Convergence and Spectral Radius

COMP 767 - Reinforcement Learning

Mathieu Nassif

McGill University

January 27, 2017

Policy Evaluation: Iterative Method (Matrix Form)

Bellman equation

$$v_{k+1}(s) = \sum_{a} \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_k(s')]$$

Policy Evaluation: Iterative Method (Matrix Form)

Bellman equation

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_k(s')]$$

$$v_{k+1} = R_\pi + \gamma P_\pi v_k$$

Convergence

Optimal Solution

$$v^* = (I - \gamma P_\pi)^{-1} R_\pi$$

Convergence

Optimal Solution

$$v^* = (I - \gamma P_\pi)^{-1} R_\pi$$

Approximation Error

$$v^* - v_{k+1} = R_\pi + \gamma P_\pi v^* - R_\pi - \gamma P_\pi v_k$$
$$v^* - v_{k+1} = \gamma P_\pi (v^* - v_k)$$
$$v^* - v_{k+1} = (\gamma P_\pi)^{k+1} (v^* - v_0)$$
$$v^* - v_{k+1} = e_{k+1} = (\gamma P_\pi)^{k+1} e_0$$

Convergence

Optimal Solution

$$v^* = (I - \gamma P_\pi)^{-1} R_\pi$$

Approximation Error

$$v^* - v_{k+1} = R_\pi + \gamma P_\pi v^* - R_\pi - \gamma P_\pi v_k$$
$$v^* - v_{k+1} = \gamma P_\pi (v^* - v_k)$$
$$v^* - v_{k+1} = (\gamma P_\pi)^{k+1} (v^* - v_0)$$
$$v^* - v_{k+1} = e_{k+1} = (\gamma P_\pi)^{k+1} e_0$$

Convergence

$$\lim_{k \to \infty} v_k = v^* \text{ if } \lim_{k \to \infty} (\gamma P_\pi)^k = 0$$

Spectral Radius

For a linear operator T, the spectral radius is the maximal magnitude of its eigenvalues.

$$\rho(T) = \max\{\lambda | \exists \mathbf{v} \text{ s.t. } T\mathbf{v} = \lambda\mathbf{v}\}$$

Spectral Radius

For a linear operator T, the spectral radius is the maximal magnitude of its eigenvalues.

$$\rho(T) = \max\{\lambda | \exists \mathbf{v} \text{ s.t. } T\mathbf{v} = \lambda \mathbf{v}\}$$

In our case,

$$\rho(\gamma P_\pi) = \gamma < 1$$

because P_π is stochastic!

Proof of Convergence

Proof of Convergence

Assume the eigenvectors of P_π form a basis of the vector space. We can express the original error, e_0 as a linear combination of the eigenvectors.

$$e_0 = c_1 x_1 + \ldots + c_N x_N$$

Proof of Convergence

Assume the eigenvectors of P_π form a basis of the vector space. We can express the original error, e_0 as a linear combination of the eigenvectors.

$$e_0 = c_1 x_1 + \ldots + c_N x_N$$

This assumption is not necessary, but it helps visualize the proof. Without this assumption, the proof would be essentially the same, at a more abstract level.

Proof of Convergence

Starting point

$$e_0 = c_1 x_1 + \ldots + c_N x_N$$

Proof of Convergence

Starting point

$$e_0 = c_1 x_1 + \ldots + c_N x_N$$

When we multiply...

$$\begin{aligned} e_1 &= \gamma P_\pi (c_1 x_1 + \ldots + c_N x_N) \\ &= c_1 \gamma P_\pi x_1 + \ldots + c_N \gamma P_\pi x_N \\ &= c_1 \lambda_1 x_1 + \ldots + c_N \lambda_N x_N \end{aligned}$$

Proof of Convergence

Starting point

$$e_0 = c_1 x_1 + \ldots + c_N x_N$$

When we multiply...

$$\begin{aligned} e_1 &= \gamma P_\pi (c_1 x_1 + \ldots + c_N x_N) \\ &= c_1 \gamma P_\pi x_1 + \ldots + c_N \gamma P_\pi x_N \\ &= c_1 \lambda_1 x_1 + \ldots + c_N \lambda_N x_N \end{aligned}$$

Again...

$$e_k = c_1 \lambda_1^k x_1 + \ldots + c_N \lambda_N^k x_N$$

Proof of Convergence

To Infinity!

