, J. and K. Kim (2011). "Inverse reinforcement learning in partially observable environments". In: *J. Machine Learning Res.* 12, pages 691–730.

Chrisman, L. (1992). "Reinforcement learning with perceptual aliasing: The perceptual distinctions approach". In: *Proc. 10th AAAI Int. Conf. Artificial Intelligence*, pages 183–188.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals, and Systems* 2, pages 303–314.

Dempster, A., N. Laird, and D. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *J. Royal Statistical Society B* 39 (1), pages 1–38.

Duflo, M. (1997). *Random Iterative Models*. Springer.

Durbin, R. et al. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press.

Ellis, H., M. Jiang, and R. Corotis (1995). "Inspection, maintenance, and repair with partial observability". In: *J. Infrastructure Systems* 1 (2), pages 92–99.

Ellsberg, D. (1961). "Risk, ambiguity, and the savage axioms". In: *Quarterly Journal of Economics* 75 (4), pages 643–669.

Farias, D. (2002). "The linear programming approach to approximate dynamic programming: Theory and application". PhD thesis. Stanford Univ.

Farias, D. and B. Van Roy (2003). "The linear programming approach to approximate dynamic programming". In: *Operations Research* 51 (6), pages 850–865.

Feinberg, E., J. Huang, and B. Scherrer (2014). "Modified policy iteration algorithms are not strongly polynomial for discounted dynamic programming". In: *Operations Research Letters* 42, pages 429–431.

Ferns, N., P. Panangaden, and D. Precup (2004). "Metrics for finite Markov decision processes". In: *Proc. 20th Conf. Uncertainty in Artificial Intelligence*, pages 162–169.

Ferns, N. and D. Precup (2014). "Bisimulation metrics are optimal value functions". In: *Proc. 30th Conf. Uncertainty in Artificial Intelligence*, pages 210–219.

Finn, C., S. Levine, and P. Abbeel (2016). "Guided cost learning: Deep inverse optimal control via policy optimization". In: *CoRR* abs/1603.00448.

Fishburn, P. (1970). *Utility Theory for Decision Making*. John Wiley & Sons.

— (1982). *The Foundations of Expected Utility*. D. Reidel Publishing.

Flach, P. (2012). *Machine Learning*. Cambridge University Press.

Friedman, J., R. Tibshirani, and T. Hastie (2001). *The Elements of Statistical Learning*. Springer.

Geman, S. and D. Geman (1984). "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". In: *IEEE Trans. Pattern Analysis and Machine Intelligence* 6, pages 721–741.

Gilbert, E. (1959). "On the identifiability problem for functions of finite Markov chains". In: *Annals of Mathematical Statistics* 30 (3), pages 688–697.

Gilboa, I. (2009). *Theory of Decision Under Uncertainty*. Cambridge University Press.

Givan, R., T. Dean, and M. Greig (2003). "Equivalence notions and model minimization in Markov decision processes". In: *Artificial Intelligence* 147 (1-2), pages 163–223.

Gluss, B. (1959). "An optimum policy for detecting a fault in a complex system". In: *Operations Research* 24, pages 468–477.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press.

Gordon, G. (1995). "Stable function approximation in dynamic programming". In: *Proc. 12th Int. Conf. Machine Learning*, pages 261–268.

Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press.

Halmos, P. (1987). *Finite Dimensional Vector Spaces*. Springer.

Hansen, E. (1997). "An improved policy iteration algorithm for partially observable MDPs". In: *Adv. Neural Information Proc. Systems 10*, pages 1015–1021.

— (1998). "Solving POMDPs by searching in policy space". In: *Proc. 14th Conf. Uncertainty in Artificial Intelligence*, pages 211–219.

Hastings, W. (1970). "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57, pages 97–109.

Hauskrecht, M. (2000). "Value-function approximations for partially observable Markov decision processes". In: *J. Artificial Intelligence Res.* 13 (1), pages 33–94.

Hauskrecht, M. and H. Fraser (2000). "Planning treatment of ischemic heart disease with partially observable Markov decision processes". In: *Artificial Intelligence in Medicine* 18 (3), pages 221–244.

Ho, J. and S. Ermon (2016). "Generative adversarial imitation learning". In: *Adv. Neural Information Proc. Systems 29*, pages 4565–4573.

Hoffman, M. et al. (2007). "Trans-dimensional MCMC for Bayesian policy learning". In: *Adv. Neural Information Proc. Systems 20*, pages 665–672.

