REFERENCES

[1] Jacob Abernethy, Kevin A Lai, and Andre Wibisono. 2019. Last-iterate convergence rates for min-max optimization. *arXiv preprint arXiv:1906.02027* (2019).

[2] James P Bailey, Gauthier Gidel, and Georgios Piliouras. 2019. Finite Regret and Cycles with Fixed Step-Size via Alternating Gradient Descent-Ascent. *arXiv preprint arXiv:1907.04392* (2019).

[3] James P Bailey and Georgios Piliouras. 2018. Multiplicative Weights Update in Zero-Sum Games. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM, 321–338.

[4] Marc G Bellemare, Georg Ostrovski, Arthur Guez, Philip S Thomas, and Rémi Munos. 2016. Increasing the Action Gap: New Operators for Reinforcement Learning.. In *AAAI*. 1476–1483.

[5] Dimitri Bertsekas. 1982. Distributed dynamic programming. *IEEE transactions on Automatic Control* 27, 3 (1982), 610–616.

[6] Dimitri P. Bertsekas and John Tsitsiklis. 1996. *Neuro-Dynamic Programming*. Athena Scientifit.

[7] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. 2015. Heads-up limit hold'em poker is solved. *Science* 347, 6218 (2015), 145–149.

[8] Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. 2018. Deep Counterfactual Regret Minimization. *arXiv preprint arXiv:1811.00164* (2018).

[9] Noam Brown and Tuomas Sandholm. 2019. Solving imperfect-information games via discounted regret minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1829–1836.

[10] Noam Brown and Tuomas Sandholm. 2019. Superhuman AI for multiplayer poker. *Science* 365, 6456 (2019), 885–890.

[11] Yun Kuen Cheung and Georgios Piliouras. 2019. Vortices Instead of Equilibria in MinMax Optimization: Chaos and Butterfly Effects of Online Learning in Zero-Sum Games. *arXiv preprint arXiv:1905.08396* (2019).

[12] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2017. Training GANs with Optimism. *arXiv preprint arXiv:1711.00141* (2017).

[13] Constantinos Daskalakis and Ioannis Panageas. 2018. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. *arXiv preprint arXiv:1807.04252* (2018).

[14] Nasrollah Etemadi. 2006. Convergence of weighted averages of random variables revisited. *Proc. Amer. Math. Soc.* 134, 9 (2006), 2739–2744.

[15] Eyal Even-Dar, Sham M Kakade, and Yishay Mansour. 2005. Experts in a Markov decision process. In *Advances in neural information processing systems*. 401–408.

[16] Eyal Even-Dar, Sham M Kakade, and Yishay Mansour. 2009. Online Markov decision processes. *Mathematics of Operations Research* 34, 3 (2009), 726–736.

[17] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. 2002. PAC bounds for multi-armed bandit and Markov decision processes. In *International Conference on Computational Learning Theory*. Springer, 255–270.

[18] Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. 2019. Stable-Predictive Optimistic Counterfactual Regret Minimization. *arXiv preprint arXiv:1902.04982* (2019).

[19] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2018. Composability of Regret Minimizers. *arXiv preprint arXiv:1811.02540* (2018).

[20] Richard G Gibson, Marc Lanctot, Neil Burch, Duane Szafron, and Michael Bowling. 2012. Generalized Sampling and Variance in Counterfactual Regret Minimization.. In *AAAI*.

[21] Eyal Gofer and Yishay Mansour. 2016. Lower bounds on individual sequence regret. *Machine Learning* 103, 1 (2016), 1–26.

[22] David Gondek, Amy Greenwald, and Keith Hall. 2004. QNR-Learning in Markov Games. (2004).

[23] Amy Greenwald, Keith Hall, and Roberto Serrano. 2003. Correlated Q-learning. In *ICML*, Vol. 3. 242–249.

[24] Amy Greenwald and Amir Jafari. 2003. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Learning Theory and Kernel Machines*. Springer, 2–12.

[25] Sergiu Hart and Andreu Mas-Colell. 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica* 68, 5 (2000), 1127–1150.

[26] Johannes Heinrich and David Silver. 2016. Deep reinforcement learning from self-play in imperfect-information games. *arXiv preprint arXiv:1603.01121* (2016).

[27] 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.

[28] Chi Jin, Zeyuan Allen-Zhu, Sebastien Bubeck, and Michael I Jordan. 2018. Is q-learning provably efficient?. In *Advances in Neural Information Processing Systems*. 4863–4873.

[29] Peter H Jin, Sergey Levine, and Kurt Keutzer. 2017. Regret Minimization for Partially Observable Deep Reinforcement Learning. *arXiv preprint arXiv:1710.11424* (2017).

