

COMP767 - Reinforcement Learning

Spectral radius and convergence

Mathieu Tanneau

École polytechnique de Montréal

January 27, 2017

1 Introduction

2 Theory

3 A numerical example

4 References

Recall Bellman equation :

$$v_{\pi}(s) = \sum_a \pi(a|s) \sum_{s', r} p(s' r | s, a) [r + \gamma v_{\pi}(s')]$$

or under its matrix form:

$$v_{\pi} = r_{\pi} + \gamma P_{\pi} v_{\pi}$$

which leads to the *policy evaluation* algorithm:

$$\forall k \geq 1, v_k = r_{\pi} + \gamma P_{\pi} v_{k-1}$$

Our topic: convergence analysis of policy evaluation

Idea:

$$\begin{aligned}
 v_k &= r_\pi + \gamma P_\pi v_{k-1} \\
 &= r_\pi + \gamma P_\pi r_\pi + \gamma^2 P_\pi^2 v_{k-1} \\
 &\vdots \\
 &= \underbrace{\left(\sum_{i=0}^{k-1} \gamma^i P_\pi^i \right) r_\pi}_{\rightarrow (I - \gamma P_\pi)^{-1} r_\pi} + \underbrace{\gamma^k P_\pi^k v_0}_{\rightarrow 0}
 \end{aligned}$$

1 Introduction

2 Theory

- Definitions
- Spectral radius
- Convergence analysis

3 A numerical example

4 References

Definition

- S set of possible states
- V the set of real-valued, bounded functions on S

$$\forall v \in V, \|v\| := \sup_{s \in S} |v(s)|$$

- $L(V)$ the set of bounded linear transformations on V

$$\forall Q \in L(V), \exists K > 0 : \forall v \in V, \|Qv\| \leq K\|v\| \quad (Q \text{ bounded})$$

$$\forall Q \in L(V), \|Q\| := \sup\{\|Qv\| \mid v \in V, \|v\| \leq 1\}$$

Theorem

*With norms as defined above,
if V is a Banach space, then $L(V)$, is a Banach space*

Definition

Let $Q \in L(V)$, the *spectral radius* of Q , noted $\sigma(Q)$, is:

$$\sigma(Q) := \lim_{n \rightarrow +\infty} |||Q^n|||^{1/n}$$

For any $Q \in L(V)$, we have $\sigma(Q) \leq |||Q|||$

Theorem

Let $Q \in L(V)$, with $\sigma(Q) < 1$. Then $(I - Q)^{-1}$ exists and:

$$(I - Q)^{-1} = \sum_{n=0}^{\infty} Q^n$$

See Puterman 1994 for proof

Recall the policy evaluation algorithm:

$$\begin{aligned}v_0 &= 0 \\v_k &= r_\pi + \gamma P_\pi v_{k-1} \\&= \left(\sum_{i=0}^{k-1} \gamma^i P_\pi^i \right) r_\pi + \gamma^k P_\pi^k v_0\end{aligned}$$

Hence the algorithm converges and:

$$\begin{aligned}v_\pi &= \lim_{k \rightarrow \infty} v_k \\&= (I - \gamma P_\pi)^{-1} r_\pi \\&= r_\pi + \gamma P_\pi v_\pi\end{aligned}$$

Definition

Define the error vector at iteration k as: $\epsilon_k = v_k - v_\pi$

Then $\epsilon_k = \gamma P_\pi \epsilon_{k-1} = \gamma^k P_\pi^k \epsilon_0$

Note that, since $\|P_\pi\| = 1$, then $\|\epsilon_k\| \leq \gamma^k \|\epsilon_0\|$

Definition (Average rate of convergence)

Define the average rate of convergence after k iterations:

$$R_k = -\ln \left[(\|\gamma^k P_\pi^k\|)^{1/k} \right]$$

Theorem

Assume $|\gamma| < 1$, then :

$$\lim_k R_k = -\ln [\sigma(\gamma P_\pi)]$$

- 1 Introduction
- 2 Theory
- 3 A numerical example**
- 4 References

cf Notebook

- 1 Introduction
- 2 Theory
- 3 A numerical example
- 4 References**



M.L. Puterman

Markov Decision Processes

1994



R.S Varga

Matrix Iterative Analysis

2000