Hornik, K. (1991). "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4, pages 251–257.

Hsu, D., W. Lee, and N. Rong (2007). "What makes some POMDP problems easy to approximate?" In: *Adv. Neural Information Proc. Systems 20*, pages 689–696.

Ignall, E. and P. Kolesar (1974). "Optimal dispatching of an infinite-capacity shuttle: Control at a single terminal". In: *Operations Research* 22, pages 1008–1024.

Ji, S. et al. (2007). "Point-based policy iteration". In: *Proc. 22nd AAAI Conf. Artificial Intelligence*, pages 1243–1249.

Kaelbling, L., M. Littman, and A. Cassandra (1998). "Planning and acting in partially observable stochastic domains". In: *Artificial Intelligence* 101, pages 99–134.

Kendall, D. (1959). "Unitary dilations of Markov transition operators and the corresponding integral representation for transition probability matrices". In: *Probability and Statistics: The Harald Cramér Volume*. Edited by U. Grenander. Almqvist and Wiksell, pages 139–161.

Khachiyan, L. (1980). "Polynomial algorithms in linear programming". In: *USSR Computational Mathematics and Mathematical Physics* 20 (1), pages 53–72.

Khalil, H. (2001). *Nonlinear Systems*. 3rd. Prentice Hall.

Kochenderfer, Mykel J. (2015). *Decision Making Under Uncertainty: Theory and Application*. MIT Press.

Krogh, A., I.S. Mian, and D. Haussler (1994). "A hidden Markov model that finds genes in *E.coli* DNA". In: *Nucleic Acids Research* 22 (22), pages 4768–4778.

Kuhn, H. (1955). "The Hungarian method for the assignment problem". In: *Naval Research Logistics Quarterly* 2, pages 83–97.

Kurniawati, H., D. Hsu, and W. Lee (2008). "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces". In: *Proc. Robotics: Science and Systems IV*, pages 65–72.

Kushner, H. and G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Springer.

Lefèvre, C. (1981). "Optimal control of a birth and death epidemic process". In: *Operations Research* 29 (5), pages 971–982.

Littman, M. (Dec. 1994). *The Witness algorithm: Solving partially observable Markov decision processes*. Technical report CS-94-40. Dep. Computer Science, Brown University.

— (1995). "On the complexity of solving Markov decision problems". In: *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, pages 394–402.

— (Mar. 1996). "Algorithms for sequential decision-making". PhD thesis. Dep. Computer Science, Brown University.

Littman, M., A. Cassandra, and L. Kaelbling (1995). "Learning policies for partially observable environments: Scaling up". In: *Proc. 12th Int. Conf. Machine Learning*, pages 362–370.

Liu, J., W. Wong, and A. Kong (1995). "Covariance structure and convergence rate of the Gibbs sampler with various scans". In: *J. Royal Statistical Society B* 57 (1), pages 157–169.

Lopes, M., F. S. Melo, B. Kenward, et al. (2009). "A computational model of social-learning mechanisms". In: *Adaptive Behaviour* 17 (6), pages 467–483.

Lopes, M., F. S. Melo, and L. Montesano (2009). "Active learning for reward estimation in inverse reinforcement learning". In: *Proc. 2009 Eur. Conf. Machine Learning and Practice of Knowledge Discovery in Databases*, pages 31–46.

Lovejoy, W. (1991). "Computationally feasible bounds for partially observed Markov decision processes". In: *Operations Research* 39 (1), pages 162–175.

Luenberger, D. (1979). *Introduction to Dynamical Systems*. John Wiley and Sons.

Lusena, C., J. Goldsmith, and M. Mundhenk (2001). "Nonapproximability results for partially observable Markov decision processes". In: *J. Artificial Intelligence Res.* 14, pages 83–103.

Madani, O., S. Hanks, and A. Condon (2003). "On the undecidability of probabilistic planning and related stochastic optimization problems". In: *Artificial Intelligence* 147, pages 5–34.

Mausam and A. Kolobov (2012). *Planning with Markov Decision Processes: An AI Perspective*. Morgan & Claypool Publishers.

McAllester, D. and S. Singh (1999). "Approximate planning for factored POMDPs using belief state simplification". In: *Proc. 15th Annual Conf. Uncertainty in Artificial Intelligence*, pages 409–416.

McCallum, A. (1992). *First results with utile distinction memory for reinforcement learning*. Technical report 446. Computer Science Department, University of Rochester.

