

# Chapter 1

## Reinforcement Learning

### 1.1 The Reinforcement Learning Problem

### 1.2 Markov Decision Processes

### 1.3 Policy Gradient

#### 1.3.1 Finite Differences

#### 1.3.2 Monte-Carlo Policy Gradient

#### 1.3.3 Policy Gradient Theorem

#### 1.3.4 Actor-Critic Methods

#### 1.3.5 Natural Policy Gradient

#### 1.3.6 Policy Gradient with Parameter Exploration



# Bibliography

- [1] AGARWAL, A., BARTLETT, P., AND DAMA, M. Optimal allocation strategies for the dark pool problem. *arXiv preprint arXiv:1003.2245* (2010).
- [2] ALMGREN, R., AND CHRISS, N. Optimal execution of portfolio transactions. *Journal of Risk* 3 (2001), 5–40.
- [3] BÄUERLE, N., AND RIEDER, U. *Markov decision processes with applications to finance*. Springer Science & Business Media, 2011.
- [4] BEKIROU, S. D. Heterogeneous trading strategies with adaptive fuzzy actor-critic reinforcement learning: A behavioral approach. *Journal of Economic Dynamics and Control* 34, 6 (2010), 1153–1170.
- [5] BERTOLUZZO, F., AND CORAZZA, M. Testing different reinforcement learning configurations for financial trading: Introduction and applications. *Procedia Economics and Finance* 3 (2012), 68–77.
- [6] BERTOLUZZO, F., AND CORAZZA, M. Q-learning-based financial trading systems with applications. *University Ca’Foscari of Venice, Dept. of Economics Working Paper Series No 15* (2014).
- [7] BERTOLUZZO, F., AND CORAZZA, M. Reinforcement learning for automated financial trading: Basics and applications. In *Recent Advances of Neural Network Models and Applications*. Springer, 2014, pp. 197–213.
- [8] BERTSEKAS, D. P. *Dynamic programming and optimal control*, vol. 1. Athena Scientific, Belmont, 1995.
- [9] BERTSEKAS, D. P., AND TSITSIKLIS, J. N. *Neuro-Dynamic Programming*, 1 ed. Optimization and neural computation series. Athena Scientific, Belmont, 1996.
- [10] BISHOP, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.

- 
- [11] BUSONI, L., BABUSKA, R., DE SCHUTTER, B., AND ERNST, D. *Reinforcement learning and dynamic programming using function approximators*, vol. 39. CRC press, 2010.
  - [12] CASQUEIRO, P. X., AND RODRIGUES, A. J. Neuro-dynamic trading methods. *European Journal of Operational Research* 175, 3 (2006), 1400–1412.
  - [13] CHAPADOS, N., AND BENGIO, Y. Cost functions and model combination for var-based asset allocation using neural networks. *IEEE Transactions on Neural Networks* 12 (2001), 890–906.
  - [14] CHOEY, M., AND WEIGEND, A. S. Nonlinear trading models through sharpe ratio maximization. *International Journal of Neural Systems* 8 (1997), 417–431.
  - [15] CHOW, Y., TAMAR, A., MANNOR, S., AND PAVONE, M. Risk-sensitive and robust decision-making: a cvar optimization approach. In *Advances in Neural Information Processing Systems* (2015), pp. 1522–1530.
  - [16] CORAZZA, M., AND SANGALLI, A. Q-learning vs. sarsa: comparing two intelligent stochastic control approaches for financial trading.
  - [17] CUMMING, J., ALRAJEH, D., AND DICKENS, L. An investigation into the use of reinforcement learning techniques within the algorithmic trading domain.
  - [18] DEMPSTER, M. A., AND LEEMANS, V. An automated fx trading system using adaptive reinforcement learning. *Expert Systems with Applications* 30, 3 (2006), 543–552.
  - [19] DEMPSTER, M. A. H., AND ROMAHI, Y. S. *Intraday FX trading: An evolutionary reinforcement learning approach*. Springer, 2002.
  - [20] DENG, Y., BAO, F., KONG, Y., REN, Z., AND DAI, Q. Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems* (2016).
  - [21] DENG, Y., KONG, Y., BAO, F., AND DAI, Q. Sparse coding inspired optimal trading system for hft industry. *IEEE Transactions on Industrial Informatics* 11, 2 (2015), 467–475.
  - [22] DU, X., ZHAI, J., AND LV, K. Algorithm trading using q-learning and recurrent reinforcement learning. *positions* 1, 1.

- [23] ELDER, T. Creating algorithmic traders with hierarchical reinforcement learning.
- [24] FELDKAMP, L. A., PROKHOROV, D. V., EAGEN, C. F., AND YUAN, F. Enhanced multi-stream kalman filter training for recurrent networks. In *Nonlinear Modeling*. Springer, 1998, pp. 29–53.
- [25] GANCHEV, K., NEVMYVAKA, Y., KEARNS, M., AND VAUGHAN, J. W. Censored exploration and the dark pool problem. *Communications of the ACM* 53, 5 (2010), 99–107.
- [26] GOLD, C. Fx trading via recurrent reinforcement learning. In *IEEE 2003 IEEE International Conference on Computational Intelligence for Financial Engineering. Proceedings* (2003).
- [27] GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. Deep learning. Book in preparation for MIT Press, 2016.
- [28] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
- [29] HENDRICKS, D., AND WILCOX, D. A reinforcement learning extension to the almgren-chriss framework for optimal trade execution. In *IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (2014), pp. 457–464.
- [30] JAEGER, H. *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach*. GMD-Forschungszentrum Informationstechnik, 2002.
- [31] KAMIJO, K., AND TANIGAWA, T. Stock price pattern recognition-a recurrent neural network approach. In *IEEE 1990 IJCNN International Joint Conference on Neural Networks* (1990).
- [32] KEARNS, M., AND NEVMYVAKA, Y. Machine learning for market microstructure and high frequency trading. *High-Frequency Trading—New Realities for Traders, Markets and Regulators* (2013), 91–124.
- [33] KONDA, V. R., AND TSITSIKLIS, J. N. Actor-critic algorithms. In *NIPS* (1999), vol. 13, pp. 1008–1014.
- [34] LARUELLE, S., LEHALLE, C.-A., AND PAGES, G. Optimal split of orders across liquidity pools: a stochastic algorithm approach. *SIAM Journal on Financial Mathematics* 2, 1 (2011), 1042–1076.

