References

1. R. Agrawal, "Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem," *Advances in Applied Probability*, vol.27, 1054–1078, 1995.
2. E. Altman and G. Koole, "On submodular value functions and complex dynamic programming," *Stochastic Models*, vol.14, 1051–1072, 1998.
3. A. Arapostathis, V.S. Borkar, E. Fernández-Gaucherand, M.K. Ghosh, and S.I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: a survey," *SIAM Journal on Control and Optimization*, vol.31, no.2, 282–344, 1993.
4. P. Auer, N. Cesa-Bianchi, and P. Fisher, "Finite-time analysis of the multi-armed bandit problem," *Machine Learning*, vol.47, 235–256, 2002.
5. M. Baglietto, T. Parisini, and R. Zoppoli, "Neural approximators and team theory for dynamic routing: a receding horizon approach," *Proceedings of the 38th IEEE Conference on Decision and Control*, 3283–3288, 1999.
6. V. Balakrishnan and A.L. Tits, "Numerical optimization-based design," *The Control Handbook*, W.S. Levine, ed., CRC Press, Boca Raton, FL, 749–758, 1996.
7. J. Banks, editor, *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, John Wiley & Sons, New York, NY, 1998.
8. D. Barash, "A genetic search in policy space for solving Markov decision processes," *AAAI Spring Symposium on Search Techniques for Problem Solving under Uncertainty and Incomplete Information,* Stanford University, 1999.
9. A. Barto, R. Sutton, and C. Anderson. "Neuron-like elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man and Cybernetics*, vol.13, 835–846, 1983.
10. D.P. Bertsekas, *Dynamic Programming and Optimal Control, Volumes 1 and 2,* Athena Scientific, Belmont, MA, 1995.
11. D.P. Bertsekas, "Differential training of rollout policies," *Proceedings of the 35th Allerton Conference on Communication, Control, and Computing*, 1997.
12. D.P. Bertsekas, "Dynamic programming and suboptimal control: A survey from ASP to MPC," *European Journal of Control,* vol.11, 310–334, 2005.
13. D.P. Bertsekas and D.A. Castanon, "Adaptive aggregation methods for infinite horizon dynamic programming," *IEEE Transactions on Automatic Control*, vol.34, no.6, 589–598, 1989.

14. D.P. Bertsekas and D.A. Castanon, "Rollout algorithms for stochastic scheduling problems," *Journal of Heuristics,* vol.5, 89–108, 1999.
15. D.P. Bertsekas and S.E. Shreve, *Stochastic Control: The Discrete Time Case*, Academic Press, New York, NY, 1978.
16. D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.
17. C. Bes and J.B. Lasserre, "An on-line procedure in discounted infinite-horizon stochastic optimal control," *Journal of Optimization Theory and Applications*, vol.50, 61–67, 1986.
18. S. Bhatnagar, M.C. Fu, and S.I. Marcus, "An optimal structured feedback policy for ABR flow control using two timescale SPSA," *IEEE/ACM Transactions on Networking*, vol.9, 479–491, 2001.
19. S. Bhulai and G. Koole, "On the structure of value functions for threshold policies in queueing models," Technical Report 2001-4, Department of Stochastics, Vrije Universiteit, Amsterdam, 2001.
20. V.D. Blondel and J.N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol.36, 1249–1274, 2000.
21. V.S. Borkar, "White-noise representations in stochastic realization theory," *SIAM Journal on Control and Optimization*, vol.31, 1093–1102, 1993.
22. V.S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes: Methods and Applications*, E.A. Feinberg and A. Shwartz, eds., Kluwer, Boston, MA, 2002.
23. P. Bratley, B.L. Fox, and L.E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York, NY, 1983.
24. S. Chand, V.N. Hsu, and S. Sethi, "Forecast, solution, and rolling horizons in operations management problems: a classified bibliography," *Manufacturing & Service Operations Management*, vol.4, no.1, 25–43, 2003.
25. H.S. Chang, "Multi-policy iteration with a distributed voting," *Mathematical Methods of Operations Research*, vol.60, no.2, 299–310, 2004.
26. H.S. Chang, "On ordinal comparison of policies in Markov reward processes," *Journal of Optimization Theory and Applications*, vol.122, no.1, 207–217, 2004.