Melo, F. S. and M. Lopes (2010). "Learning from demonstrations using MDP induced metrics". In: *Proc. 2010 Eur. Conf. Machine Learning and Practice of Knowledge Discovery in Databases*, pages 385–401.

Melo, F. S., M. Lopes, and R. Ferreira (2010). "Analysis of inverse reinforcement learning with perturbed demonstrations". In: *Proc. 19th European Conf. Artificial Intelligence*, pages 349–354.

Melo, F.S. and M.I. Ribeiro (2006). "Transition entropy in partially observable Markov decision processes". In: *Proc. 9th Int. Conf. Intelligent Autonomous Systems*, pages 282–289.

Metropolis, N. et al. (1953). "Equations of state calculations by fast computing machines". In: *J. Chemical Physics* 21 (6), pages 1087–1091.

Meuleau, N. et al. (1999). "Solving POMDPs by searching the space of finite policies". In: *Proc. 15th Conf. Uncertainty in Artificial Intelligence*, pages 417–426.

Meyn, S. and R. Tweedie (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Mikusiński, P. and M. Taylor (2002). *An Introduction to Multivariable Analysis: From Vector to Manifold*. Springer Science.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

Monahan, G. (1982). "A survey of partially observable Markov decision processes: Theory, models, and algorithms". In: *Management Science* 28 (1), pages 1–16.

Mundhenk, M. (2000). *The complexity of planning with partially-observable Markov decision processes*. Technical report TR2000-376. Computer Science Dep., Dartmouth College.

Mundhenk, M. et al. (2000). "Complexity of finite-horizon Markov decision process problems". In: *J. ACM* 37 (4), pages 681–720.

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Neu, G. and C. Szepesvári (2007). "Apprenticeship learning using inverse reinforcement learning and gradient methods". In: *Proc. 23rd Conf. Uncertainty in Artificial Intelligence*, pages 295–302.

— (2009). "Training parsers by inverse reinforcement learning". In: *Machine Learning* 77 (2-3), pages 303–337.

Newell, A. and H. Simon (June 1956). *The logic theory machine: A complex information processing system*. Technical report. The RAND corporation.

Ng, A. and S. Russell (2000). "Algorithms for inverse reinforcement learning". In: *Proc. 17th Int. Conf. Machine Learning*, pages 663–670.

Nikovski, D. and I. Nourbakhsh (2002). "Learning probabilistic models for state tracking of mobile robots". In: *Proc. 2002 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 1026–1031.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.

Nummelin, E. and R. Tweedie (1978). "Geometric ergodicity and R-positivity for general Markov chains". In: *Annals of Probability* 6 (3), pages 404–420.

Ong, S. et al. (2010). "Planning under uncertainty for robotic tasks with mixed observability". In: *Int. J. Robotics Research* 29 (8), pages 1053–1068.

Papadimitriou, C. and J. Tsitsiklis (1987). "The complexity of Markov decision processes". In: *Mathematics of Operations Research* 12 (3), pages 441–450.

Paquet, S., L. Tobin, and B. Chaib-draa (2005). "An online POMDP algorithm for complex multiagent environments". In: *Proc. 4th Int. Conf. Autonomous Agents and Multiagent Systems*, pages 970–977.

Parks, D. (1981). "Concurrency and automata on infinite sequences". In: *Proc. 5th GI-Conference on Theoretical Computer Science*, pages 167–183.

Pineau, J., G. Gordon, and S. Thrun (2003). "Point-based value iteration: An anytime algorithm for POMDPs". In: *Proc. 18th Int. Joint Conf. Artificial Intelligence*, pages 1025–1032.

Popov, N. (1977). "Conditions for geometric ergodicity of countable Markov chains". In: *Soviet Math. Doklady* 18 (3), pages 676–679.

Poupart, P. (2005). "Exploiting structure to efficiently solve large scale partially observable Markov decision processes". PhD thesis. Dep. Computer Science, Univ. Toronto.

Poupart, P. and C. Boutilier (2003). "Bounded finite state controllers". In: *Adv. Neural Information Proc. Systems 16*, pages 823–830.

— (2004). "VDCBPI: An approximate scalable algorithm for large POMDPs". In: *Adv. Neural Information Proc. Systems 17*, pages 1081–1088.

