The geometry of discounted stationary distributions of Markov decision processes

Johannes Müller

Max-Planck-Institut für Mathematik in den Naturwissenschaften,
Mathematical Machine Learning, Leipzig, Germany

The problem of reward maximisation in Markov decision processes is equivalent to the maximisation of a linear function over the set of all discounted stationary distributions. Therefore, we study the geometry of this set and show that in the case of full observability, the discounted stationary distributions form a polytope, which is combinatorially equivalent to the policy polytope under mild assumptions. We see that the orientation of this polytope is solely determined by the discount factor and the transition mechanism of the process, whereas the affine part additionally depends on the initial distribution. Further, we compute the rational degree of the mapping from policies to discounted stationary distributions and see that it essentially agrees with the degree of observability of the system. Finally, we discuss how linear constraints of the policy model correspond to polynomial inequalities in the polytope of discounted stationary distributions.