Class 11

Shikhar Saxena

February 14, 2023

Contents

Policy Iteration Algorithm

1. Set $n = 0$ and pick an arbitrary stationary policy π^0.
2. (Policy Evaluation) Obtain V^{π_n} by repeatedly applying T_{π_n} on arbitrary u. Or alternatively,

$$V^{\pi_n} = [I - \alpha P^{\pi_n}]^{-1} r^{\pi_n}$$

3. (Policy Improvement) Find policy π^{n+1} that improves upon policy π^n by being greedy on V^{π_n}.
4. If $\pi^{n+1} = \pi^n$ then stop else update the policy and n and repeat step two onwards.

Corollary 1. *If f is not α-optimal, I can always improve it or if you cannot improve a policy for atleast one state, then it is already α-optimal.*

Contraction Mappings

For any function $u \in B(\mathcal{S})$ let $\|u\| = \sup_{i\geq 0} |u(i)|$

Note. sup *used when the set is unbounded for example $u(i) = 1 - \frac{1}{i}$. Here* max *doesn't make sense as we never reach 1 (only closer to it).*

Definition 1. *A mapping $T : B(\mathcal{S}) \to B(\mathcal{S})$ is said to be a* ***contraction mapping*** *if*

$$\|Tu - Tv\| \leq \beta \|u - v\|$$

for some $\beta < 1$, $\forall u, v \in B(\mathcal{S})$.

Called contraction as the vectors we get after appling the operator are closer than they were before. (norm decreases).

Theorem 1. *If $T : B(\mathcal{S}) \to B(\mathcal{S})$ is a contraction mapping then there exists a unique function $g \in B(\mathcal{S})$ such that $Tg = g$. Furthermore $\forall u \in B(\mathcal{S})$, $T^n u \to g$ as $n \to \infty$.*

Proof. Check slides. □

This theorem is a special case of Banach fixed-point theorem. Other fixed-point theorems are Brouwer and Kakutani.

Optimality Operator T_α

Theorem 2. *The optimality operator T_α*

$$(T_\alpha u)(i) = \max_a \left\{ r(i,a) + \alpha \sum_j P_{ij}(a) u(j) \right\}$$

is a contraction mapping.

Proof. Check slides. □