$$\begin{aligned}\lim_{k\to\infty} ||e_k|| &= \lim_{k\to\infty} ||c_1\lambda_1^k x_1 + \ldots + c_N\lambda_N^k x_N|| \\ &\leqslant |c_1|(\lim_{k\to\infty} |\lambda_1|^k)||x_1|| + \ldots + |c_N|(\lim_{k\to\infty} |\lambda_N|^k)||x_N||\end{aligned}$$

Proof of Convergence

To Infinity!

$$\begin{aligned}\lim_{k\to\infty} ||e_k|| &= \lim_{k\to\infty} ||c_1\lambda_1^k x_1 + \ldots + c_N\lambda_N^k x_N|| \\ &\leqslant |c_1|(\lim_{k\to\infty} |\lambda_1|^k)||x_1|| + \ldots + |c_N|(\lim_{k\to\infty} |\lambda_N|^k)||x_N||\end{aligned}$$

Use the spectral radius

$$\begin{aligned}\rho(\gamma P_\pi) &< 1 & & \Leftrightarrow \\ |\lambda_i| &< 1 & \forall i & \Leftrightarrow \\ \lim_{k\to\infty} |\lambda_i|^k &= 0 & \forall i &\end{aligned}$$

Proof of Convergence

To Infinity!

$$\begin{aligned}\lim_{k\to\infty} ||e_k|| &= \lim_{k\to\infty} ||c_1\lambda_1^k x_1 + \ldots + c_N\lambda_N^k x_N|| \\ &\leqslant |c_1|(\lim_{k\to\infty} |\lambda_1|^k)||x_1|| + \ldots + |c_N|(\lim_{k\to\infty} |\lambda_N|^k)||x_N||\end{aligned}$$

Use the spectral radius

$$\begin{aligned}\rho(\gamma P_\pi) &< 1 & & \Leftrightarrow \\ |\lambda_i| &< 1 & \forall i & \Leftrightarrow \\ \lim_{k\to\infty} |\lambda_i|^k &= 0 & \forall i &\end{aligned}$$

Conclusion

$$\lim_{k\to\infty} ||e_k|| \leqslant |c_1| \times 0 \times ||x_1|| + \ldots + |c_N| \times 0 \times ||x_N|| = 0$$

Notes

A few points to note

Notes

A few points to note

- Relies on the powers of the matrix becoming smaller and smaller.

Notes

A few points to note

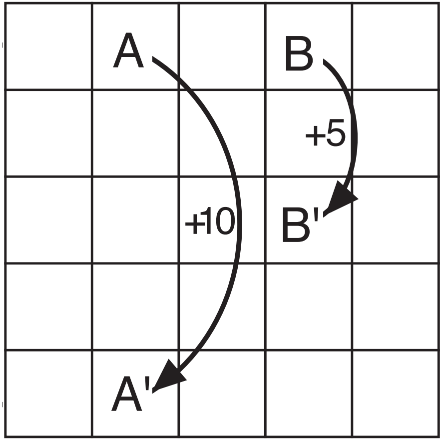
- Relies on the powers of the matrix becoming smaller and smaller.
- Convergence is linear at worst (and in the general case)

Notes

A few points to note

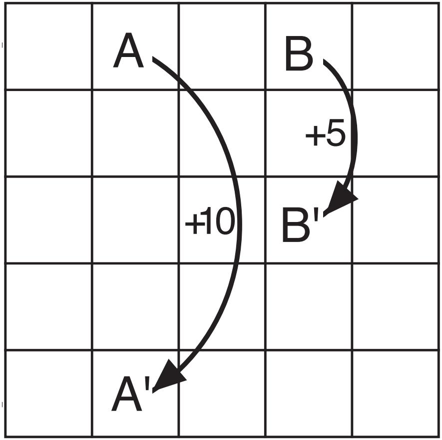
- Relies on the powers of the matrix becoming smaller and smaller.
- Convergence is linear at worst (and in the general case)
- Small spectral radius (γ) = Fast convergence

Experience: Gridworld (Example 3.8 of Sutton & Barto)



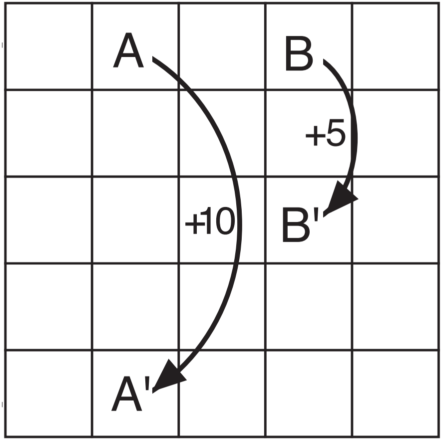
- 25 states
- 4 actions
- Rewards: -1/0/+5/+10

