

















## REFERENCES

- [1] Ilan Adler. 2013. The equivalence of linear programs and zero-sum games. *International Journal of Game Theory* 42, 1 (2013), 165–177.
- [2] Bikramjit Banerjee and Jing Peng. 2004. Performance bounded reinforcement learning in strategic interactions. In *AAAI*, Vol. 4. 2–7.
- [3] Ilai Bistritz and Amir Leshem. 2018. Distributed multi-player bandits—a game of thrones approach. In *Advances in Neural Information Processing Systems*. 7222–7232.
- [4] Walter Bossert and Guofu Tan. 1995. An arbitration game and the egalitarian bargaining solution. *Social Choice and Welfare* 12, 1 (1995), 29–41.
- [5] Michael Bowling. 2005. Convergence and no-regret in multiagent learning. In *Advances in neural information processing systems*. 209–216.
- [6] Michael Bowling and Manuela Veloso. 2001. Convergence of Gradient Dynamics with a Variable Learning Rate. In *In Proceedings of the Eighteenth International Conference on Machine Learning*. 27–34.
- [7] Ronen I Brafman and Moshe Tennenholtz. 2002. R-max—a general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research* 3, Oct (2002), 213–231.
- [8] Ronen I Brafman and Moshe Tennenholtz. 2003. Efficient learning equilibrium. In *Advances in Neural Information Processing Systems*. 1635–1642.
- [9] Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. 2019. Delay and cooperation in nonstochastic bandits. *The Journal of Machine Learning Research* 20, 1 (2019), 613–650.
- [10] Nicolò Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge university press.
- [11] Vincent Conitzer and Tuomas Sandholm. 2007. AWESOME: A general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Machine Learning* 67, 1-2 (2007), 23–43.
- [12] Jacob W Crandall and Michael A Goodrich. 2011. Learning to compete, coordinate, and cooperate in repeated games using reinforcement learning. *Machine Learning* 82, 3 (2011), 281–314.
- [13] George B Dantzig. 1951. A proof of the equivalence of the programming problem and the game problem. *Activity analysis of production and allocation* 13 (1951), 330–338.
- [14] Madalina M Drugan and Ann Nowe. 2013. Designing multi-objective multi-armed bandits algorithms: A study. *learning* 8 (2013), 9.
- [15] Amy Greenwald and Keith Hall. 2003. Correlated-Q Learning. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning (ICML'03)*. AAAI Press, 242–249.
- [16] Eshcar Hillel, Zohar S Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. 2013. Distributed exploration in multi-armed bandits. In *Advances in Neural Information Processing Systems*. 854–862.
- [17] Junling Hu and Michael P Wellman. 2003. Nash Q-learning for general-sum stochastic games. *Journal of machine learning research* 4, Nov (2003), 1039–1069.
- [18] Junling Hu, Michael P Wellman, et al. 1998. Multiagent reinforcement learning: theoretical framework and an algorithm. In *ICML*, Vol. 98. Citeseer, 242–250.
- [19] Haruo Imai. 1983. Individual monotonicity and lexicographic maxmin solution. *Econometrica: Journal of the Econometric Society* (1983), 389–401.
- [20] Thomas Jaksch, Ronald Ortner, and Peter Auer. 2010. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research* 11, Apr (2010), 1563–1600.
- [21] Ehud Kalai. 1977. Proportional solutions to bargaining situations: interpersonal utility comparisons. *Econometrica: Journal of the Econometric Society* (1977), 1623–1630.
- [22] Ehud Kalai, Meir Smorodinsky, et al. 1975. Other solutions to Nash’s bargaining problem. *Econometrica* 43, 3 (1975), 513–518.
- [23] Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. 2016. Distributed cooperative decision-making in multiarmed bandits: Frequentist and Bayesian algorithms. In *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 167–172.
- [24] Michael L Littman. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*. Elsevier, 157–163.
- [25] Michael L Littman. 2001. Friend-or-foe Q-learning in general-sum games. In *ICML*, Vol. 1. 322–328.
- [26] Michael L. Littman and Peter Stone. 2003. A Polynomial-time Nash Equilibrium Algorithm for Repeated Games. In *Proceedings of the 4th ACM Conference on Electronic Commerce (EC '03)*. ACM, New York, NY, USA, 48–54. <https://doi.org/10.1145/779928.779935>
- [27] Michael L Littman and Csaba Szepesvári. 1996. A generalized reinforcement-learning model: Convergence and applications. In *ICML*, Vol. 96. 310–318.
- [28] Enrique Munoz de Cote and Michael L. Littman. 2008. A Polynomial-time Nash Equilibrium Algorithm for Repeated Stochastic Games. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press, Corvallis, Oregon, 419–426.
- [29] John F Nash Jr. 1950. The bargaining problem. *Econometrica: Journal of the Econometric Society* (1950), 155–162.
- [30] Martin J Osborne and Ariel Rubinstein. 1994. *A course in game theory*.
- [31] Rob Powers, Yoav Shoham, and Thuc Vu. 2007. A general criterion and an algorithmic framework for learning in multi-agent systems. *Machine Learning* 67, 1 (01 May 2007), 45–76.
- [32] Haiyan Qiao, Jerzy Rozenblit, Ferenc Szidarovszky, and Lizhi Yang. 2006. Multi-agent learning model with bargaining. In *Proceedings of the 2006 winter simulation conference*. IEEE, 934–940.
- [33] John Rawls. 1971. *A theory of justice*. Harvard university press.
- [34] Yuzuru Sato, Eizo Akiyama, and J Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences* 99, 7 (2002), 4748–4751.
- [35] Shahin Shahrampour, Alexander Rakhlin, and Ali Jadbabaie. 2017. Multi-armed bandits in multi-agent networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2786–2790.
- [36] Satinder Singh, Michael Kearns, and Yishay Mansour. 2000. Nash Convergence of Gradient Dynamics in General-Sum Games. In *In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. Morgan, 541–548.
- [37] Paul N Somerville. 1954. Some problems of optimum sampling. *Biometrika* 41, 3/4 (1954), 420–429.
- [38] Jeff L Stimpson and Michael A Goodrich. 2003. Learning to cooperate in a social dilemma: A satisficing approach to bargaining. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 728–735.
- [39] William Thomson. 1981. Nash’s bargaining solution and utilitarian choice rules. *Econometrica: Journal of the Econometric Society* (1981), 535–538.
- [40] Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. 2017. Online Reinforcement Learning in Stochastic Games. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). 4987–4997.
- [41] Martin Zinkevich, Amy Greenwald, and Michael L. Littman. 2006. Cyclic Equilibria in Markov Games. In *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. C. Platt (Eds.). MIT Press, 1641–1648.