

References

- [ALZ08] Asmuth, J., Littman, M. L. and Zinkov, R., 2008. “Potential-Based Shaping in Model-Based Reinforcement Learning,” Proc. of 23rd AAAI Conference, pp. 604-609.
- [AMS07] Antos, A., Munos, R., and Szepesvari, C., 2007. “Fitted Q- Iteration in Continuous Action-Space MDPs,” Proc. of NIPS, pp. 9-16.
- [Abr90] Abramson, B., 1990. “Expected-Outcome: A General Model of Static Evaluation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, pp. 182-193.
- [AnM79] Anderson, B. D. O., and Moore, J. B., 1979. Optimal Filtering, Prentice-Hall, Englewood Cliffs, N. J.
- [BBD10] Busoniu, L., Babuska, R., De Schutter, B., and Ernst, D., 2010. Reinforcement Learning and Dynamic Programming Using Function Approximators, CRC Press, N. Y.
- [BBG13] Bertazzi, L., Bosco, A., Guerriero, F., and Lagana, D., 2013. “A Stochastic Inventory Routing Problem with Stock-Out,” Transportation Research, Part C, Vol. 27, pp. 89-107.
- [BBM17] Borelli, F., Bemporad, A., and Morari, M., 2017. Predictive Control for Linear and Hybrid Systems, Cambridge Univ. Press, Cambridge, UK.
- [BBN04] Bertsekas, D. P., Borkar, V., and Nedić, A., 2004. “Improved Temporal Difference Methods with Linear Function Approximation,” in Learning and Approximate Dynamic Programming, by J. Si, A. Barto, W. Powell, and D. Wunsch, (Eds.), IEEE Press, N. Y.
- [BBS87] Bean, J. C., Birge, J. R., and Smith, R. L., 1987. “Aggregation in Dynamic Programming,” Operations Research, Vol. 35, pp. 215-220.
- [BCN18] Bottou, L., Curtis, F. E., and Nocedal, J., 2018. “Optimization Methods for Large-Scale Machine Learning,” SIAM Review, Vol. 60, pp. 223-311.
- [BGM95] Bertsekas, D. P., Guerriero, F., and Musmanno, R., 1995. “Parallel Shortest Path Methods for Globally Optimal Trajectories,” High Performance Computing: Technology, Methods, and Applications, (J. Dongarra et al., Eds.), Elsevier.
- [BGM96] Bertsekas, D. P., Guerriero, F., and Musmanno, R., 1996. “Parallel Label Correcting Methods for Shortest Paths,” J. Optimization Theory Appl., Vol. 88, 1996, pp. 297-320.
- [BKM05] de Boer, P. T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. 2005. “A Tutorial on the Cross-Entropy Method,” Annals of Operations Research, Vol. 134, pp. 19-67.

- [BMO14] Boyd, S., Mueller, M. T., O'Donoghue, B., and Wang, Y., 2014. "Performance Bounds and Suboptimal Policies for Multi-Period Investment," *Foundations and Trends in Optimization*, Vol. 1, pp. 1-72.
- [BNO03] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2003. *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA.
- [BPP13] Bhatnagar, S., Prasad, H., and Prashanth, L. A., 2013. *Stochastic Recursive Algorithms for Optimization*, *Lecture Notes in Control and Information Sciences*, Springer, N. Y.
- [BSA83] Barto, A. G., Sutton, R. S., and Anderson, C. W., 1983. "Neuronlike Elements that Can Solve Difficult Learning Control Problems," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 13, pp. 835-846.
- [BTW97] Bertsekas, D. P., Tsitsiklis, J. N., and Wu, C., 1997. "Rollout Algorithms for Combinatorial Optimization," *Heuristics*, Vol. 3, pp. 245-262.
- [BYB94] Bradtke, S. J., Ydstie, B. E., and Barto, A. G., 1994. "Adaptive Linear Quadratic Control Using Policy Iteration," *Proc. IEEE American Control Conference*, Vol. 3, pp. 3475-3479.
- [Bar81] Bar-Shalom, Y., 1981. "Stochastic Dynamic Programming: Caution and Probing," *IEEE Trans. on Automatic Control*, Vol. AC-26, pp. 1184-1195.
- [BeC99] Bertsekas, D. P., and Castanon, D. A., 1999. "Rollout Algorithms for Stochastic Scheduling Problems," *Heuristics*, Vol. 5, pp. 89-108.
- [BeC08] Besse, C., and Chaib-draa, B., 2008. "Parallel Rollout for Online Solution of DEC-POMDPs," *Proc. of 21st International FLAIRS Conference*, pp. 619-624.
- [BeD62] Bellman, R., and Dreyfus, S., 1962. *Applied Dynamic Programming*, Princeton Univ. Press, Princeton, N. J.
- [BeG92] Bertsekas, D. P., and Gallager, R. G., 1992. *Data Networks*, 2nd Edition, Prentice-Hall, Englewood Cliffs, N. J.
- [BeL14] Beyme, S., and Leung, C., 2014. "Rollout Algorithm for Target Search in a Wireless Sensor Network," *80th Vehicular Technology Conference (VTC2014)*, IEEE, pp. 1-5.
- [BeI96] Bertsekas, D. P., and Ioffe, S., 1996. "Temporal Differences-Based Policy Iteration and Applications in Neuro-Dynamic Programming," *Lab. for Info. and Decision Systems Report LIDS-P-2349*, Massachusetts Institute of Technology.
- [BeR71a] Bertsekas, D. P., and Rhodes, I. B., 1971. "On the Minimax Reachability of Target Sets and Target Tubes," *Automatica*, Vol. 7, pp. 233-247.
- [BeR71b] Bertsekas, D. P., and Rhodes, I. B., 1971. "Recursive State Estimation for a Set-Membership Description of the Uncertainty," *IEEE Trans. Automatic Control*, Vol. AC-16, pp. 117-128.
- [BeR73] Bertsekas, D. P., and Rhodes, I. B., 1973. "Sufficiently Informative Functions and the Minimax Feedback Control of Uncertain Dynamic Systems," *IEEE Trans. Automatic Control*, Vol. AC-18, pp. 117-124.
- [BeS78] Bertsekas, D. P., and Shreve, S. E., 1978. *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, N. Y.; republished by Athena Scientific, Belmont, MA, 1996 (can be downloaded from the author's website).
- [BeT89] Bertsekas, D. P., and Tsitsiklis, J. N., 1989. *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, N. J.; republished by Athena