- [35] LARUELLE, S., LEHALLE, C.-A., AND PAGES, G. Optimal posting price of limit orders: learning by trading. *Mathematics and Financial Economics* 7, 3 (2013), 359–403.
- [36] LI, H., DAGLI, C. H., AND ENKE, D. Short-term stock market timing prediction under reinforcement learning schemes. In *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning* (2007), pp. 233–240.
- [37] MOODY, J., AND SAFFELL, M. Learning to trade via direct reinforcement. *Neural Networks, IEEE Transactions on* 12, 4 (2001), 875–889.
- [38] MOODY, J., SAFFELL, M., LIAO, Y., AND WU, L. Reinforcement learning for trading systems. In *Decision Technologies for Computational Finance: Proceedings of the fifth International Conference Computational Finance* (2013), vol. 2, Springer Science & Business Media, p. 129.
- [39] MOODY, J., AND WU, L. Optimization of trading systems and portfolios. In *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)* (1997), pp. 300–307.
- [40] MOODY, J., WU, L., LIAO, Y., AND SAFFELL, M. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting* 17 (1998), 441–470.
- [41] NEVMYVAKA, Y., FENG, Y., AND KEARNS, M. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd international conference on Machine learning* (2006), ACM, pp. 673–680.
- [42] NOCEDAL, J., AND WRIGHT, S. *Numerical optimization*. Springer Science & Business Media, 2006.
- [43] O, J., LEE, J., LEE, J. W., AND ZHANG, B.-T. Adaptive stock trading with dynamic asset allocation using reinforcement learning. *Information Sciences* 176, 15 (2006), 2121–2147.
- [44] PETERS, J., AND SCHAAL, S. Policy gradient methods for robotics. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (2006), IEEE, pp. 2219–2225.
- [45] PETERS, J., AND SCHAAL, S. Reinforcement learning of motor skills with policy gradients. *Neural networks* 21, 4 (2008), 682–697.

- [46] SAAD, E. W., PROKHOROV, D. V., AND WUNSCH, D. C. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks* 9, 6 (1998), 1456–1470.
- [47] SEHNKE, F., ET AL. *Parameter exploring policy gradients and their implications*. PhD thesis, Technische Universität München, 2012.
- [48] SEHNKE, F., OSENDORFER, C., RÜCKSTIESS, T., GRAVES, A., PETERS, J., AND SCHMIDHUBER, J. Policy gradients with parameter-based exploration for control. In *Artificial Neural Networks-ICANN 2008*. Springer, 2008, pp. 387–396.
- [49] SEHNKE, F., OSENDORFER, C., RÜCKSTIESS, T., GRAVES, A., PETERS, J., AND SCHMIDHUBER, J. Parameter-exploring policy gradients. *Neural Networks* 23, 4 (2010), 551–559.
- [50] SILVER, D., LEVER, G., HEES, N., DEGRIS, T., WIERSTRA, D., AND RIEDMILLER, M. Deterministic policy gradient algorithms. In *ICML* (2014).
- [51] SUTTON, R. S., AND BARTO, A. G. *Introduction to reinforcement learning*, vol. 135. MIT Press Cambridge, 1998.
- [52] SUTTON, R. S., MCALLESTER, D. A., SINGH, S. P., MANSOUR, Y., ET AL. Policy gradient methods for reinforcement learning with function approximation. In *NIPS* (1999), vol. 99, pp. 1057–1063.
- [53] SZEPESVÁRI, C. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning* 4, 1 (2010), 1–103.
- [54] TAMAR, A., CASTRO, D. D., AND MANNOR, S. Temporal difference methods for the variance of the reward to go. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* (2013), pp. 495–503.
- [55] TAMAR, A., CHOW, Y., GHAVAMZADEH, M., AND MANNOR, S. Policy gradient for coherent risk measures. In *Advances in Neural Information Processing Systems* (2015), pp. 1468–1476.
- [56] TAMAR, A., DI CASTRO, D., AND MANNOR, S. Policy gradients with variance related risk criteria. In *Proceedings of the 29th International Conference on Machine Learning* (2012), pp. 387–396.

- 
- [57] TAN, Z., QUEK, C., AND CHENG, P. Y. Stock trading with cycles: A financial application of anfis and reinforcement learning. *Expert Systems with Applications* 38, 5 (2011), 4741–4755.
  - [58] TSAY, R. S. *Analysis of financial time series*, vol. 543. John Wiley & Sons, 2005.
  - [59] WERBOS, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* 78, 10 (1990), 1550–1560.
  - [60] WIERING, M., AND VAN OTTERLO, M. *Reinforcement Learning: State-of-the-Art*, 1 ed., vol. 12 of *Adaptation, Learning, and Optimization*. Springer, 2012.
  - [61] WILLIAMS, R. J., AND ZIPSER, D. A learning algorithm for continually running fully recurrent neural networks. *Neural computation* 1, 2 (1989), 270–280.
  - [62] YANG, S., PADDRIK, M., HAYES, R., TODD, A., KIRILENKO, A., BELING, P., AND SCHERER, W. Behavior based learning in identifying high frequency trading strategies. In *IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (2012).