Bellman Expectation Equation as an Operator (1)

Fix a policy π and consider the *iterative policy evaluation* (value iteration)

$$v_{k+1}(s) = \sum_{a \in \mathcal{A}} \pi(a \mid s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_k(s') \right), \quad \text{for all } s \in \mathcal{S}.$$

We can write this in vector/matrix form

$$v^{k+1} = \boldsymbol{\mathcal{R}}^{\boldsymbol{\pi}} + \gamma \boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}} v^k,$$

wher $v^{k+1}, v^k \in \mathbb{R}^|\mathcal{S}|$ and

$$\boldsymbol{\mathcal{R}}^{\boldsymbol{\pi}} = \sum_{a \in \mathcal{A}} \pi(a \mid s) \mathcal{R}_s^a,$$

$$\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}} = \sum_{s' \in \mathcal{S}} \sum_{a \in \mathcal{A}} \pi(a \mid s) \mathcal{P}_{ss'}^a.$$

Bellman Expectation Equation as an Operator (2)

For the fixed policy π and discount factor $\gamma \in [0, 1)$, define the operator $T^{\pi} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$ with

$$T^{\pi}(v) = \boldsymbol{\mathcal{R}^{\pi}} + \gamma \boldsymbol{\mathcal{P}^{\pi}} v.$$

Given a policy π, T^{π} takes as input a value function $v \in \mathbb{R}^{|\mathcal{S}|}$, performs one Bellman expectation update according to policy π, and returns as output another value function $T^{\pi}(v) \in \mathbb{R}^{|\mathcal{S}|}$.

Bellman Expectation Update is a Contraction: Intuition

Consider two arbitrary value functions $v, u \in \mathbb{R}^{|\mathcal{S}|}$. Then,

$$T^{\pi}(v) = \boldsymbol{\mathcal{R}^{\pi}} + \gamma \boldsymbol{\mathcal{P}^{\pi}} v,$$
$$T^{\pi}(u) = \boldsymbol{\mathcal{R}^{\pi}} + \gamma \boldsymbol{\mathcal{P}^{\pi}} u.$$

Question: how do u, v look after the update (application of operator T^{π})? In particular, can we say something about whether they look more or less similar?

Intuition:

- T^{π} shifts u, v by a constant, $\boldsymbol{\mathcal{R}^{\pi}}$, which leaves their distance unaffected but also,
- updates u, v by applying one time the same policy π. A policy stipulates how to average over previous values. Moreover, the average of previous values shrinks by a weight less than 1 (discount factor γ).

The last step makes u, v more similar, i.e., it brings them *closer*.

To formalize this statement, we need a *metric* to measure *distance* in the space, $\mathbb{R}^{|\mathcal{S}|}$, of value functions.

Definitions: Metric and Contraction

We will measure distance in $\mathbb{R}^{|\mathcal{S}|}$ with the ∞-norm $d\left(v, u\right) := \|v - u\|_{\infty}$, where

$$\|v - u\|_{\infty} := \max_{s \in \mathcal{S}} \left| v\left(s\right) - u\left(s\right) \right|,$$

$$\|T^{\pi}\left(v\right) - T^{\pi}\left(u\right)\|_{\infty} := \max_{s \in \mathcal{S}} \left| T^{\pi}\left(v\right)\left(s\right) - T^{\pi}\left(u\right)\left(s\right) \right|.$$

This makes $\left(\mathbb{R}^{|\mathcal{S}|}, \|\cdot\|_{\infty}\right)$ a complete metric space[1].

We say that an operator $T^{\pi} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$ is a contraction mapping, iff

$$d\left(T^{\pi}\left(v\right), T^{\pi}\left(u\right)\right) \leq \lambda d\left(v, u\right),$$

for some $\lambda \in [0, 1)$. In our case, this is equivalent to

$$\|T^{\pi}\left(v\right) - T^{\pi}\left(u\right)\|_{\infty} \leq \lambda \|v - u\|_{\infty}$$

[1]See Appendix for precise definition.

