MATH(4/7)406 (Control Theory)
Solution to Quiz 3 (Unit 5) - October 7, 2014.
Prepared by Brendan Patch Yoni Nazarathy

This quiz is directly based on HW5. This is a theorem from [Put94]:

Theorem 1 *Denote by V the normed linear space associated with discounted MDP and let $v^0 \in V$. Let $\{v^n\}$ denote iterates of value iteration. Let $\lambda \in (0,1)$ denote the discount factor. The value of the MDP is denoted by v^*_λ.*

The following global convergence rate properties hold for the value iteration algorithm:

1. *Convergence is linear at rate λ.*
2. *Its asymptotic average rate of convergence equals λ.*
3. *It converges $O(\lambda^n)$.*
4. *For all n,*
$$||v^n - v^*_\lambda|| \leq \frac{\lambda^n}{1-\lambda}||v^1 - v^0||.$$

Q1: What is the mathematical definition of "convergence is linear at rate λ"? What is the meaning in words of this? Write this in one or two sentences at most.

Q2: How about "asymptotic rate of convergence equals λ"? (Mathematical definition and meaning in words).

Q3: How about "it converges $O(\lambda^n)$"? (Mathematical definition and meaning in words).

Q4: Prove 4.

Solution:

1. The mathematical definition of "convergence is linear at rate λ" is:
   There exists a constant $\alpha \geq 1$ such that for all $n \in \mathbb{N}$,
$$||v^{n+1} - v^*_\lambda|| \ \leq \ K\,||v^n - v^*_\lambda||^\alpha\,, \tag{1}$$
   where λ is the smallest K such that (1) holds. [5]
   Some ways this could be expressed in words are:
   - "At each iteration of the sequence the difference between the current iterate and the asymptotic iterate is at most the difference between the previous iterate and the value of the MDP to the power of α greater than 0 and then multiplied by λ."

   [2]

2. The mathematical definition of "asymptotic average rate of convergence equals λ" is:

$$\limsup_{n\to\infty} \left[\frac{||\upsilon^n - \upsilon^*_\lambda||}{||\upsilon^0 - \upsilon^*_\lambda||} \right]^{1/n} = \lambda \, ,$$

where we assume $\upsilon_0 \neq \upsilon^*$. [5]

Some ways this could be expressed in words are:

- "The lowest upper bound on the geometric mean of the ratio of the difference between each iterate and the asymptotic iterate and the the difference between the initial iterate and the value of the MDP as the number of iterates goes to infinity is equal to λ."; OR
- "The difference between each iterate and the asymptotic iterate is eventually approximated by taking a scalar multiple of the previous term."

[2]

3. To say "it converges $O(\lambda^n)$" means, mathematically, that

$$\limsup_{n\to\infty} \frac{||\upsilon^n - \upsilon^*_\lambda||}{\lambda^n} < \infty \, .$$

[5]

Some ways this could be expressed in words are:

- "The difference between the current iterate and the value of the MDP grows slower than the function λ^n."

[2]

4. **Proof.** For any $n \geq 0$,

$$||\upsilon^n - \upsilon^*_\lambda|| = ||\upsilon^n - \upsilon^{n+1} + \upsilon^{n+1} - \upsilon^*_\lambda|| \leq \underbrace{||\upsilon^n - \upsilon^{n+1}||}_{(\clubsuit)} + \underbrace{||\upsilon^{n+1} - \upsilon^*_\lambda||}_{(\spadesuit)} \, , \qquad (2)$$

by the triangle inequality. [5]

We will make use of the Bellman operator L, defined by

$$L\,\upsilon \equiv \max_{d \in D^{MD}} \left\{ r_d + \lambda\, P_d\, \upsilon \right\} ,$$

where r_d is the reward vector associated with decision rule d and P_d is the transition probability matrix associated with the decision d. [1]

From lectures, υ^*_λ is a fixed point of V (i.e. $L\,\upsilon^*_\lambda = \upsilon^*_\lambda$). [2]

Thus, for $\spadesuit$,

$$||\upsilon^{n+1} - \upsilon^*_\lambda|| = ||L\,\upsilon^n - L\,\upsilon^*_\lambda|| \leq \lambda\, ||\upsilon^n - \upsilon^*_\lambda|| \, .$$

[2]

Further, since $\lambda \in (0,1)$ we have from lectures that L is a contraction mapping on V. [2]

Thus, for ♣,

$$
\begin{aligned}
||\upsilon^n - \upsilon^{n+1}|| = ||\upsilon^{n+1} - \upsilon^n|| = ||L\,\upsilon^n - L\,\upsilon^{n-1}|| &\le \lambda\,||\upsilon^n - \upsilon^{n-1}|| \\
&\le \lambda^2\,||\upsilon^{n-1} - \upsilon^{n-2}|| \\
&\vdots \\
&\le \lambda^n\,||\upsilon^1 - \upsilon^0||\,.
\end{aligned}
$$

[2]

Hence (2) becomes

$$||\upsilon^n - \upsilon_\lambda^*|| \le \lambda\,||\upsilon^n - \upsilon_\lambda^*|| + \lambda^n\,||\upsilon^1 - \upsilon^0||\,,$$

which, upon rearranging, yields the result. [1]