27. H.S. Chang, "Multi-policy improvement in stochastic optimization with forward recursive function criteria," *Journal of Mathematical Analysis and Applications*, vol.305, no.1, 130–139, 2005.
28. H.S. Chang, "Converging marriage in honey-bees optimization and application to stochastic dynamic programming," *Journal of Global Optimization*, vol.35, no.3, 423–441, 2006.
29. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, "An adaptive sampling algorithm for solving Markov decision processes," *Operations Research*, vol.53, no.1, 126–139, 2005.
30. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, "An asymptotically efficient simulation-based algorithm for finite horizon stochastic dynamic programming," *IEEE Transactions on Automatic Control*, accepted for publication, 2006.
31. H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus, "Recursive learning automata approach to Markov decision processes," *IEEE Transactions on Automatic Control*, submitted.
32. H.S. Chang, R. Givan, and E.K.P. Chong, "Parallel rollout for on-line solution of partially observable Markov decision processes," *Discrete Event Dynamic Systems: Theory and Application*, vol.15, no.3, 309–341, 2004.

33. H.S. Chang, H-G. Lee, M.C. Fu, and S.I. Marcus, "Evolutionary policy iteration for solving Markov decision processes," *IEEE Transactions on Automatic Control*, vol.50, no.11, 1804-1808, 2005.
34. H.S. Chang and S.I. Marcus, "Approximate receding horizon approach for Markov decision processes: average reward case," *Journal of Mathematical Analysis and Applications*, vol.286, no.2, 636–651, 2003.
35. H.S. Chang and S.I. Marcus, "Two-person zero-sum Markov games: receding horizon approach," *IEEE Transactions on Automatic Control*, vol.48, no.11, 1951–1961, 2003.
36. H. Chin and A. Jafari, "Genetic algorithm methods for solving the best stationary policy of finite Markov decision processes," *Proceedings of the 30th Southeastern Symposium on System Theory*, 538–543, 1998.
37. E.K.P. Chong, R. Givan, and H.S. Chang, "A framework for simulation-based network control via hindsight optimization," *Proceedings of the 39th IEEE Conference on Decision and Control*, 1433–1438, 2000.
38. W.L. Cooper, S.G. Henderson, and M.E. Lewis, "Convergence of simulation-based policy iteration," *Probability in the Engineering and Informational Sciences*, vol.17, no.2, 213–234, 2003.
39. A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm," *ACM Transactions on Mathematical Software*, vol.13, no.3, 262–280, 1987.
40. L. Dai, "Convergence properties of ordinal comparison in the simulation of discrete event dynamic systems," *Journal of Optimization Theory and Applications*, vol.91, 363–388, 1996.
41. P.T. De Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of Operation Research,* vol.134, 19–67, 2005.
42. D.P. de Farias and B. Van Roy, "The linear programming approach to approximate dynamic programming," *Operations Research,* vol.51, no.6, 850–865, 2003.
43. K.A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems,* Ph.D. Thesis, U. Michigan, Ann Arbor, MI, 1975.
44. L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, NY, 1986; also available for free download at `http://cg.scs.carleton.ca/~luc/rnbookindex.html` (accessed July 20, 2006).
45. M. Dorigo and L.M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, vol.1, 53–66, 1997.
46. E. Even-Dar, S. Mannor, and Y. Mansour, "PAC bounds for multi-armed bandit and Markov decision processes," *Proceedings of the 15th Annual Conf. on Computational Learning Theory*, 255–270, 2002.
47. H. Fang and X. Cao, "Potential-based on-line policy iteration algorithms for Markov decision processes," *IEEE Transactions on Automatic Control*, vol.49, 493–505, 2004.
48. A. Federgruen and M. Tzur, "Detection of minimal forecast horizons in dynamic programs with multiple indicators of the future," *Naval Research Logistics*, vol.43, 169–189, 1996.
49. E.A. Feinberg and A. Shwartz, editors, *Handbook of Markov Decision Processes: Methods and Applications*, Kluwer, Boston, MA, 2002.

50. E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, "On the average cost optimality equation and the structure of optimal policies for partially observable Markov processes," *Annals of Operations Research*, vol.29, 471–512, 1991.