Poupart, P., K. Kim, and D. Kim (2011). "Closing the gap: Improved bounds on optimal POMDP solutions". In: *Proc. 21st Int. Conf. Automated Planning and Scheduling*, pages 194–201.

Powers, M. (2015). "Paradox-proof utility functions for heavy-tailed payoffs: Two instructive two-envelope problems". In: *Risks* 3, pages 26–34.

Puterman, M. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Rabiner, L. (1989). "A tutorial to hidden Markov models and selected applications in speech recognition". In: *Proc. IEEE* 77 (2), pages 257–286.

Ramachandran, D. and E. Amir (2007). "Bayesian inverse reinforcement learning". In: *Proc. 20th Int. Joint Conf. Artificial Intelligence*, pages 2586–2591.

Ravindran, B. (Feb. 2004). "An algebraic approach to abstraction in reinforcement learning". PhD thesis. University of Massachusetts Amherst.

Ravindran, B. and A. Barto (2002). "Model minimization in hierarchical reinforcement learning". In: *Proc. 5th Int. Symp. Abstraction, Reformulation and Approximation*, pages 196–211.

Robert, C. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer.

Roberts, G. and J. Rosenthal (2004). "General state space Markov chains and MCMC algorithms". In: *Probability Surveys* 1, pages 20–71.

Roberts, G. and R. Tweedie (1996). "Geometric convergence and central limit theorems for multivariate Hastings and Metropolis algorithms". In: *Biometrika* 83, pages 95–110.

Roberts, L. (1975). "ALOHA packet system with and without slots and capture". In: *Computer Communications Review* 5 (2), pages 28–42.

Ross, S. and B. Chaib-draa (2007). "AEMS: An anytime online search algorithm for approximate policy refinement in large POMDPs". In: *Proc. 20th Int. Joint Conf. Artificial Intelligence*, pages 2592–2598.

Ross, S., G. Gordon, and J. Bagnell (2011). "A reduction of imitation learning and structured prediction to no-regret online learning". In: *Proc. 14th Int. Conf. Artificial Intelligence and Statistics*, pages 627–635.

Ross, S., J. Pineau, et al. (2008). "Online planning algorithms for POMDPs". In: *J. Artificial Intelligence Res.* 32, pages 663–704.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.

Russell, S. and P. Norvig (2010). *Artificial Intelligence, a Modern Approach*. Prentice Hall.

Samuel, A. (1959). "Some studies in machine learning using the game of checkers". In: *IBM J. Research and Development* 3 (3). Reprinted in IBM J. Res. Devel. 44:1/2, pp. 206-226, 2000., pages 210–229.

— (1967). "Some studies in machine learning using the game of checkers II: Recent progress". In: *IBM J. Research and Development* 11, pages 601–617.

Savage, L. (1972). *The Foundations of Statistics*. Dover Publications.

Shani, G., R. Brafman, and S. Shimony (2007). "Forward search value iteration for POMDPs". In: *Proc. 20th Int. Joint Conf. Artificial Intelligence*, pages 2619–2624.

Shani, G., J. Pineau, and R. Kaplow (2013). "A survey of point-based POMDP solvers". In: *J. Autonomous Agents and Multi-Agent Systems* 27, pages 1–51.

Shoham, Y. and K. Leyton-Brown (2009). *Multiagent Systems: Algorithmic, Game-theoretic, and Logical Foundations*. Cambridge University Press.

Sigaud, O. and O. Buffet, editors (2013). *Markov Decision Processes in Artificial Intelligence*. John Wiley & Sons.

Silver, D. and J. Veness (2010). "Monte-Carlo planning in large POMDPs". In: *Adv. Neural Information Proc. Systems 23*. Volume 2164-2172.

Smith, T. and R. Simmons (2005). "Point-based POMDP algorithms: Improved analysis and implementation". In: *Proc. 21st Conf. Uncertainty in Artificial Intelligence*, pages 542–547.

Sondik, E. (June 1971). "The optimal control of partially observable Markov processes". PhD thesis. Stanford Univ.

— (1978). "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs". In: *Operations Research* 25 (2), pages 282–304.

Spaan, M. and N. Vlassis (2005). "PERSEUS: Randomized point-based value iteration for POMDPs". In: *J. Artificial Intelligence Res.* 24, pages 195–220.

Stigler, G. (1950a). "The development of utility theory, part I". In: *J. Political Economy* 58 (4), pages 307–327.

— (1950b). "The development of utility theory, part II". In: *J. Political Economy* 63 (5), pages 373–396.