- Scientific, Belmont, MA, 1997.
- [BeT91] Bertsekas, D. P., and Tsitsiklis, J. N., 1991. "An Analysis of Stochastic Shortest Path Problems," *Math. Operations Res.*, Vol. 16, pp. 580-595.
- [BeT96] Bertsekas, D. P., and Tsitsiklis, J. N., 1996. *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.
- [BeT97] Bertsimas, D., and Tsitsiklis, J. N., 1997. *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA.
- [BeT00] Bertsekas, D. P., and Tsitsiklis, J. N., 2000. "Gradient Convergence of Gradient Methods with Errors," *SIAM J. on Optimization*, Vol. 36, pp. 627-642.
- [BeT08] Bertsekas, D. P., and Tsitsiklis, J. N., 2008. *Introduction to Probability*, 2nd Edition, Athena Scientific, Belmont, MA.
- [BeY09] Bertsekas, D. P., and Yu, H., 2009. "Projected Equation Methods for Approximate Solution of Large Linear Systems," *J. of Computational and Applied Mathematics*, Vol. 227, pp. 27-50.
- [BeY10] Bertsekas, D. P., and Yu, H., 2010. "Asynchronous Distributed Policy Iteration in Dynamic Programming," *Proc. of Allerton Conf. on Communication, Control and Computing*, Allerton Park, Ill, pp. 1368-1374.
- [BeY12] Bertsekas, D. P., and Yu, H., 2010. "Q-Learning and Enhanced Policy Iteration in Discounted Dynamic Programming," *Math. of Operations Research*, Vol. 37, pp. 66-94.
- [BeY16] Bertsekas, D. P., and Yu, H., 2016. "Stochastic Shortest Path Problems Under Weak Conditions," *Lab. for Information and Decision Systems Report LIDS-2909*, MIT.
- [Bel57] Bellman, R., 1957. *Dynamic Programming*, Princeton University Press, Princeton, N. J.
- [Ber71] Bertsekas, D. P., 1971. "Control of Uncertain Systems With a Set-Membership Description of the Uncertainty," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA (available from the author's website).
- [Ber72a] Bertsekas, D. P., 1972. "Infinite Time Reachability of State Space Regions by Using Feedback Control," *IEEE Trans. Automatic Control*, Vol. AC-17, pp. 604-613.
- [Ber72b] Bertsekas, D. P., 1972. "On the Solution of Some Minimax Control Problems," *Proc. 1972 IEEE Decision and Control Conf.*, New Orleans, LA.
- [Ber75] Bertsekas, D. P., 1975. "Convergence of Discretization Procedures in Dynamic Programming," *IEEE Trans. Automatic Control*, Vol. AC-20, pp. 415-419.
- [Ber76] Bertsekas, D. P., 1976. *Dynamic Programming and Stochastic Control*, Academic Press, N. Y.
- [Ber82] Bertsekas, D. P., 1982. "Distributed Dynamic Programming," *IEEE Trans. Automatic Control*, Vol. AC-27, pp. 610-616.
- [Ber91] Bertsekas, D. P., 1991. *Linear Network Optimization: Algorithms and Codes*, M.I.T. Press, Cambridge, MA.
- [Ber95] Bertsekas, D. P., 1995. "A Counterexample to Temporal Differences Learning," *Neural Computation*, Vol. 7, pp. 270-279.
- [Ber97a] Bertsekas, D. P., 1997. "A New Class of Incremental Gradient Methods for Least Squares Problems," *SIAM J. on Optimization*, Vol. 7, pp. 913-926.
- [Ber97b] Bertsekas, D. P., 1997. "Differential Training of Rollout Policies," *Proc. of the*

- 35th Allerton Conference on Communication, Control, and Computing, Allerton Park, Ill.
- [Ber98] Bertsekas, D. P., 1998. *Network Optimization: Continuous and Discrete Models*, Athena Scientific, Belmont, MA.
- [Ber05a] Bertsekas, D. P., 2005. “Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC,” *European J. of Control*, Vol. 11, pp. 310-334.
- [Ber05b] Bertsekas, D. P., 2005. “Rollout Algorithms for Constrained Dynamic Programming,” LIDS Report 2646, MIT.
- [Ber07] Bertsekas, D. P., 2007. “Separable Dynamic Programming and Approximate Decomposition Methods,” *IEEE Trans. on Aut. Control*, Vol. 52, pp. 911-916.
- [Ber09] Bertsekas, D. P., 2009. *Convex Optimization Theory*, Athena Scientific, Belmont, MA.
- [Ber10] Bertsekas, D. P., 2010. “Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey,” Lab. for Information and Decision Systems Report LIDS-P-2848, MIT; this is an extended version of a paper by the same title in the edited volume *Optimization for Machine Learning*, by S. Sra, S. Nowozin, and S. J. Wright, MIT Press, Cambridge, MA, 2012, pp. 85-119.
- [Ber11a] Bertsekas, D. P., 2011. “Approximate Policy Iteration: A Survey and Some New Methods,” *J. of Control Theory and Applications*, Vol. 9, pp. 310-335.
- [Ber11b] Bertsekas, D. P., 2011. “ λ -Policy Iteration: A Review and a New Implementation,” Lab. for Information and Decision Systems Report LIDS-P-2874, MIT; in *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, by F. Lewis and D. Liu (eds.), IEEE Press, Computational Intelligence Series, 2012.
- [Ber11c] Bertsekas, D. P., 2011. “Temporal Difference Methods for General Projected Equations,” *IEEE Trans. on Aut. Control*, Vol. 56, pp. 2128-2139.
- [Ber12] Bertsekas, D. P., 2012. *Dynamic Programming and Optimal Control*, Vol. II, 4th Edition, Athena Scientific, Belmont, MA.
- [Ber13] Bertsekas, D. P., 2013. “Rollout Algorithms for Discrete Optimization: A Survey,” *Handbook of Combinatorial Optimization*, Springer.
- [Ber14] Bertsekas, D. P., 2014. “Robust Shortest Path Planning and Semicontractive Dynamic Programming,” Lab. for Information and Decision Systems Report LIDS-P-2915, MIT; arXiv preprint arXiv:1608.01670; to appear in *Naval Research Logistics*.
- [Ber15a] Bertsekas, D. P., 2015. *Convex Optimization Algorithms*, Athena Scientific, Belmont, MA.
- [Ber15b] Bertsekas, D. P., 2015. “Incremental Aggregated Proximal and Augmented Lagrangian Algorithms,” Lab. for Information and Decision Systems Report LIDS-P-3176, MIT, July 2015; revised September 2015; arXiv preprint arXiv:1507.1365936.
- [Ber16a] Bertsekas, D. P., 2016. *Nonlinear Programming*, 3rd Edition, Athena Scientific, Belmont, MA.
- [Ber16b] Bertsekas, D. P., 2016. “Affine Monotonic and Risk-Sensitive Models in Dynamic Programming,” Lab. for Information and Decision Systems Report LIDS-3204, MIT; arXiv preprint arXiv:1608.01393.
- [Ber16c] Bertsekas, D. P., 2016. “Proximal Algorithms and Temporal Differences for Large Linear Systems: Extrapolation, Approximation, and Simulation,” arXiv preprint arXiv:1610.05427.