51. G.S. Fishman, *Monte Carlo Methods: Concepts, Algorithms, and Applications*, Springer, New York, NY, 1996.
52. G.S. Fishman, *A First Course in Monte Carlo*, Duxbury, Thomson Brooks/Cole, Belmont, CA, 2006.
53. Y. Freund and R. Schapire, "Adaptive game playing using multiplicative weights," *Games and Economic Behavior*, vol.29, 79–103, 1999.
54. M.C. Fu and K.J. Healy, "Techniques for simulation optimization: an experimental study on an (s, S) inventory system," *IIE Transactions*, vol.29, 191–199, 1997.
55. M.C. Fu, J. Hu, and S.I. Marcus, "Model-based randomized methods for global optimization," *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems,* Kyoto, Japan, July 2006.
56. M.C. Fu and X. Jin, "On the convergence rate of ordinal comparisons of random variables," *IEEE Transactions on Automatic Control*, vol.46, 1950–1954, 2001.
57. M.C. Fu, S.I. Marcus, and I-J. Wang, "Monotone optimal policies for a transient queueing staffing problem," *Operations Research*, vol.46, 327–331, 2000.
58. R. Givan, E.K.P. Chong, and H.S. Chang, "Scheduling multiclass packet streams to minimize weighted loss," *Queueing Systems*, vol.41, no.3, 241–270, 2002.
59. R. Givan, S. Leach, and T. Dean, "Bounded Markov decision processes," *Artificial Intelligence*, vol.122, 71–109, 2000.
60. F. Glover, "Tabu search: a tutorial," *Interfaces*, vol.20, no.4, 74-94, 1990.
61. D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Boston, MA, 1989.
62. A. Gosavi, *Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*, Kluwer, 2003.
63. R.C. Grinold, "Finite horizon approximations of infinite horizon linear programs," *Mathematical Programming*, vol.12, 1–17, 1997.
64. R. Hartley, "Inequalities for a class of sequential stochastic decision processes," in *Stochastic Programming*, M.A.H. Dempster, ed., Academic Press, San Diego, CA, 109–123, 1980.
65. D.B. Hausch and W.T. Ziemba, "Bounds on the value of information in uncertain decision problems," *Stochastics*, vol.10, 181–217, 1983.
66. Y. He and E.K.P. Chong, "Sensor scheduling for target tracking: a Monte Carlo sampling approach," *Digital Signal Processing*, vol.16, no.5, pp. 533–545, 2006.
67. Y. He, M.C. Fu, and S.I. Marcus, "Simulation-based algorithms for average cost Markov decision processes," in *Computing Tools for Modeling, Optimization and Simulation, Interfaces in Computer Science and Operations Research*, M. Laguna and J.L. González Velarde, eds., Kluwer, 161–182, 2000.
68. S.G. Henderson and B.L. Nelson, editors, *Handbooks in Operations Research and Management Science: Simulation*, North-Holland/Elsevier, Amsterdam, 2006.
69. O. Hernández-Lerma and J.B. Lasserre, "A forecast horizon and a stopping rule for general Markov decision processes," *Journal of Mathematical Analysis and Applications*, vol.132, 388–400, 1988.

70. O. Hernández-Lerma and J.B. Lasserre, "Error bounds for rolling horizon policies in discrete-time Markov control processes," *IEEE Transactions on Automatic Control*, vol.35, 1118–1124, 1990.
71. O. Hernández-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, NY, 1996.
72. W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol.58, 13–30, 1963.
73. T. Homem-de-Mello, "A study on the cross-entropy method for rare-event probability estimation," *INFORMS Journal on Computing,* accepted for publication.
74. L.J. Hong and B.L. Nelson, "Discrete optimization via simulation using COMPASS," *Operations Research,* vol.54, 115-129, 2006.
75. J. Hu, M.C. Fu, and S.I. Marcus, "A model reference adaptive search method for global optimization," *Operations Research,* in press, 2007.
76. J. Hu, M.C. Fu, and S.I. Marcus, "A model reference adaptive search method for stochastic global optimization," submitted for publication, 2006.
77. J. Hu, M.C. Fu, V. Ramezani, and S.I. Marcus, "An evolutionary random policy search algorithm for solving Markov decision processes," *INFORMS Journal on Computing*, accepted for publication, 2005.
78. S. Iwamoto, "Stochastic optimization of forward recursive functions," *Journal of Mathematical Analysis and Applications*, vol.292, 73–83, 2004.