Strang, G. (2009). *Introduction to Linear Algebra*. Wellesley Cambridge Press.

Sutton, R. (1988). "Learning to predict by the methods of temporal differences". In: *Machine Learning* 3, pages 9–44.

Sutton, R. and A. Barto (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Sutton, R., D. Precup, and S. Singh (1999). "Between MDPs and Semi-MDPs: A framework for temporal abstraction in reinforcement learning". In: *Artificial Intelligence* 112, pages 181–211.

Syed, U. and R. Schapire (2007a). "A game-theoretic approach to apprenticeship learning". In: *Adv. Neural Information Proc. Systems 20*, pages 1449–1456.

— (2007b). "Imitation learning with a value-based prior". In: *Proc. 23rd Conf. Uncertainty in Artificial Intelligence*, pages 384–391.

— (2010). "A reduction from apprenticeship learning to classification". In: *Adv. Neural Information Proc. Systems 23*, pages 2253–2261.

Szepesvári, C. and M. Littman (1999). "A unified analysis of value-function-based reinforcement-learning algorithms". In: *Neural Computation* 11 (8), pages 2017–2060.

Taylor, J., D. Precup, and P. Panangaden (2008). "Bounding performance loss in approximate MDP homomorphisms". In: *Adv. Neural Information Processing Systems 21*, pages 1649–1656.

Tierney, L. (1994). "Markov Chains for exploring posterior distributions". In: *Annals of Statistics* 22, pages 1701–1762.

Toussaint, M., S. Harmeling, and A. Storkey (Dec. 2006). *Probabilistic inference for solving (PO)MDPs*. Technical report Informatics Research Report 0934. School of Informatics, University of Edimburgh.

Toussaint, M. and A. Storkey (2006). "Probabilistic inference for solving discrete and continuous state Markov decision processes". In: *Proc. 23rd Int. Conf. Machine Learning*, pages 945–952.

Tsitsiklis, J. (2002). "On the convergence of optimistic policy iteration". In: *J. Machine Learning Res.* 3, pages 59–72.

— (2007). "NP-Hardness of checking the unichain condition in average cost MDPs". In: *Operations Research Letters* 35, pages 319–323.

Tsitsiklis, J. and B. Van Roy (1996). "Feature-based methods for large-scale dynamic programming". In: *Machine Learning* 22, pages 59–94.

Vere-Jones, D. (1962). "Geometric ergodicity in denumerable Markov chains". In: *Quart. J. Math. Oxford* 13 (2), pages 7–28.

Viterbi, A. (1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". In: *IEEE Transactions on Information Theory* 13 (2), pages 260–269.

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Vroman, M. (Oct. 2014). "Maximum likelihood inverse reinforcement learning". PhD thesis. Rutgers University.

Watkins, C. (May 1989). "Learning from delayed rewards". PhD thesis. King's College, University of Cambridge.

White, D. (1993). "A survey of applications of Markov decision processes". In: *J. Operational Research Soc.* 44 (11), pages 1073–1096.

Ye, Y. (2005). "A new complexity result on solving the Markov decision problem". In: *Mathematics of Operations Research* 30 (3), pages 733–749.

— (2011). "The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate". In: *Mathematics of Operations Research* 36 (4), pages 593–603.

Yonezaki, T., K. Yoshida, and T. Yagi (1998). "An error correction approach based on the MAP algorithm combined with hidden Markov models". In: *Proc. 1998 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pages 33–36.

Young, S. et al. (2013). "POMDP-based statistical spoken dialog systems: A review". In: *Proc. IEEE* 101 (5), pages 1160–1179.

Zhang, N. and W. Liu (1996). *Planning in stochastic domains: Problem characteristics and approximation*. Technical report HKUST-CS96-31. Dep. Computer Science, Hong Kong University of Science and Technology.

Zhang, Z., D. Hsu, and W. Lee (2014). "Covering number for efficient heuristic-based POMDP planning". In: *Proc. 31st Int. Conf. Machine Learning*, pages 28–36.

Zhang, Z., M. Littman, and X. Chen (2012). "Covering number as a complexity measure for POMDP planning and learning". In: *Proc. 16th AAAI Conf. Artificial Intelligence*, pages 1853–1859.

Ziebart, B. et al. (2008). "Maximum entropy inverse reinforcement learning". In: *Proc. 23rd AAAI Conf. Artificial Intelligence*, pages 1433–1438.