- [Ber17] Bertsekas, D. P., 2017. *Dynamic Programming and Optimal Control*, Vol. I, 2nd Edition, Athena Scientific, Belmont, MA.
- [Ber18a] Bertsekas, D. P., 2018. *Abstract Dynamic Programming*, 2nd Edition, Athena Scientific, Belmont, MA.
- [Ber18b] Bertsekas, D. P., 2018. “Feature-Based Aggregation and Deep Reinforcement Learning: A Survey and Some New Implementations,” Lab. for Information and Decision Systems Report, MIT; arXiv preprint arXiv:1804.04577 (a version is published in IEEE/CAA Journal of Automatica Sinica).
- [Ber18c] Bertsekas, D. P., 2018. “Biased Aggregation, Rollout, and Enhanced Policy Improvement for Reinforcement Learning,” Lab. for Information and Decision Systems Report, MIT.
- [Ber18d] Bertsekas, D. P., 2018. “Proximal Algorithms and Temporal Difference Methods for Solving Fixed Point Problems,” *Computational Optim. Appl.*, Vol. 70, pp. 709-736.
- [BiL97] Birge, J. R., and Louveaux, 1997. *Introduction to Stochastic Programming*, Springer, New York, N. Y.
- [Bis95] Bishop, C. M., 1995. *Neural Networks for Pattern Recognition*, Oxford University Press, N. Y.
- [Bor08] Borkar, V. S., 2008. *Stochastic Approximation: A Dynamical Systems Viewpoint*, Cambridge Univ. Press.
- [Boy02] Boyan, J. A., 2002. “Technical Update: Least-Squares Temporal Difference Learning,” *Machine Learning*, Vol. 49, pp. 1-15.
- [BrB96] Bradtko, S. J., and Barto, A. G., 1996. “Linear Least-Squares Algorithms for Temporal Difference Learning,” *Machine Learning*, Vol. 22, pp. 33-57.
- [CFH05] Chang, H. S., Hu, J., Fu, M. C., and Marcus, S. I., 2005. “An Adaptive Sampling Algorithm for Solving Markov Decision Processes,” *Operations Research*, Vol. 53, pp. 126-139.
- [CFH13] Chang, H. S., Hu, J., Fu, M. C., and Marcus, S. I., 2013. *Simulation-Based Algorithms for Markov Decision Processes*, (2nd Ed.), Springer, N. Y.
- [CFH16] Chang, H. S., Hu, J., Fu, M. C., and Marcus, S. I., 2016. “Google DeepMind’s AlphaGo,” *ORMS Today*, Informs, Vol. 43.
- [CGC04] Chang, H. S., Givan, R. L., and Chong, E. K. P., 2004. “Parallel Rollout for Online Solution of Partially Observable Markov Decision Processes,” *Discrete Event Dynamic Systems*, Vol. 14, pp. 309-341.
- [CHH02] Campbell, M., Hoane, A. J. and Hsu, F. H., 2002. “Deep Blue,” *Artificial Intelligence*, Vol. 134, pp. 57-83.
- [CaB04] Camacho, E. F., and Bordons, C., 2004. *Model Predictive Control*, 2nd Edition, Springer, New York, N. Y.
- [Cao07] Cao, X. R., 2007. *Stochastic Learning and Optimization: A Sensitivity-Based Approach*, Springer, N. Y.
- [ChT89] Chow, C.-S., and Tsitsiklis, J. N., 1989. “The Complexity of Dynamic Programming,” *J. of Complexity*, Vol. 5, pp. 466-488.
- [ChT91] Chow, C.-S., and Tsitsiklis, J. N., 1991. “An Optimal One-Way Multigrid Algorithm for Discrete-Time Stochastic Control,” *IEEE Trans. on Automatic Control*, Vol. AC-36, 1991, pp. 898-914.