79. R. Jain and P. Varaiya, "Simulation-based uniform value function estimates of Markov decision processes," *SIAM Journal on Control and Optimization*, vol.45, no.5, 1633–1656, 2006.
80. L. Johansen, *Lectures on Macroeconomic Planning*, North-Holland, Amsterdam, 1977.
81. L. Kaelbling, M. Littman, and A. Moore, "Reinforcement learning: a survey," *Artificial Intelligence*, vol.4, 237–285, 1996.
82. L. Kallenberg, "Finite state and action MDPs," in *Handbook of Markov Decision Processes: Methods and Applications*, E.A. Feinberg and A. Shwartz, eds., Kluwer, Boston, MA, 2002.
83. M. Kearns, Y. Mansour, and A.Y. Ng, "A sparse sampling algorithm for near-optimal planning in large Markov decision processes," *Machine Learning*, vol.49, 193–208, 2001.
84. S.S. Keerthi and E.G. Gilbert, "Optimal infinite horizon feedback laws for a general class of constrained discrete time systems: stability and moving-horizon approximations," *Journal of Optimization Theory and Applications*, vol.57, 265–293, 1988.
85. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol.220, 45–54, 1983.
86. M.Y. Kitaev and V.V. Rykov, *Controlled Queueing Systems,* CRC Press, Boca Raton, FL, 1995.
87. A. Kolarov and J. Hui, "On computing Markov decision theory-based cost for routing in circuit-switched broadband networks," *Journal of Network and Systems Management*, vol.3, no.4, 405–425, 1995.
88. D. Koller and R. Parr, "Policy iteration for factored MDPs," *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, 326–334, 2000.
89. V.R. Konda and J.N. Tsitsiklis, "Actor-critic algorithms," *SIAM Journal on Control and Optimization*, vol.42, no.4, 1143–1166, 2003.

90. G. Koole, "The deviation matrix of the $M/M/1/\infty$ and $M/M/1/N$ queue, with applications to controlled queueing models," *Proceedings of the 37th IEEE Conference on Decision and Control*, 56–59, 1998.
91. G. Koole and P. Nain, "On the value function of a priority queue with an application to a controlled pollying model," *Queueing Systems: Theory and Applications*, accepted for publication.
92. P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
93. T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol.6, 4–22, 1985.
94. A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, 3rd ed., McGraw-Hill, New York, NY, 2000.
95. A.Z.-Z. Lin, J. Bean, and C. White, III, "A hybrid genetic/optimization algorithm for finite horizon partially observed Markov decision processes," *INFORMS Journal on Computing*, vol.16, no.1, 27–38, 2004.
96. N. Littlestone and M.K. Warmnuth, "The weighted majority algorithm," *Information and Computation*, vol.108, 212–261, 1994.
97. M. Littman, T. Dean, and L. Kaelbling, "On the complexity of solving Markov decision problems," *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence*, 394–402, 1995.
98. J. MacQueen, "A modified dynamic programming method for Markovian decision problems," *Journal of Mathematical Analysis and Applications,* vol.14, 38–43, 1966.
99. A. Madansky, "Inequalities for stochastic linear programming problems," *Management Science*, vol.6, 197–204, 1960.
100. S. Mannor, R.Y. Rubinstein, and Y. Gat, "The cross-entropy method for fast policy search," *International Conference on Machine Learning*, 512–519, 2003.
101. P. Marbach and J.N. Tsitsiklis, "Simulation-based optimization of Markov reward processes," *IEEE Transactions on Automatic Control*, vol.46, no.2, 191–209, 2001.
102. P. Marbach and J.N. Tsitsiklis, "Approximate gradient methods in policy-space optimization of Markov reward processes," *Discrete Event Dynamic Systems: Theory and Applications*, vol.13, 111–148, 2003.
103. D.Q. Mayne and H. Michalska, "Receding horizon control of nonlinear system," *IEEE Transactions on Automatic Control*, vol.38, 814–824, 1990.
104. M. Morari and J.H. Lee, "Model predictive control: past, present, and future," *Computers and Chemical Engineering*, vol.23, 667–682, 1999.
105. C.N. Morris, "Natural exponential families with quadratic variance functions," *The Annals of Statistics,* vol.10, 65–80, 1982.
