Tosic Stefan

Mtk.: 11817912

Algorithm 1: Iterative Policy Evaluation, for estimating $V \approx v_\pi$

Input: π, the policy to be evaluated
Data: a small threshold $\theta > 0$ determining accuracy of estimation
Output: $V \approx v_\pi$
initialize $V(s)$ arbitrarily for all $s \in \mathcal{S}$, and $V(\text{terminal})$ to 0;
repeat
  $\Delta \leftarrow 0$;
  foreach $s \in \mathcal{S}$ **do**
    $v_{old} \leftarrow V(s)$;
    $V(s) \leftarrow r(s) + \gamma \sum_{s'} p(s'|s)\, V(s')$;
    $\Delta \leftarrow \max(\Delta, |v_{old} - V(s)|)$;
  end
until $\Delta < \theta$;

Your task: Prove that for $\gamma < 1$, Iterative Policy Evaluation (Algorithm 1) always converges to v_π, for any MDP, any policy π and any initialization of $V(s)$. The key step is to interpret the value function as a $|\mathcal{S}|$-dimensional Euclidean vector and show that the Bellman equation is a contraction for any $\gamma < 1$. When this has been shown, the rest of the proof (which should be provided) will follow via Banach's fixed point theorem.

Proof that $\gamma < 1$ always converges to v_π!

↳ allows us to compute the state-value functions $v_\pi(s)$ for any arbitrary policy π.

Theorem 1: Let V be a Banach space and $T: V \to V$ be a contraction mapping. Then T has a unique fixed point v^*. Furthermore, for any $v_0 \in V$, let $(v_n; n \geq 0)$ be a sequence of vectors defined via $v_{n+1} = T(v_n)$. For any v_0, this sequence converges to v^*.

show that T is a contraction. A mapping $T: V \to V$ is called L-Lipschitz if for any $v_1, v_2 \in V$

$$\| T(v_1) - T(v_2) \| \leq L \, \| v_1 - v_2 \|$$

T is called a contraction if it is L-Lipschitz with $L < 1$

Considering the metric space (V, d), where V is the vector space over the value function vectors and d is a metric induced by an L_∞-norm:

$$\forall v \in V: \| v \|_\infty = \max_{s \in S} |v(s)|$$

S ... finite set of states

$$\forall v_1, v_2 \in V: d(T(v_1), T(v_2)) = \| v_1 - v_2 \|_\infty = \max_{s \in S} |v_1(s) - v_2(s)|$$

The operator T is a γ-contraction which means that:

$$\forall v_1, v_2 \in V: d(T(v_1, T(v_2)) \leq \gamma \, d(v_1, v_2)$$

$$\left(V(s) = r(s) + \gamma \sum_{s'} p(s'|s)\, V(s') \right)$$

Proof :

$$\|T(v_1) - T(v_2)\|_\infty = \|(r(s) + \gamma P(s)\bar{v}_1) - (r(s) + \gamma P(s)\bar{v}_2)\|_\infty$$

$$= \|\gamma \cdot P(s)(\bar{v}_1 - \bar{v}_2)\|_\infty$$

$$= \left\|\gamma \cdot P(s) \begin{pmatrix} \bar{v}_1(s_1) - \bar{v}_2(s_1) \\ \bar{v}_1(s_2) - \bar{v}_2(s_1) \\ \vdots \\ \bar{v}_1(s_{|S|}) - \bar{v}_2(s_{|S|}) \end{pmatrix}\right\|_\infty$$

$$\leq \left\|\gamma \cdot P(s) \begin{pmatrix} \|\bar{v}_1 - \bar{v}_2\|_\infty \\ \|\bar{v}_1 - v_2\|_\infty \\ \vdots \\ \|\bar{v}_1 - \bar{v}_2\|_\infty \end{pmatrix}\right\|_\infty$$

