A Tighter Analysis of Randomised Policy Iteration

Meet Taraviya and **Shivaram Kalyanakrishnan**
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
{mtaraviya, shivaram}@cse.iitb.ac.in

Abstract

Policy Iteration (PI) is a popular family of algorithms to compute an optimal policy for a given Markov Decision Problem (MDP). Starting from an arbitrary initial policy, PI repeatedly performs locally-improving switches until an optimal policy is found. The exact form of the switching rule gives rise to different variants of PI. Two decades ago, Mansour and Singh [1999] provided the first non-trivial "strong" upper bound on the number of iterations taken by "Howard's PI" (HPI), a widely-used variant of PI (strong bounds depend only on the number of states and actions in the MDP). They also proposed a randomised variant (RPI) and showed an even tighter strong upper bound. Their bounds for HPI and RPI have not been improved subsequently.

We revisit the algorithms and analysis of Mansour and Singh [1999]. We prove a novel result on the structure of the policy space for k-action MDPs, $k \geq 2$, which generalises a result known for $k = 2$. Also proposing a new counting argument, we obtain a strong bound of $(O(\sqrt{k \log k}))^n$ iterations for an algorithm akin to RPI, improving significantly upon Mansour and Singh's original bound of roughly $O((k/2)^n)$. Similar analysis of a randomised variant of HPI also yields a strong upper bound of $(O(\sqrt{k \log k}))^n$ iterations, registering the first exponential improvement for HPI over the trivial bound of k^n. Our other contributions include a lower bound of $\Omega(n)$ iterations for RPI and an upper bound of 1.6001^n iterations for a randomised variant of "Batch-Switching PI" [Kalyanakrishnan et al., 2016] on 2-action MDPs—the tightest strong upper bound shown yet for the PI family.

References

[Kalyanakrishnan et al., 2016] Shivaram Kalyanakrishnan, Utkarsh Mall, and Ritish Goyal. Batch-switching policy iteration. In *Proceedings of the Twenty-fifth International Joint Conference on Artificial Intelligence (IJCAI 2016)*, pages 3147–3153. AAAI Press, 2016.

[Mansour and Singh, 1999] Yishay Mansour and Satinder Singh. On the complexity of policy iteration. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI 1999)*, pages 401–408. Morgan Kaufmann, 1999.