Experience: Gridworld (Example 3.8 of Sutton & Barto)



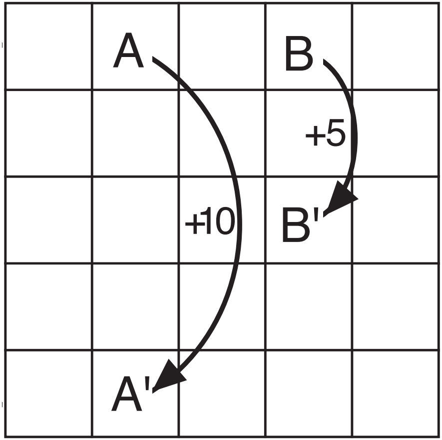
- 25 states
- 4 actions
- Rewards: -1/0/+5/+10
- Policy: Random (uniform)

Experience: Gridworld (Example 3.8 of Sutton & Barto)



- 25 states
- 4 actions
- Rewards: -1/0/+5/+10
- Policy: Random (uniform)
- v_0: all zeros

Experience: Gridworld (Example 3.8 of Sutton & Barto)



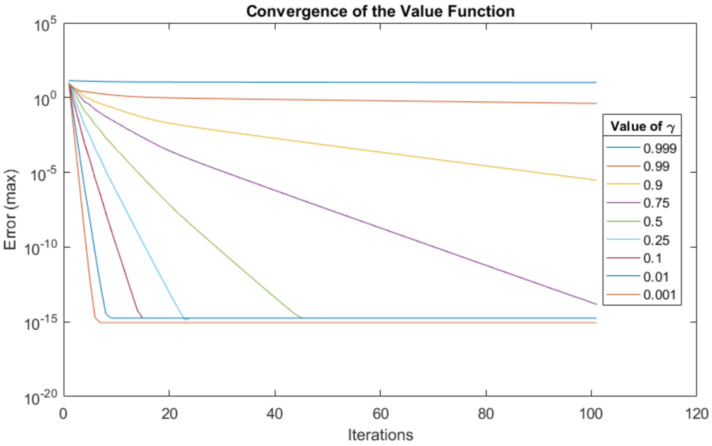
- 25 states
- 4 actions
- Rewards: -1/0/+5/+10
- Policy: Random (uniform)
- v_0: all zeros
- γ: variable

Successive Iterations of the Algorithm

Depending on γ, how does the error evolves for successive iterations?

Successive Iterations of the Algorithm

Depending on γ, how does the error evolves for successive iterations?

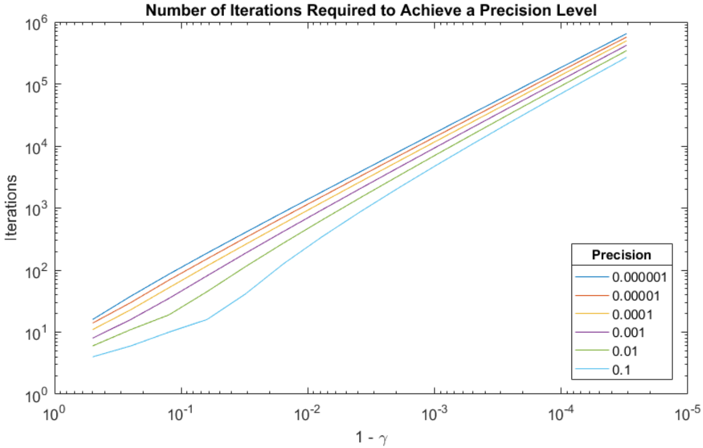


Number of Iterations to Obtain a Reasonable estimate

How many iterations are required to obtain a given precision?

Number of Iterations to Obtain a Reasonable estimate

How many iterations are required to obtain a given precision?



Proof of Convergence (General Case)

Relies on the following result

$$\rho(M) < 1 \qquad \Leftrightarrow \qquad \lim_{k\to\infty} M^k = 0$$

Then,

$$\lim_{k\to\infty} e_k = \lim_{k\to\infty} (\gamma P_\pi)^k e_0 = 0$$

Proof of Convergence (General Case)

Express M in its Jordan normal form

$$M = VJV^{-1}$$

$$M^k = VJ^kV^{-1}$$

$$\lim_{k\to\infty} J^k = 0$$

$$\lim_{k\to\infty} M^k = 0$$