Bellman Expectation Update is a Contraction: Formal

We will show that T^{π} is a contraction mapping whenever $\gamma \in [0, 1)$.

$$
\begin{aligned}
\| T^{\pi}(v) - T^{\pi}(u) \|_{\infty} &= \| \boldsymbol{\mathcal{R}^{\pi}} + \gamma \boldsymbol{\mathcal{P}^{\pi}} v - \boldsymbol{\mathcal{R}^{\pi}} - \gamma \boldsymbol{\mathcal{P}^{\pi}} u \|_{\infty} \\
&= \| \gamma \boldsymbol{\mathcal{P}^{\pi}} v - \gamma \boldsymbol{\mathcal{P}^{\pi}} u \|_{\infty} = \gamma \| \boldsymbol{\mathcal{P}^{\pi}} (v - u) \|_{\infty}.
\end{aligned}
$$

Let $n = |\mathcal{S}|$, then

$$
\boldsymbol{\mathcal{P}^{\pi}}(v - u) = \begin{pmatrix} p^{\pi}_{11} & p^{\pi}_{12} & \dots & p^{\pi}_{1n} \\ p^{\pi}_{21} & p^{\pi}_{22} & \dots & p^{\pi}_{2n} \\ \dots & \dots & \ddots & \dots \\ p^{\pi}_{n1} & p^{\pi}_{n2} & \dots & p^{\pi}_{nn} \end{pmatrix} \cdot \begin{pmatrix} v_1 - u_1 \\ v_2 - u_2 \\ \vdots \\ v_n - u_n \end{pmatrix} = \begin{pmatrix} \sum_{s=1}^{n} p^{\pi}_{1s}(v_s - u_s) \\ \sum_{s=1}^{n} p^{\pi}_{1s}(v_s - u_s) \\ \vdots \\ \sum_{s=1}^{n} p^{\pi}_{ns}(v_s - u_s) \end{pmatrix}
$$

$$
\leq \max_{s \in \mathcal{S}} |v_s - u_s| \cdot \mathbf{1}_n = \| v - u \|_{\infty} \cdot \mathbf{1}_n,
$$

since $\boldsymbol{\mathcal{P}^{\pi}}$ is a probability matrix and its rows sum up to 1.

Bellman Expectation Update is a Contraction: Formal

We will show that T^π is a contraction mapping whenever $\gamma \in [0, 1)$.

$$\|T^\pi(v) - T^\pi(u)\|_\infty = \|\cancel{\mathcal{R}^\pi} + \gamma \mathcal{P}^\pi v - \cancel{\mathcal{R}^\pi} - \gamma \mathcal{P}^\pi u\|_\infty$$

$$= \|\gamma \mathcal{P}^\pi v - \gamma \mathcal{P}^\pi u\|_\infty = \gamma \|\mathcal{P}^\pi (v - u)\|_\infty.$$

Let $n = |\mathcal{S}|$, then

$$\mathcal{P}^\pi (v - u) = \begin{pmatrix} p^\pi_{11} & p^\pi_{12} & \dots & p^\pi_{1n} \\ p^\pi_{21} & p^\pi_{22} & \dots & p^\pi_{2n} \\ \dots & \dots & \ddots & \dots \\ p^\pi_{n1} & p^\pi_{n2} & \dots & p^\pi_{nn} \end{pmatrix} \cdot \begin{pmatrix} v_1 - u_1 \\ v_2 - u_2 \\ \vdots \\ v_n - u_n \end{pmatrix} = \begin{pmatrix} \sum_{s=1}^n p^\pi_{1s}(v_s - u_s) \\ \sum_{s=1}^n p^\pi_{1s}(v_s - u_s) \\ \vdots \\ \sum_{s=1}^n p^\pi_{ns}(v_s - u_s) \end{pmatrix}$$

$$\leq \max_{s \in \mathcal{S}} |v_s - u_s| \cdot \mathbf{1}_n = \|v - u\|_\infty \cdot \mathbf{1}_n,$$

since $\mathcal{P}^\pi$ is a probability matrix and its rows sum up to 1.