[30] Michael Johanson, Nolan Bard, Marc Lanctot, Richard Gibson, and Michael Bowling. 2012. Efficient Nash equilibrium approximation through Monte Carlo counterfactual regret minimization. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 837–846.

[31] Ian A Kash, Michael Sullins, and Katja Hofmann. 2019. Combining no-regret and Q-learning. *arXiv preprint arXiv:1910.03094* (2019).

[32] Michael J Kearns and Satinder P Singh. 1999. Finite-sample convergence rates for Q-learning and indirect algorithms. In *Advances in neural information processing systems*. 996–1002.

[33] Vojtěch Kovařík and Viliam Lisỳ. 2018. Analysis of hannan consistent selection for monte carlo tree search in simultaneous move games. *arXiv preprint arXiv:1804.09045* (2018).

[34] Marc Lanctot, Richard Gibson, Neil Burch, Martin Zinkevich, and Michael Bowling. 2012. No-regret learning in extensive-form games with imperfect recall. *arXiv preprint arXiv:1205.0622* (2012).

[35] Marc Lanctot, Kevin Waugh, Martin Zinkevich, and Michael Bowling. 2009. Monte Carlo sampling for regret minimization in extensive games. In *Advances in neural information processing systems*. 1078–1086.

[36] Hui Li, Kailiang Hu, Zhibang Ge, Tao Jiang, Yuan Qi, and Le Song. 2018. Double Neural Counterfactual Regret Minimization. *arXiv preprint arXiv:1812.10607* (2018).

[37] Michael L Littman. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*. Elsevier, 157–163.

[38] Yao Ma, Hao Zhang, and Masashi Sugiyama. 2015. Online Markov decision processes with policy iteration. *arXiv preprint arXiv:1510.04454* (2015).

[39] Shie Mannor and Nahum Shimkin. 2003. The empirical Bayes envelope and regret minimization in competitive Markov decision processes. *Mathematics of Operations Research* 28, 2 (2003), 327–345.

[40] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2703–2717.

[41] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356, 6337 (2017), 508–513.

[42] Gergely Neu, Anders Jonsson, and Vicenç Gómez. 2017. A unified view of entropy-regularized markov decision processes. *arXiv preprint arXiv:1705.07798* (2017).

[43] Goran Radanovic, Rati Devidze, David Parkes, and Adish Singla. 2019. Learning to collaborate in markov decision processes. *arXiv preprint arXiv:1901.08029* (2019).

[44] Sriram Srinivasan, Marc Lanctot, Vinicius Zambaldi, Julien Pérolat, Karl Tuyls, Rémi Munos, and Michael Bowling. 2018. Actor-critic policy optimization in partially observable multiagent environments. In *Advances in Neural Information Processing Systems*. 3422–3435.

[45] Oskari Tammelin. 2014. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042* (2014).

[46] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. 2015. Solving Heads-Up Limit Texas Hold'em.. In *IJCAI*. 645–652.

[47] Gerald Tesauro and Jeffrey O Kephart. 2002. Pricing in agent economies using multi-agent Q-learning. *Autonomous Agents and Multi-Agent Systems* 5, 3 (2002), 289–304.

[48] Harm Van Seijen, Hado Van Hasselt, Shimon Whiteson, and Marco Wiering. 2009. A theoretical and empirical analysis of Expected Sarsa. In *Adaptive Dynamic Programming and Reinforcement Learning, 2009. ADPRL'09. IEEE Symposium on*. IEEE, 177–184.

[49] Kevin Waugh, Dustin Morrill, James Andrew Bagnell, and Michael Bowling. 2015. Solving Games with Functional Regret Estimation.. In *AAAI*, Vol. 15. 2138–2144.

[50] Yasin Abbasi Yadkori, Peter L Bartlett, Varun Kanade, Yevgeny Seldin, and Csaba Szepesvári. 2013. Online learning in Markov decision processes with adversarially chosen transition probability distributions. In *Advances in neural information processing systems*. 2508–2516.

[51] Jia Yuan Yu and Shie Mannor. 2009. Online learning in Markov decision processes with arbitrarily changing rewards and transitions. In *2009 International Conference on Game Theory for Networks*. IEEE, 314–322.

[52] Jia Yuan Yu, Shie Mannor, and Nahum Shimkin. 2009. Markov decision processes with arbitrary reward processes. *Mathematics of Operations Research* 34, 3 (2009), 737–757.

[53] Martin Zinkevich, Amy Greenwald, and Michael L Littman. 2006. Cyclic equilibria in Markov games. In *Advances in Neural Information Processing Systems*. 1641–1648.

[54] Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. 2008. Regret minimization in games with incomplete information. In *Advances in neural information processing systems*. 1729–1736.