- [ChV12] Chacon, A., and Vladimirovsky, A., 2012. "Fast Two-Scale Methods for Eikonal Equations," *SIAM J. on Scientific Computing*, Vol. 34, pp. A547-A578.
- [ChV13] Chacon, A., and Vladimirovsky, A., 2013. "A Parallel Heap-Cell Method for Eikonal Equations," *arXiv preprint arXiv:1306.4743*.
- [ChV15] Chacon, A., and Vladimirovsky, A., 2015. "A Parallel Two-Scale Method for Eikonal Equations," *SIAM J. on Scientific Computing*, Vol. 37, pp. A156-A180.
- [Cou06] Coulom, R., 2006. "Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search," *International Conference on Computers and Games*, Springer, pp. 72-83.
- [Cyb89] Cybenko, 1989. "Approximation by Superpositions of a Sigmoidal Function," *Math. of Control, Signals, and Systems*, Vol. 2, pp. 303-314.
- [DFV03] de Farias, D. P., and Van Roy, B., 2003. "The Linear Programming Approach to Approximate Dynamic Programming," *Operations Research*, Vol. 51, pp. 850-865.
- [DFV04] de Farias, D. P., and Van Roy, B., 2004. "On Constraint Sampling in the Linear Programming Approach to Approximate Dynamic Programming," *Mathematics of Operations Research*, Vol. 29, pp. 462-478.
- [DeF04] De Farias, D. P., 2004. "The Linear Programming Approach to Approximate Dynamic Programming," in *Learning and Approximate Dynamic Programming*, by J. Si, A. Barto, W. Powell, and D. Wunsch, (Eds.), IEEE Press, N. Y.
- [DeK11] Devlin, S., and Kudenko, D., 2011. "Theoretical Considerations of Potential-Based Reward Shaping for Multi-Agent Systems," In *Proceedings of AAMAS*.
- [Dre65] Dreyfus, S. D., 1965. *Dynamic Programming and the Calculus of Variations*, Academic Press, N. Y.
- [EGW06] Ernst, D., Geurts, P., and Wehenkel, L., 2006. "Tree-Based Batch Mode Reinforcement Learning," *J. of Machine Learning Research*, Vol. 6, pp. 503-556.
- [Eck68] Eckles, J. E., 1968. "Optimum Maintenance with Incomplete Information," *Operations Research*, Vol. 16, pp. 1058-1067.
- [FGL13] Festa, P., Guerriero, F., Lagana, D. and Musmanno, R., 2013. "Solving the Shortest Path Tour Problem," *European J. of Operational Research*, Vol. 230, pp. 464-474.
- [FYG06] Fern, A., Yoon, S. and Givan, R., 2006. "Approximate Policy Iteration with a Policy Language Bias: Solving Relational Markov Decision Processes," *J. of Artificial Intelligence Research*, Vol. 25, pp. 75-118.
- [FeV02] Ferris, M. C., and Voelker, M. M., 2002. "Neuro-Dynamic Programming for Radiation Treatment Planning," *Numerical Analysis Group Research Report NA-02/06*, Oxford University Computing Laboratory, Oxford University.
- [FeV04] Ferris, M. C., and Voelker, M. M., 2004. "Fractionation in Radiation Treatment Planning," *Mathematical Programming B*, Vol. 102, pp. 387-413.
- [Fer10] Fernau, H., 2010. "Minimum Dominating Set of Queens: A Trivial Programming Exercise?," *Discrete Applied Mathematics*, Vol. 158, pp. 308-318.
- [Fu18] Fu, M. C., 2018. "Markov Decision Processes, AlphaGo, and Monte Carlo Tree Search: Back to the Future," *INFORMS Tutorials in Operations Research*.
- [Fun89] Funahashi, K., 1989. "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, Vol. 2, pp. 183-192.

- [GBC16] Goodfellow, I., Bengio, J., and Courville, A., Deep Learning, MIT Press, Cambridge, MA.
- [GGS13] Gabillon, V., Ghavamzadeh, M., and Scherrer, B., 2013. "Approximate Dynamic Programming Finally Performs Well in the Game of Tetris," in Advances in Neural Information Processing Systems, pp. 1754-1762.
- [GOP15a] Gurbuzbalaban, M., Ozdaglar, A., and Parrilo, P., 2015. "On the Convergence Rate of Incremental Aggregated Gradient Algorithms," arXiv preprint arXiv:1506.02081.
- [GOP15b] Gurbuzbalaban, M., Ozdaglar, A., and Parrilo, P., 2015. "Convergence Rate of Incremental Gradient and Newton Methods," arXiv preprint arXiv:1510.08562.
- [GOP15c] Gurbuzbalaban, M., Ozdaglar, A., and Parrilo, P., 2015. "Why Random Reshuffling Beats Stochastic Gradient Descent," arXiv preprint arXiv:1510.08560.
- [GTO15] Goodson, J. C., Thomas, B. W., and Ohlmann, J. W., 2015. "Restocking-Based Rollout Policies for the Vehicle Routing Problem with Stochastic Demand and Duration Limits," Transportation Science, Vol. 50, pp. 591-607.
- [Gal13] Gallager, R. G., 2013. Stochastic Processes, Cambridge Univ. Press.
- [GoR85] Gonzalez, R., and Rofman, E., 1985. "On Deterministic Control Problems: An Approximation Procedure for the Optimal Cost, Parts I, II," SIAM J. Control Optimization, Vol. 23, pp. 242-285.
- [GoS84] Goodwin, G. C., and Sin, K. S. S., 1984. Adaptive Filtering, Prediction, and Control, Prentice-Hall, Englewood Cliffs, N. J.
- [Gor95] Gordon, G. J., 1995. "Stable Function Approximation in Dynamic Programming," in Machine Learning: Proceedings of the Twelfth International Conference, Morgan Kaufmann, San Francisco, CA.
- [Gos15] Gosavi, A., 2015. Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning, 2nd Edition, Springer, N. Y.
- [Grz17] Grzes, M., 2017. "Reward Shaping in Episodic Reinforcement Learning," in Proc. of the 16th Conference on Autonomous Agents and MultiAgent Systems, pp. 565-573.
- [GuM03] Guerriero, F., and Mancini, M., 2003. "A Cooperative Parallel Rollout Algorithm for the Sequential Ordering Problem," Parallel Computing, Vol. 29, pp. 663-677.
- [HJG16] Huang, Q., Jia, Q. S., and Guan, X., 2016. "Robust Scheduling of EV Charging Load with Uncertain Wind Power Integration, IEEE Transactions on Smart Grid.
- [HNR72] Hart, P. E., Nilsson, N. J. and Raphael, B., 1972. "Correction to a Formal Basis for the Heuristic Determination of Minimum Cost Paths," ACM SIGART Bulletin, Vol. 37, pp. 28-29.
- [HPC96] Helmsen, J., Puckett, E. G., Colella, P., and Dorr, M., 1996. "Two New Methods for Simulating Photolithography Development," SPIE's 1996 International Symposium on Microlithography, pp. 253-261.
- [HSW89] Hornick, K., Stinchcombe, M., and White, H., 1989. "Multilayer Feedforward Networks are Universal Approximators," Neural Networks, Vol. 2, pp. 359-159.
- [Hay09] Haykin, S., 2009. Neural Networks and Learning Machines, (3rd Edition), Prentice-Hall, Englewood-Cliffs, N. J.
- [HoK71] Hoffman, K., and Kunze, R., 1971. Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, N. J.

