

- 1063.
- Szepesvári, Cs., 2010. *Algorithms for Reinforcement Learning*. Morgan & Claypool Publishers.
- Thiery, C., Scherrer, B., 21–24 June 2010. Least-squares λ policy iteration: Bias-variance trade-off in control problems. In: *Proceedings 27th International Conference on Machine Learning (ICML-10)*. Haifa, Israel, pp. 1071–1078.
- Thrun, S. B., 1992. Efficient exploration in reinforcement learning. Tech. rep., Pittsburgh, PA, USA.
- Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning 4* (2), 26–31.
- Tolić, D., Fierro, R., Ferrari, S., 2012. Optimal self-triggering for nonlinear systems via approximate dynamic programming. In: *IEEE Multi-Conference on Systems and Control*. pp. 879–884.
- Tolić, D., Palunko, I., 2017. Learning suboptimal broadcasting intervals in multi-agent systems. *IFAC-PapersOnLine* 50 (1), 4144 – 4149, 20th IFAC World Congress.
- Tsitsiklis, J. N., Van Roy, B., 1997. An analysis of temporal difference learning with function approximation. *IEEE Transactions on Automatic Control* 42 (5), 674–690.
- Tucker, G., Bhupatiraju, S., Gu, S., Turner, R. E., Ghahramani, Z., Levine, S., 2018. The mirage of action-dependent baselines in reinforcement learning. arXiv preprint arXiv:1802.10031.
- van Hasselt, H., 2010. Double q-learning. In: *Advances in Neural Information Processing Systems*. pp. 2613–2621.
- van Hasselt, H., Guez, A., Hessel, M., Mnih, V., Silver, D., 2016a. Learning values across many orders of magnitude. In: *Neural Information Processing Systems (NIPS)*.
- van Hasselt, H., Guez, A., Silver, D., 2016b. Deep reinforcement learning with double q-learning. In: *Conf. Artificial Intelligence (AAAI)*.
- Wang, Y., O’Donoghue, B., Boyd, S., 2014. Approximate dynamic programming via iterated bellman inequalities. *International Journal of Robust and Nonlinear Control* 25 (10), 1472–1496.
- Wang, Z., Bapst, V., Heess, N., Mnih, V., Munos, R., Kavukcuoglu, K., de Freitas, N., 2016. Sample efficient actor-critic with experience replay. arXiv preprint arXiv:1611.01224.
- Watkins, C. J. C. H., 1989. Learning from delayed rewards. Ph.D. thesis, King’s College, Cambridge.
- Watkins, C. J. C. H., Dayan, P., 1992. Q-learning. *Machine Learning* 8, 279–292.
- Watter, M., Springenberg, J., Boedecker, J., Riedmiller, M., 2015. Embed to control: A locally linear latent dynamics model for control from raw images. In: *Neural Information Processing Systems (NIPS)*.
- Weinstein, A., Littman, M. L., 25–19 June 2012. Bandit-based planning and learning in continuous-action Markov decision processes. In: *Proceedings 22nd International Conference on Automated Planning and Scheduling (ICAPS-12)*. São Paulo, Brazil.
- Wiering, M., van Otterlo, M. (Eds.), 2012. *Reinforcement Learning: State of the Art*. Vol. 12. Springer.
- Williams, R. J., 1992. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Machine Learning* 8, 229–256.
- Williams, R. J., Peng, J., 1991. Function optimization using connectionist reinforcement learning algorithms. *Connection Science* 3 (3), 241–268.
- Wiskott, L., Sejnowski, T. J., 2002. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation* 14 (4), 715–770.
- Yang, X., Liu, D., Wei, Q., 2014. Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming. *IET Control Theory Applications* 8 (16), 1676–1688.
- Yu, H., Bertsekas, D. P., 2009. Convergence results for some temporal difference methods based on least squares. *IEEE Transactions on Automatic Control* 54 (7), 1515–1531.
- Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O., 2016. Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.
- Zhang, H., Cui, L., Zhang, X., Luo, Y., Dec 2011. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks* 22 (12), 2226–2236.
- Zhang, J., Springenberg, J. T., Boedecker, J., Burgard, W., 2017. Deep reinforcement learning with successor features for navigation across similar environments. In: *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE, pp. 2371–2378.