106. H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distributions I. Binary parameters," in *Proceedings of the 4th International Conference on Parallel Problem Solving from Nature*, H. Voigt, W. Ebeling, I. Rechenberg, and H. Schwefel, eds., Springer-Verlag, Berlin, 178–187, 1996.
107. K.S. Narendra and A.L. Thathachar, *Learning Automata: An Introduction*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
108. A.Y. Ng, R. Parr, and D. Koller "Policy search via density estimation," *Advances in Neural Information Processing Systems 12*, NIPS 1999, S.A. Solla, T.K. Leen, and K.-R. Müller, eds., MIT Press, 1022-1028, 2000.
109. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

110. B.J. Oommen and J.K. Lanctot, "Discrete pursuit learning automata," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.20, 931-938, 1990.
111. T.J. Ott and K.R. Krishnan, "Separable routing: a scheme for state-dependent routing of circuit switched telephone traffic," *Annals of Operations Research*, vol.35, 43–68, 1992.
112. W.N. Patten and L.W. White, "A sliding horizon feedback control problem with feedforward and disturbance," *Journal Mathematical Systems, Estimation, and Control*, vol.7, 1–33, 1997.
113. J.M. Peha and F.A. Tobagi, "Evaluating scheduling algorithms for traffic with heterogeneous performance objectives," *Proceedings of the IEEE GLOBECOM*, 21–27, 1990.
114. E.L. Porteus, "Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs," *Journal of Optimization Theory and Applications*, vol.36, 419–432, 1982.
115. W.B. Powell, *Approximate Dynamic Programming*, Wiley, New York, NY, 2007, in preparation.
116. A.S. Poznyak and K. Najim, *Learning Automata and Stochastic Optimization*, Springer-Verlag, New York, NY, 1997.
117. A.S. Poznyak, K. Najim, and E. Gomez-Ramirez, *Self-Learning Control of Finite Markov Chains,* Marcel Dekker, New York, NY, 2000.
118. M.L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, New York, NY, 1994.
119. K. Rajaraman and P.S. Sastry, "Finite time analysis of the pursuit algorithm for learning automata," *IEEE Transactions on Systems, Man, and Cybernetics,* Part B, vol.26, no.4, 590–598, 1996.
120. S.M. Ross, *Applied Probability Models with Optimization Applications*, Dover, Mineola, NY, 1992; originally published by Holden-Day, San Francisco, CA, 1970.
121. S.M. Ross, *Stochastic Processes*, 2nd ed., John Wiley & Sons, New York, NY, 1996.
122. R.Y. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley & Sons, New York, NY, 1993.
123. R.Y. Rubinstein and D.P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation, and Machine Learning,* Springer, New York, NY, 2004.
124. J. Rust, "Structural estimation of Markov decision processes," R. Engle and D. McFadden, eds., *Handbook of Econometrics,* North-Holland/Elsevier, Amsterdam, 1994.
125. J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica,* vol.65, no.3, 487–516, 1997.
126. G. Santharam, P.S. Sastry, and M.A.L. Thathachar, "Continuous action set learning automata for stochastic optimization," *The Franklin Institute*, vol.331B, no.5, 607–628, 1994.
127. U. Savagaonkar, E.K.P. Chong, and R.L. Givan, "Online pricing for bandwidth provisioning in multi-class networks," *Computer Networks*, vol.44, no.6, 835–853, 2004.
128. N. Secomandi, "Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands," *Computers and Operations Research*, vol.27, 1201–1225, 2000.

129. L.I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*, Wiley, New York, NY, 1999.
130. J.G. Shanthikumar and D.D. Yao, "Stochastic monotonicity in general queueing networks," *Journal of Applied Probability*, vol.26, 413–417, 1989.
131. L. Shi and S. Ólafsson, "Nested partitions method for global optimization," *Operations Research*, vol.48, 390–407, 2000.
132. L. Shi and S. Ólafsson, "Nested partitions method for stochastic optimization," *Methodology and Computing in Applied Probability*, vol.2, 271–291, 2000.
133. A.N. Shiryaev, *Probability,* 2nd ed., Springer-Verlag, New York, NY, 1995.
134. J. Si, A.G. Barto, W.B. Powell, and D.W. Wunsch, editors, *Handbook of Learning and Approximate Dynamic Programming*, IEEE Press, Piscataway, NJ, 2004.