- [IEE71] IEEE Trans. Automatic Control, 1971. Special Issue on Linear-Quadratic Gaussian Problem, Vol. AC-16.
- [Iva68] Ivakhnenko, A. G., 1968. "The Group Method of Data Handling: A Rival of the Method of Stochastic Approximation," Soviet Automatic Control, Vol. 13, pp. 43-55.
- [Iva71] Ivakhnenko, A. G., 1971. "Polynomial Theory of Complex Systems," IEEE Transactions on Systems, Man and Cybernetics, Vol. 4, pp. 364-378.
- [Jon90] Jones, L. K., 1990. "Constructive Approximations for Neural Networks by Sigmoidal Functions," Proceedings of the IEEE, Vol. 78, pp. 1586-1589.
- [KGB82] Kimemia, J., Gershwin, S. B., and Bertsekas, D. P., 1982. "Computation of Production Control Policies by a Dynamic Programming Technique," in Analysis and Optimization of Systems, A. Bensoussan and J. L. Lions (eds.), Springer, N. Y., pp. 243-269.
- [KLC98] Kaelbling, L.P., Littman, M. L., and Cassandra, A. R., 1998. "Planning and Acting in Partially Observable Stochastic Domains," Artificial Intelligence, Vol. 101, pp. 99-134.
- [KaD66] Karush, W., and Dear, E. E., 1966. "Optimal Stimulus Presentation Strategy for a Stimulus Sampling Model of Learning," J. Math. Psychology, Vol. 3, pp. 15-47.
- [KaK58] Kalman, R. E., and Koepcke, R. W., 1958. "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," Trans. ASME, Vol. 80, pp. 1820-1826.
- [KaW94] Kall, P., and Wallace, S. W., 1994. Stochastic Programming, Wiley, Chichester, UK.
- [KeG88] Keerthi, S. S., and Gilbert, E. G., 1988. "Optimal, Infinite Horizon Feedback Laws for a General Class of Constrained Discrete Time Systems: Stability and Moving-Horizon Approximations," J. Optimization Theory Appl., Vol. 57, pp. 265-293.
- [Kim82] Kimemia, J., 1982. "Hierarchical Control of Production in Flexible Manufacturing Systems," Ph.D. Thesis, Dep. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
- [KoC15] Kouvaritakis, B., and Cannon, M., 2015. Model Predictive Control: Classical, Robust and Stochastic, Springer, N. Y.
- [KoS06] Kocsis, L., and Szepesvari, C., 2006. "Bandit Based Monte-Carlo Planning," Proc. of 17th European Conference on Machine Learning, Berlin, pp. 282-293.
- [KuL82] Kumar, P. R., and Lin, W., 1982. "Optimal Adaptive Controllers for Unknown Markov Chains," IEEE Trans. Automatic Control, Vol. AC-27, pp. 765-774.
- [KuV86] Kumar, P. R., and Varaiya, P. P., 1986. Stochastic Systems: Estimation, Identification, and Adaptive Control, Prentice-Hall, Englewood Cliffs, N. J.
- [KuY97] Kushner, H. J., and Yin, G., 1997. Stochastic Approximation Methods, Springer-Verlag, N. Y.
- [Kum83] Kumar, P. R., 1983. "Optimal Adaptive Control of Linear-Quadratic-Gaussian Systems," SIAM J. on Control and Optimization, Vol. 21, pp. 163-178.
- [Kum85] Kumar, P. R., 1985. "A Survey of Some Results in Stochastic Adaptive Control," SIAM J. on Control and Optimization, Vol. 23, pp. 329-380.
- [LGW16] Lan, Y., Guan, X., and Wu, J., 2016. "Rollout Strategies for Real-Time Multi-Energy Scheduling in Microgrid with Storage System," IET Generation, Transmission and Distribution, Vol. 10, pp. 688-696.

- [LLL08] Lewis, F. L., Liu, D., and Lendaris, G. G., 2008. Special Issue on Adaptive Dynamic Programming and Reinforcement Learning in Feedback Control, *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, Vol. 38, No. 4.
- [LLP93] Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S., 1993. "Multilayer Feed-forward Networks with a Nonpolynomial Activation Function can Approximate any Function," *Neural Networks*, Vol. 6, pp. 861-867.
- [LWW17] Liu, D., Wei, Q., Wang, D., and Yang, X., 2017. *Adaptive Dynamic Programming with Applications in Optimal Control*, Springer, Berlin.
- [LeL13] Lewis, F. L., and Liu, D., (Eds), 2013. *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, Wiley, Hoboken, N. J.
- [LiR06] Lincoln, B., and Rantzer, A., 2006. "Relaxing Dynamic Programming," *IEEE Trans. Automatic Control*, Vol. 51, pp. 1249-1260.
- [LiW15] Li, H. and Womer, N. K., 2015. "Solving Stochastic Resource-Constrained Project Scheduling Problems by Closed-Loop Approximate Dynamic Programming," *European J. of Operational Research*, Vol. 246, pp. 20-33.
- [LoS01] Longstaff, F. A., and Schwartz, E. S., 2001. "Valuing American Options by Simulation: A Simple Least-Squares Approach," *Review of Financial Studies*, Vol. 14, pp. 113-147.
- [Lov91a] Lovejoy, W. S., 1991. "Computationally Feasible Bounds for Partially Observed Markov Decision Processes," *Operations Research*, Vol. 39, pp. 1621-175.
- [Lov91b] Lovejoy, W. S., 1991. "A Survey of Algorithmic Methods for Partially Observed Markov Decision Processes," *Annals of Operations Research*, Vol. 18, pp. 47-66.
- [MHK98] Meuleau, N., Hauskrecht, M., Kim, K.-E., Peshkin, L., Kaelbling, L. K., and Dean, T., 1998. "Solving Very Large Weakly Coupled Markov Decision Processes," *Proc. of the Fifteenth National Conference on Artificial Intelligence*, Madison, WI, pp. 165-172.
- [MMB02] McGovern, A., Moss, E., and Barto, A., 2002. "Building a Basic Building Block Scheduler Using Reinforcement Learning and Rollouts," *Machine Learning*, Vol. 49, pp. 141-160.
- [MPP04] Meloni, C., Pacciarelli, D., and Pranzo, M., 2004. "A Rollout Metaheuristic for Job Shop Scheduling Problems," *Annals of Operations Research*, Vol. 131, pp. 215-235.
- [MYF03] Moriyama, H., Yamashita, N., and Fukushima, M., 2003. "The Incremental Gauss-Newton Algorithm with Adaptive Stepsize Rule," *Computational Optimization and Applications*, Vol. 26, pp. 107-141.
- [MaJ15] Mastin, A., and Jaillet, P., 2015. "Average-Case Performance of Rollout Algorithms for Knapsack Problems," *J. of Optimization Theory and Applications*, Vol. 165, pp. 964-984.
- [Mac02] Maciejowski, J. M., 2002. *Predictive Control with Constraints*, Addison-Wesley, Reading, MA.
- [May14] Mayne, D. Q., 2014. "Model Predictive Control: Recent Developments and Future Promise," *Automatica*, Vol. 50, pp. 2967-2986.
- [Mey07] Meyn, S., 2007. *Control Techniques for Complex Networks*, Cambridge Univ. Press, N. Y.
- [MoL99] Morari, M., and Lee, J. H., 1999. "Model Predictive Control: Past, Present, and Future," *Computers and Chemical Engineering*, Vol. 23, pp. 667-682.

