Convergence and Optimality of Value Iteration

Value Iteration: Existence, Uniqueness and Convergence

Theorem 2: Contraction Mapping

Let

$$T(V)(i) = \max_u \left(E[r(i,u)] + \alpha \sum_j P_{ij}(u) V(j) \right)$$

Then the value iteration algorithm can be written as

$$V_{k+1} = T(V_k)$$

and the Bellman Equation can be written as

$$V^* = T(V^*).$$

T is a contraction mapping given $\alpha \in [0, 1)$.

Value Iteration: Existence, Uniqueness and Convergence

Proof of Theorem 2

Two facts that are easy to prove

- Fact 1: If $V_1 \leq V_2$ (entry-wise), then

$$\tilde{V}_1 = T(V_1) \leq T(V_2) = \tilde{V}_2$$

$$\begin{pmatrix} V_1(1) \\ V_1(2) \end{pmatrix} \leq \begin{pmatrix} V_2(1) \\ V_2(2) \end{pmatrix}$$

$$V_1(1) \leq V_2(1)$$
$$V_1(2) \leq V_2(2)$$

- Fact 2: Let $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$, then

$$T(V + \gamma e) = T(V) + \alpha\gamma e \quad \forall \gamma \text{ (scalar)}.$$

Value Iteration: Existence, Uniqueness and Convergence

Proof of Theorem 2

Define $b = \max_i |V_1(i) - V_2(i)| = \|V_1 - V_2\|_\infty$, then

$$V_2 - be \leq V_1 \leq V_2 + be$$

$$T(V_2 - be) \leq T(V_1) \leq T(V_2 + be) \quad \text{(Fact 1)}$$

$$T(V_2) - \alpha be \leq T(V_1) \leq T(V_2) + \alpha be \quad \text{(Fact 2)}$$

$$\|T(V_1) - T(V_2)\|_\infty \leq \alpha b = \alpha \|V_1 - V_2\|_\infty$$

(Contraction mapping)

Thus, $V^* = T(V^*)$ has a unique solution.

Value Iteration: Existence, Uniqueness and Convergence

Theorem 3: Obtain an Optimal Policy

$$\mu^*(i) \in \arg\max_u \left(E[r(i,u)] + \alpha \sum_j P_{ij}(u) V^*(j) \right)$$

where $V^* = T(V^*)$. Then μ^* is an optimal policy.

Value Iteration: Existence, Uniqueness and Convergence

Proof of Theorem 3

Considering stationary policy μ^*,

$$T_{\mu^*}(V^*)(i) = E[r(i, \mu^*(i))] + \alpha \sum_j P_{ij}(\mu^*(i)) V^*(j)$$

$$= \max_u E[r(i,u)] + \alpha \sum_j P_{ij}(u) V^*(j) = V^*(i). \quad \text{(Theorem 1)}$$

$$\implies T_{\mu^*}(V^*) = V^*$$

Therefore, V^* is a fixed point of T_{μ^*}. Since T_{μ^*} has a unique fixed point,

$$V_{\mu^*} = V^*.$$

Reference

- This lecture is based on R. Srikant's lecture notes on *MDPs with discounted cost* available at `https://sites.google.com/illinois.edu/mdps-and-rl/lectures?authuser=1`