135. J.E. Smith and K.F. McCardle, "Structural properties of stochastic dynamic programs," *Operations Research*, vol.50, 796–809, 2002.
136. M. Srinivas, and L.M. Patnaik, "Genetic algorithms: a survey," *IEEE Computer*, vol.27, no.6, 17–26, 1994.
137. S. Stidham and R. Weber, "A survey of Markov decision models for control of networks of queues," *Queueing Systems*, vol.13, 291–314, 1993.
138. R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 1998.
139. G. Tesauro and G.R. Galperin, "On-line policy improvement using Monte-Carlo search," in *Advances in Neural Information Processing Systems 9*, NIPS 1996, M. Mozer, M.I. Jordan, and T. Petsche, eds., MIT Press, 1068–1074, 1997.
140. M.A.L. Thathachar and P.S. Sastry, "A class of rapidly converging algorithms for learning automata," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.SMC-15, 168-175, 1985.
141. M.A.L. Thathachar and P.S. Sastry, "Varieties of learning automata: an overview," *IEEE Transactions on Systems, Man, and Cybernetics,* Part B, vol.32, no.6, 711–722, 2002.
142. P. Tinnakornsrisuphap, S. Vanichpun, and R. La, "Dynamic resource allocation of GPS queues under leaky buckets," *Proceedings of IEEE GLOBECOM*, 3777–3781, 2003.
143. F. Topsoe, "Bounds for entropy and divergence for distributions over a two-element set," *Journal of Inequalities in Pure and Applied Mathematics*, vol.2, issue 2, Article 25, 2001.
144. J.N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning," *Machine Learning*, vol.16, 185–202, 1994.
145. W.A. van den Broek, "Moving horizon control in dynamic games," *Journal of Economic Dynamics and Control.*, vol.26, 937–961, 2002.
146. B. Van Roy, "Neuro-dynamic programming: overview and recent trends," in *Handbook of Markov Decision Processes: Methods and Applications*, E.A. Feinberg and A. Shwartz, eds., Kluwer, Boston, MA, 2002.
147. C.J.C.H. Watkins, "*Q*-learning," *Machine Learning*, vol.8, no.3, 279–292, 1992.
148. R. Weber, "On the Gittins index for multiarmed bandits," *Annals in Applied Probability*, vol.2, 1024–1033, 1992.
149. C. Wells, C. Lusena, and J. Goldsmith, "Genetic algorithms for approximating solutions to POMDPs," Department of Computer Science Technical Report TR-290-99, University of Kentucky, 1999. `http://cs.engr.uky.edu/ goldsmit/papers/gen.ps`.

150. R.M. Wheeler, Jr. and K.S. Narendra, "Decentralized learning in finite Markov chains," *IEEE Transactions on Automatic Control*, vol.31, no.6, 519–526, 1986.
151. J.L. Williams, J.W. Fisher III, and A.S. Willsky, "Importance sampling actor-critic algorithms," *Proceedings of the 2006 American Control Conference*, 1625–1630, 2006.
152. R.J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol.8, 229–256, 1992.
153. D.H. Wolpert, "Finding bounded rational equilibria part I: Iterative focusing," *Proceedings of the Eleventh International Symposium on Dynamic Games and Applications, ISDG (International Society of Dynamic Games) 04,* T. Vincent, ed., 2004. `http://ic.arc.nasa.gov/people/dhw/papers/2.pdf`.
154. G. Wu, E.K.P. Chong, and R.L. Givan, "Burst-level congestion control using hindsight optimization," *IEEE Transactions on Automatic Control*, special issue on Systems and Control Methods for Communication Networks, vol.47, no.6, 979–991, 2002.
155. S. Yakowitz, P. L'Ecuyer, and F. Vázquez-Abad, "Global stochastic optimization with low-dispersion point sets," *Operations Research,* vol.48, 939–950, 2000.
156. Q. Zhang and H. Mühlenbein, "On the convergence of a class of estimation of distribution algorithm," *IEEE Transactions on Evolutionary Computation*, vol.8, no.2, 127-136, 2004.
157. M. Zlochin, M. Birattari, N. Meuleau, and M. Dorigo, "Model-based search for combinatorial optimization: a critical survey," *Annals of Operations Research*, vol.131, 373–395, 2004.

Index