- [MuS08] Munos, R., and Szepesvari, C., 2008. "Finite-Time Bounds for Fitted Value Iteration," *J. of Machine Learning Research*, Vol. 1, pp. 815-857.
- [Mun14] Munos, R., 2014. "From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning," *Foundations and Trends in Machine Learning*, Vol. 7, pp. 1-129.
- [NHR99] Ng, A. Y., Harada, D., and Russell, S. J., 1999. "Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping," in *Proc. of the 16th International Conference on Machine Learning*, pp. 278-287.
- [NeB00] Nedić, A., and Bertsekas, D. P., 2000. "Convergence Rate of Incremental Subgradient Algorithms," *Stochastic Optimization: Algorithms and Applications*, S. Uryasev and P. M. Pardalos, Eds., Kluwer, pp. 263-304.
- [NeB01] Nedić, A., and Bertsekas, D. P., 2001. "Incremental Subgradient Methods for Nondifferentiable Optimization," *SIAM J. on Optimization*, Vol. 12, pp. 109-138.
- [NeB03] Nedić, A., and Bertsekas, D. P., 2003. "Least-Squares Policy Evaluation Algorithms with Linear Function Approximation," *J. of Discrete Event Systems*, Vol. 13, pp. 79-110.
- [New75] Newborn, M., 1975. *Computer Chess*, Academic Press, N. Y.
- [Nil71] Nilsson, N. J., 1971. *Problem-Solving Methods in Artificial Intelligence*, McGraw-Hill, N. Y.
- [Nil80] Nilsson, N. J., 1980. *Principles of Artificial Intelligence*, Morgan-Kaufmann, San Mateo, Ca.
- [OrS02] Ormoneit, D., and Sen, S., 2002. "Kernel-Based Reinforcement Learning," *Machine Learning*, Vol. 49, pp. 161-178.
- [PBT98] Polymenakos, L. C., Bertsekas, D. P., and Tsitsiklis, J. N., 1998. "Efficient Algorithms for Continuous-Space Shortest Path Problems," *IEEE Trans. on Automatic Control*, Vol. 43, pp. 278-283.
- [Pat01] Patek, S. D., 2001. "On Terminating Markov Decision Processes with a Risk Averse Objective Function," *Automatica*, Vol. 37, pp. 1379-1386.
- [Pea84] Pearl, J., 1984. *Heuristics*, Addison-Wesley, Reading, MA.
- [Pow11] Powell, W. B., 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd Edition, J. Wiley and Sons, Hoboken, N. J.
- [Pre95] Prekopa, A., 1995. *Stochastic Programming*, Kluwer, Boston.
- [Put94] Puterman, M. L., 1994. *Markovian Decision Problems*, J. Wiley, N. Y.
- [RSS12] Runarsson, T. P., Schoenauer, M., and Sebag, M., 2012. "Pilot, Rollout and Monte Carlo Tree Search Methods for Job Shop Scheduling," in *Learning and Intelligent Optimization*, Springer, Berlin, pp. 160-174.
- [Ros12] Ross, S. M., 2012. *Simulation*, 5th Edition, Academic Press, Orlando, Fla.
- [RuK04] Rubinstein, R. Y., and Kroese, D. P., 2004. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization*, Springer, N. Y.
- [RuK17] Rubinstein, R. Y., and Kroese, D. P., 2017. *Simulation and the Monte Carlo Method*, 3rd Edition, J. Wiley, N. Y.
- [SBP04] Si, J., Barto, A., Powell, W., and Wunsch, D., (Eds.) 2004. *Learning and Approximate Dynamic Programming*, IEEE Press, N. Y.

- [SGC02] Savagaonkar, U., Givan, R., and Chong, E. K. P., 2002. "Sampling Techniques for Zero-Sum, Discounted Markov Games," in Proc. 40th Allerton Conference on Communication, Control and Computing, Monticello, Ill.
- [SGG15] Scherrer, B., Ghavamzadeh, M., Gabillon, V., Lesner, B., and Geist, M., 2015. "Approximate Modified Policy Iteration and its Application to the Game of Tetris," J. of Machine Learning Research, Vol. 16, pp. 1629-1676.
- [SHB15] Simroth, A., Holfeld, D., and Brunsch, R., 2015. "Job Shop Production Planning under Uncertainty: A Monte Carlo Rollout Approach," Proc. of the International Scientific and Practical Conference, Vol. 3, pp. 175-179.
- [SHM16] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., and Dieleman, S., 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search," Nature, Vol. 529, pp. 484-489.
- [SHS17] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T. and Lillicrap, T., 2017. "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," arXiv preprint arXiv:1712.01815.
- [SLJ13] Sun, B., Luh, P. B., Jia, Q. S., Jiang, Z., Wang, F., and Song, C., 2013. "Building Energy Management: Integrated Control of Active and Passive Heating, Cooling, Lighting, Shading, and Ventilation Systems," IEEE Trans. on Automation Science and Engineering, Vol. 10, pp. 588-602.
- [SZL08] Sun, T., Zhao, Q., Lun, P., and Tomastik, R., 2008. "Optimization of Joint Replacement Policies for Multipart Systems by a Rollout Framework," IEEE Transactions on Automation Science and Engineering, Vol. 5, pp. 609-619.
- [Sch97] Schaeffer, J., 1997. One Jump Ahead, Springer, N. Y.
- [Sch13] Scherrer, B., 2013. "Performance Bounds for Lambda Policy Iteration and Application to the Game of Tetris," J. of Machine Learning Research, Vol. 14, pp. 1181-1227.
- [Sch15] Schmidhuber, J., 2015. "Deep Learning in Neural Networks: An Overview," Neural Networks, pp. 85-117.
- [Sec00] Secomandi, N., 2000. "Comparing Neuro-Dynamic Programming Algorithms for the Vehicle Routing Problem with Stochastic Demands," Computers and Operations Research, Vol. 27, pp. 1201-1225.
- [Sec01] Secomandi, N., 2001. "A Rollout Policy for the Vehicle Routing Problem with Stochastic Demands," Operations Research, Vol. 49, pp. 796-802.
- [Sec03] Secomandi, N., 2003. "Analysis of a Rollout Approach to Sequencing Problems with Stochastic Routing Applications," J. of Heuristics, Vol. 9, pp. 321-352.
- [Sha50] Shannon, C., 1950. "Programming a Digital Computer for Playing Chess," Phil. Mag., Vol. 41, pp. 356-375.
- [SIL91] Slotine, J.-J. E., and Li, W., Applied Nonlinear Control, Prentice-Hall, Englewood Cliffs, N. J.
- [Spa03] Spall, J. C., 2003. Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, J. Wiley, Hoboken, N. J.
- [StW91] Stewart, B. S., and White, C. C., 1991. "Multiobjective A*," J. ACM, Vol. 38, pp. 775-814.
- [Str76] Strang, G., 1976. Linear Algebra and its Applications, Academic Press, N. Y.