Bellman Expectation Update is a Contraction: Formal

So

$$\|\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}\left(v-u\right)\|_\infty \leq \|\|v-u\|_\infty \cdot \mathbf{1}_n\|_\infty = \|v-u\|_\infty \cdot \|\mathbf{1}_n\|_\infty = \|v-u\|_\infty.$$

Hence,

$$\begin{aligned}\|T^\pi\left(v\right)-T^\pi\left(u\right)\|_\infty &= \|\cancel{\boldsymbol{\mathcal{R}}^{\boldsymbol{\pi}}}+\gamma\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}v-\cancel{\boldsymbol{\mathcal{R}}^{\boldsymbol{\pi}}}-\gamma\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}u\|_\infty \\ &= \|\gamma\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}v-\gamma\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}u\|_\infty = \gamma\|\boldsymbol{\mathcal{P}}^{\boldsymbol{\pi}}\left(v-u\right)\|_\infty \\ &\leq \gamma\|v-u\|_\infty,\end{aligned}$$

which implies that T^π is a contraction mapping iff $\gamma \in [0, 1)$.

$\gamma < 1$ is Sufficient but not Necessary Condition

Remark: Value iteration may bring u, v closer even for $\gamma = 1$. Informally:

- The unique inequality in the previous calculation is generally not tight.
- Whenever all elements of the last vector are strictly less than its absolute maximum (which is easy to achieve), an application of T^π on v, u brings u, v closer even for $\gamma = 1$.

Contraction Mapping Theorem

Theorem
If $T : X \to X$ is a contraction mapping on a complete metric space (X, d), then T has a unique fixed point, i.e., there exists exactly one $x \in X$ such that $T(x) = x$.

Complete metric space: a metric space (X, d) is complete if every Cauchy sequence converges to a point in X.

Cauchy sequence: $(x_n)_{n \in N} \subseteq X$ is a Cauchy sequence if for every positive real number $\epsilon > 0$ there is a positive integer $N(\epsilon)$ such that for all positive integers $m, n > N(\epsilon)$, it holds that $d(x_m, x_n) < \epsilon$.

Contraction Mapping Theorem: Proof (1)

Proof.

Step 1: The sequence $(x_n)_{n\in\mathbb{N}}$ with

$$x_{n+1} := Tx_n$$

is a Cauchy sequence. To see this, write $x_n = T^n x_0$, then

$$d\left(x_2, x_1\right) = d(Tx_1, Tx_0) \leq \gamma d(x_1, x_0)$$

and by induction

$$d\left(x_{n+1}, x_n\right) \leq \gamma^n d(x_1, x_0), \qquad \text{for every } n \in \mathbb{N}.$$

Contraction Mapping Theorem: Proof (2)

Hence, for any $n, m \in \mathbb{N}$, we have by the triangle inequality that

$$\begin{aligned} d\left(x_m, x_n\right) &\leq d\left(x_{n+1}, x_n\right) + d\left(x_{n+2}, x_{n+1}\right) + \cdots + d\left(x_m, x_{m-1}\right) \\ &\leq [\gamma^n + \gamma^{n+1} + \cdots + \gamma^m] d\left(x_1, x_0\right) \\ &= \gamma^n \cdot \left(\sum_{k=0}^{m-n} \gamma^k\right) d\left(x_1, x_0\right) \\ &\leq \gamma^n \cdot \left(\sum_{k=0}^{\infty} \gamma^k\right) d\left(x_1, x_0\right) \\ &= \frac{\gamma^n}{1-\gamma} d\left(x_1, x_0\right). \end{aligned}$$

Since $\frac{\gamma^n}{1-\gamma} \to 0$ as $n \to \infty$, x_n is a Cauchy sequence. (Here we used $\gamma \in [0, 1)$.)

Contraction Mapping Theorem: Proof (3)

Step 2: Since x_n is a Cauchy sequence and (X, d) is complete, there exists $x \in X$ such that $\lim_{n\to\infty} x_n = x$. This x is a fixed point of T. To see this, observe that since T is a contraction mapping, then T is also continuous (why?). Hence,

$$Tx = T \lim_{n\to\infty} x_n = \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} x_{n+1} = x.$$

This shows that x is a fixed point of T.

Step 3: To show that x is the unique fixed point of T, let $y \neq x \in X$ be another fixed point of T. Then,

$$0 < d(x, y) = d(Tx, Ty) \leq \gamma d(x, y) < d(x, y)$$

since $\gamma \in [0, 1)$, which leads to the contradiction $d(x, y) < d(x, y)$. This shows that $\lim x_n = x$ is the unique fixed point of T and concludes the proof.