$$= \left\|\gamma P(s) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} (\|\bar{v}_1 - \bar{v}_2\|_\infty)\right\|_\infty$$

since:

$$P(s) \cdot (1, 1, ..)^T = (1, 1, .. 1)^T$$

$$= \gamma \left\| \begin{pmatrix} \|\bar{v}_1 - \bar{v}_2\|_\infty \\ \|\bar{v}_1 - \bar{v}_2\|_\infty \\ \vdots \\ \|\bar{v}_1 - \bar{v}_2\|_\infty \end{pmatrix} \right\|_\infty$$

$$= \gamma \cdot \|\bar{v}_1 - \bar{v}_2\|_\infty$$

$$\Rightarrow \|T(\bar{v}_1) - T(\bar{v}_2)\|_\infty \leq \gamma \cdot \|\bar{v}_1 - v_2\|_\infty$$

Now that we know T is a γ-contraction, we can use the fact to find the fixed point and show that it is unique
(by using the Banach contraction principle)

Define a sequence $\{v_k\}$ in V by :

$$v_{k+1} = T(v_k) = T^{k+1}(v_0) \,,\; k \geq 0$$

Because T is γ-contraction, we have:

$$d(v_k, v_{k+1}) = d(T(v_{k-1}), T(v_k)) \leq \gamma \cdot d(v_{k-1}, v_k)$$

$$d(v_k, v_{k+1}) \leq \gamma \cdot d(v_0, v_1)$$

For any m, n such that $m > n$ it means

$$d(v_n, v_m) \leq \sum_{i=n}^{m-1} d(v_i, v_{i+1})$$

$$\leq \sum_{i=n}^{m-1} \gamma^i d(v_0, v_1) \leq \frac{\gamma^n}{1-\gamma} d(v_0, v_1)$$

(Cauchy criterion). In a complete metric space, a sequence is Cauchy if it converges

We can find N for any $\varepsilon > 0$ such that $d(x_n, x_m) < \varepsilon \ \forall\, m, n \geq N$:

$$d(v_n, v_m) \leq \frac{\gamma^n}{1-\gamma} d(v_0, v_1) < \varepsilon$$

$$\gamma^n < \varepsilon \cdot \frac{1-\gamma}{d(v_0, v_1)}$$

$$n > \log_\gamma\left(\varepsilon \cdot \frac{1-\gamma}{d(v_0, v_1)}\right)$$

$$N = \left\lceil \log_\gamma\left(\varepsilon \cdot \frac{1-\gamma}{d(v_0, v_1)}\right) \right\rceil$$

$$\Longrightarrow d(v_n, v_m) \leq \frac{\gamma^N}{1-\gamma} \cdot d(v_0, v_1) < \varepsilon$$

Because $\{v_k\}$ is a Cauchy sequence, it satisfies the Cauchy criterion and converges

$\longrightarrow$ There exists a convergence point x^*:

$$x^* = \lim_{k\to\infty} v_k = \lim_{k\to\infty} v_{k-1} = T(\lim_{k\to\infty} v_{k-1}) = T(x^*)$$

The limit of the iterative application of T on v_0 always converges to a fixed point x^* such that $T(x^*) = x^*$. But we already know one fixed point of the mapping, it is the solution $v_\pi = v_\pi(s)\ \forall s \in S$ to the Bellman expectation equation, which is the state-value function for an arbitrary policy π.

$$\forall v_0 \in V: \lim_{k \to \infty} T^k(v_0) = v_\pi \qquad \ldots \; v_\pi \in V$$

The last thing to show is that the fixed point is unique. Let x, y be fixed points of T, then:

$$d(x,y) = d(T(x), T(y)) \leq \gamma \cdot d(x,y)$$

$$d(x,y) \leq \gamma \cdot d(x,y)$$

$$(1-\gamma) \cdot d(x,y) \leq 0$$

$(1-\gamma) > 0$, thus $d(x,y) = 0$ and $x = y$