- [SuB18] Sutton, R., and Barto, A. G., 2018. Reinforcement Learning, 2nd Edition, MIT Press, Cambridge, MA.
- [Sut88] Sutton, R. S., 1988. "Learning to Predict by the Methods of Temporal Differences," *Machine Learning*, Vol. 3, pp. 9-44.
- [SzL06] Szita, I., and Lorinz, A., 2006. "Learning Tetris Using the Noisy Cross-Entropy Method," *Neural Computation*, Vol. 18, pp. 2936-2941.
- [Sze10] Szepesvari, C., 2010. Algorithms for Reinforcement Learning, Morgan and Claypool Publishers, San Francisco, CA.
- [TGL13] Tesauro, G., Gondek, D. C., Lenchner, J., Fan, J., and Prager, J. M., 2013. "Analysis of Watson's Strategies for Playing Jeopardy!," *J. of Artificial Intelligence Research*.
- [TeG96] Tesauro, G., and Galperin, G. R., 1996. "On-Line Policy Improvement Using Monte Carlo Search," presented at the 1996 Neural Information Processing Systems Conference, Denver, CO; also in M. Mozer et al. (eds.), *Advances in Neural Information Processing Systems 9*, MIT Press (1997).
- [ThS09] Thiery, C., and Scherrer, B., 2009. "Improvements on Learning Tetris with Cross Entropy," *International Computer Games Association Journal*, Vol. 32, pp. 23-33.
- [TeG96] Tesauro, G., and Galperin, G. R., 1996. "On-Line Policy Improvement Using Monte Carlo Search," presented at the 1996 Neural Information Processing Systems Conference, Denver, CO; also in M. Mozer et al. (eds.), *Advances in Neural Information Processing Systems 9*, MIT Press (1997).
- [Tes89a] Tesauro, G. J., 1989. "Neurogammon Wins Computer Olympiad," *Neural Computation*, Vol. 1, pp. 321-323.
- [Tes89b] Tesauro, G. J., 1989. "Connectionist Learning of Expert Preferences by Comparison Training," in *Advances in Neural Information Processing Systems*, pp. 99-106.
- [Tes92] Tesauro, G. J., 1992. "Practical Issues in Temporal Difference Learning," *Machine Learning*, Vol. 8, pp. 257-277.
- [Tes94] Tesauro, G. J., 1994. "TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play," *Neural Computation*, Vol. 6, pp. 215-219.
- [Tes95] Tesauro, G. J., 1995. "Temporal Difference Learning and TD-Gammon," *Communications of the ACM*, Vol. 38, pp. 58-68.
- [Tes01] Tesauro, G. J., 2001. "Comparison Training of Chess Evaluation Functions," in *Machines that Learn to Play Games*, Nova Science Publishers, pp. 117-130.
- [Tes02] Tesauro, G. J., 2002. "Programming Backgammon Using Self-Teaching Neural Nets," *Artificial Intelligence*, Vol. 134, pp. 181-199.
- [ThS09] Thiery, C., and Scherrer, B., 2009. "Improvements on Learning Tetris with Cross-Entropy," *International Computer Games Association J.*, Vol. 32, pp. 23-33.
- [The54] Theil, H., 1954. "Econometric Models and Welfare Maximization," *Weltwirtsch. Arch.*, Vol. 72, pp. 60-83.
- [TsV96] Tsitsiklis, J. N., and Van Roy, B., 1996. "Feature-Based Methods for Large-Scale Dynamic Programming," *Machine Learning*, Vol. 22, pp. 59-94.
- [TsV97] Tsitsiklis, J. N., and Van Roy, B., 1997. "An Analysis of Temporal-Difference Learning with Function Approximation," *IEEE Trans. on Aut. Control*, Vol. 42, pp. 674-690.

- [TsV99a] Tsitsiklis, J. N., and Van Roy, B., 1999. "Average Cost Temporal-Difference Learning," *Aut. a*, Vol. 35, pp. 1799-1808.
- [TsV99b] Tsitsiklis, J. N., and Van Roy, B., 1999. "Optimal Stopping of Markov Processes: Hilbert Space Theory, Approximation Algorithms, and an Application to Pricing Financial Derivatives", *IEEE Trans. on Aut. Control*, Vol. 44, pp. 1840-1851.
- [TsV01] Tsitsiklis, J. N., and Van Roy, B., 2001. "Regression Methods for Pricing Complex American-Style Options," *IEEE Trans. on Neural Networks*, Vol. 12, pp. 694-703.
- [Tse98] Tseng, P., 1998. "Incremental Gradient(-Projection) Method with Momentum Term and Adaptive Stepsize Rule," *SIAM J. on Optimization*, Vol. 8, pp. 506-531.
- [Tsi84a] Tsitsiklis, J. N., 1984. "Convexity and Characterization of Optimal Policies in a Dynamic Routing Problem," *J. Optimization Theory Appl.*, Vol. 44, pp. 105-136.
- [Tsi84b] Tsitsiklis, J. N., 1984. "Periodic Review Inventory Systems with Continuous Demand and Discrete Order Sizes," *Management Sci.*, Vol. 30, pp. 1250-1254.
- [Tsi87] Tsitsiklis, J. N., 1987. "Analysis of a Multiaccess Control Scheme," *IEEE Trans. Automatic Control*, Vol. AC-32, pp. 1017-1020.
- [Tsi94] Tsitsiklis, J. N., 1994. "Asynchronous Stochastic Approximation and Q-Learning," *Machine Learning*, Vol. 16, pp. 185-202.
- [Tsi95] Tsitsiklis, J. N., 1995. "Efficient Algorithms for Globally Optimal Trajectories," *IEEE Trans. Automatic Control*, Vol. AC-40, pp. 1528-1538.
- [TuP03] Tu, F., and Pattipati, K. R., 2003. "Rollout Strategies for Sequential Fault Diagnosis," *IEEE Trans. on Systems, Man and Cybernetics, Part A*, pp. 86-99.
- [VVL13] Vrabie, D., Vamvoudakis, K. G., and Lewis, F. L., 2013. *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*, The Institution of Engineering and Technology, London.
- [Vla08] Vladimirsky, A., 2008. "Label-Setting Methods for Multimode Stochastic Shortest Path Problems on Graphs," *Math. of Operations Research*, Vol. 33, pp. 821-838.
- [WCG03] Wu, G., Chong, E. K. P., and Givan, R. L., 2003. "Congestion Control Using Policy Rollout," *Proc. 2nd IEEE CDC*, Maui, Hawaii, pp. 4825-4830.
- [WaB13a] Wang, M., and Bertsekas, D. P., 2013. "Stabilization of Stochastic Iterative Methods for Singular and Nearly Singular Linear Systems," *Mathematics of Operations Research*, Vol. 39, pp. 1-30.
- [WaB13b] Wang, M., and Bertsekas, D. P., 2013. "Convergence of Iterative Simulation-Based Methods for Singular Linear Systems," *Stochastic Systems*, Vol. 3, pp. 39-96.
- [WaB16] Wang, M., and Bertsekas, D. P., 2016. "Incremental Constraint Projection-Proximal Methods for Nonsmooth Convex Optimization," *SIAM Journal on Optimization*, Vol. 26, pp. 681-717.
- [Wat89] Watkins, C. J. C. H., *Learning from Delayed Rewards*, Ph.D. Thesis, Cambridge Univ., England.
- [WeB99] Weaver, L., and Baxter, J., 1999. "Reinforcement Learning From State and Temporal Differences," *Tech. Report*, Department of Computer Science, Australian National University.
- [Wen14] Wen, Z., 2014. *Efficient Reinforcement Learning with Value Function Generalization*, Ph.D. Dissertation, Stanford University.

- [WhS89] White, C. C., and Scherer, W. T., 1989. "Solution Procedures for Partially Observed Markov Decision Processes," *Operations Res.*, Vol. 30, pp. 791-797.
- [Whi69] White, D. J., 1969. *Dynamic Programming*, Holden-Day, San Francisco, CA.
- [Whi82] Whittle, P., 1982. *Optimization Over Time*, Wiley, N. Y., Vol. 1, 1982, Vol. 2, 1983.
- [Whi88] Whittle, P., 1988. "Restless Bandits: Activity Allocation in a Changing World," *J. of Applied Probability*, pp. 287-298.
- [Whi90] Whittle, P., 1990. *Risk-Sensitive Optimal Control*, Wiley, N. Y.
- [Wie03] Wiewiora, E., 2003. "Potential-Based Shaping and Q-Value Initialization are Equivalent," *J. of Artificial Intelligence Research*, Vol. 19, pp. 205-208.
- [Wil71] Willems, J., 1971. "Least Squares Stationary Optimal Control and the Algebraic Riccati Equation," *IEEE Trans. on Automatic Control*, Vol. 16, pp. 621-634.
- [Wit66a] Witsenhausen, H. S., 1966. "Minimax Control of Uncertain Systems," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- [Wit66b] Witsenhausen, H. S., 1966. "A Comparison of Closed-Loop and Open-Loop Optimum Systems," *IEEE Trans. Automatic Control*, Vol. AC-11, pp. 619-621.
- [Wit70] Witsenhausen, H. S., 1970. "On Performance Bounds for Uncertain Systems," *SIAM J. on Control*, Vol. 8, pp. 55-89.
- [Wit71] Witsenhausen, H. S., 1971. "Separation of Estimation and Control for Discrete-Time Systems," *Proc. IEEE*, Vol. 59, pp. 1557-1566.
- [WuB99] Wu, C. C., and Bertsekas, D. P., 1999. "Distributed Power Control Algorithms for Wireless Networks," unpublished report, available from the author's [www](http://www.ccs.mit.edu) site.
- [YDR04] Yan, X., Diaconis, P., Rusmevichientong, P., and Van Roy, B., 2004. "Solitaire: Man Versus Machine," *Advances in Neural Information Processing Systems*, Vol. 17, pp. 1553-1560.
- [YuB04] Yu, H., and Bertsekas, D. P., 2004. "Discretized Approximations for POMDP with Average Cost," *Proc. of 20th Conference on Uncertainty in Artificial Intelligence*, Banff, Canada.
- [YuB07] Yu, H., and Bertsekas, D. P., 2007. "A Least Squares Q-Learning Algorithm for Optimal Stopping Problems," *Proc. European Control Conference 2007*, Kos, Greece, pp. 2368-2375; an extended version appears in *Lab. for Information and Decision Systems Report LIDS-P-2731*, MIT.
- [YuB09] Yu, H., and Bertsekas, D. P., 2009. "Convergence Results for Some Temporal Difference Methods Based on Least Squares," *IEEE Trans. on Aut. Control*, Vol. 54, pp. 1515-1531.
- [YuB10] Yu, H., and Bertsekas, D. P., 2010. "Error Bounds for Approximations from Projected Linear Equations," *Mathematics of Operations Research*, Vol. 35, pp. 306-329.
- [YuB12] Yu, H., and Bertsekas, D. P., 2012. "Weighted Bellman Equations and their Applications in Dynamic Programming," *Lab. for Information and Decision Systems Report LIDS-P-2876*, MIT.
- [YuB13a] Yu, H., and Bertsekas, D. P., 2013. "Q-Learning and Policy Iteration Algorithms for Stochastic Shortest Path Problems," *Annals of Operations Research*, Vol. 208, pp. 95-132.

[YuB13b] Yu, H., and Bertsekas, D. P., 2013. “On Boundedness of Q-Learning Iterates for Stochastic Shortest Path Problems,” *Math. of OR*, Vol. 38, pp. 209-227.

[Yu12] Yu, H., 2012. “Least Squares Temporal Difference Methods: An Analysis Under General Conditions,” *SIAM J. on Control and Optimization*, Vol. 50, pp. 3310-3343.