Policy evaluation convergence proof by spectral radius

Pierre Thodoroff and Jean Harb

January 27, 2017

1 Policy Evaluation

We use the following notations : T is the bellman operator, P_d^π is the probability transition matrix under the policy π, b the reward vector, γ the discount factor and v the state values . We want to find v such that :

$$Tv = v$$

$$v = b + \gamma P_d^\pi v$$

$$(I - \gamma P_d^\pi)v = b$$

Which has the the same form as $Ax = b$ where $A = (I - \gamma P_d)$.
We solve this system using matrix splitting. We first split A such that

$$A = M - N$$

where M is non singular. From matrix splitting theory an approximate solution can be obtained using the following iterative method

$$v^{t+1} = M^{-1}Nv^t + M^{-1}b$$

We will now prove why the iterative solution converge to the solution.

2 Proof

We want to solve for A.

$$Ax = b$$

We can do so in closed form or as an iterative method by splitting A.

$$A = M - N$$

Giving us

$$x(M - N) = b$$

$$x = M^{-1}Nx + M^{-1}b$$

We can iterate this until convergence, where each iteration is given by

$$x^{(k+1)} = M^{-1}Nx^k + M^{-1}b$$

We calculate the error at iteration k as

$$e^{(k)} = x^* - x^{(k)}$$

If we isolate b, we get

$$b = Mx^{(k+1)} - Nx^k$$

and

$$b = Mx^* - Nx^*$$

giving us

$$Mx^{(k+1)} - Nx^k = Mx^* - Nx^*$$
$$Nx^* - Nx^k = Mx^* - Mx^{(k+1)}$$
$$N(x^* - x^k) = M(x^* - x^{(k+1)})$$
$$Ne^{(k)} = Me^{(k+1)}$$
$$e^{(k+1)} = M^{-1}Ne^{(k)}$$

Let's call $M^{-1}N = G$.

$$e^{(k+1)} = Ge^{(k)}$$

and as a recursion, we get

$$e^{(k)} = G^k e^{(0)}$$

We want to prove that as $k \to \infty$, $e^{(0)} \to 0$. Let's start by getting the hypothetical set of eigenvalues v of G, such that

$$Gv_i = \lambda_i v_i$$

So let's represent $e^{(0)}$ as a linear combination of the eigenvectors of G, where c_i is the coefficient weighting of each basis vector.

$$e^{(0)} = c_1 v_1 + ... + c_n v_n$$

If we apply one iteration of G on $e^{(0)}$, we get

$$Ge^{(0)} = c_1 G v_1 + ... + c_n G v_n = c_1 \lambda_1 v_1 + ... + c_n \lambda_n v_n$$

So applying G k times, we get

$$G^k e^{(0)} = c_1 G^k v_1 + ... + c_n G^k v_n = c_1 \lambda_1^k v_1 + ... + c_n \lambda_n^k v_n$$

Therefore as $k \to \infty$, each $\lambda_i \to 0$ iff $|\lambda_i| < 1$. There's a theorem saying that eigenvalues of a matrix cannot be larger than the largest sum of any row. As we're dealing with a probability matrix, the sum of all rows is 1. This isn't enough to prove convergence, as we need the eigenvalues to be strictly less than one. But if we use a $\gamma < 1$, then we get the eigenvalues of γP, which will be strictly less than one, proving convergence.

Finally, we also notice that the smaller γ is, the faster we'll converge. We show this empirically in code attached to this document.

3 Discussion

Different splitting of A yields different methods. The speed of convergence of the different method is often analyzed using the spectral radius of $M^{-1}N$.

3.1 Splitting for regular policy evaluation

The first splitting we will examine is where M = I and N = γP_D^π.

$$A = M - N = I - \gamma P_d^\pi$$

It yields the classical policy evaluation method

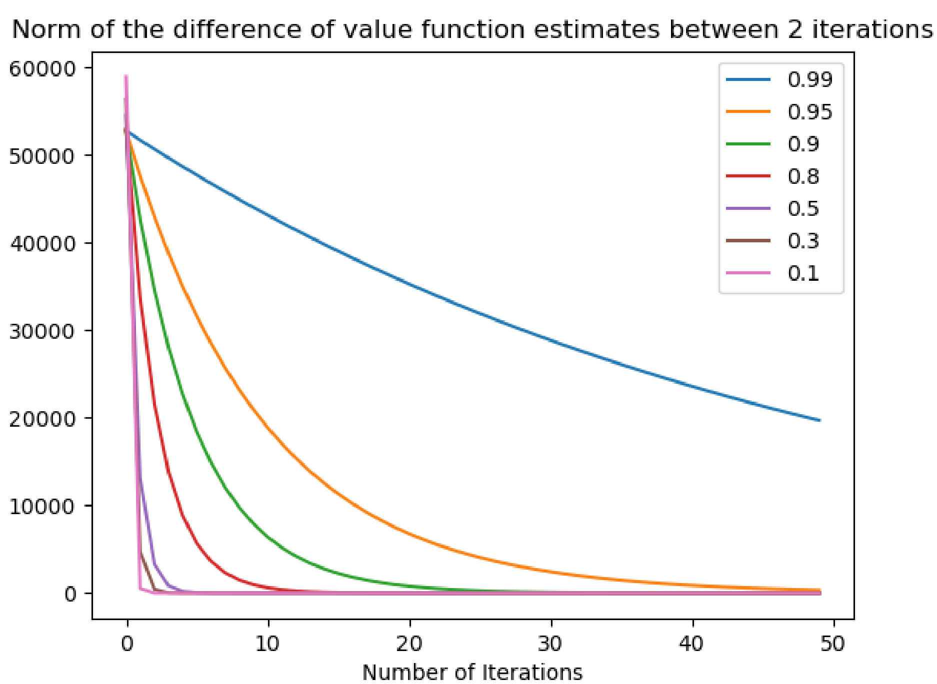
$$v^{t+1} = I^{-1}\gamma P_d^\pi v^t + I^{-1}b$$

$$v^{t+1} = \gamma P_d^\pi v^t + b$$

By studying the spectral radius of $M^{-1}N$ we can get a rate of convergence. In the classical policy evaluation case

$$\rho(M^{-1}N) = \rho(I^{-1}\gamma P_d^\pi) = \gamma\rho(P_d^\pi) = \gamma * 1$$

This is because the largest eigenvalue of a square matrix where each row sum to one is equal to 1. In this setting the spectral radius and rate of convergence is determined by γ. We then study experimentally the impact of γ on the convergence rate of a small random MDP.



3.2 Gausss-Seidel splitting

Let's define $P_d^\pi = L(P_d^\pi) + U(P_d^\pi)$ where

$$L(P_d^\pi) = \begin{matrix} 0 & 0 & 0 & 0 \\ p_{21} & 0 & 0 & 0 \\ p_{31} & p_{32} & 0 & 0 \\ p_{41} & p_{42} & p_{43} & 0 \end{matrix}$$

$$U(P_d^\pi) = \begin{matrix} p_{11} & p_{12} & p_{13} & p_{14} \\ 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & 0 & p_{44} \end{matrix}$$

We now define the splitting of A such that

$$A = M - N = (I - \gamma L(P_d^\pi)) - \gamma U(P_d^\pi)$$

where $M = (I - \gamma L(P_d^\pi))$ and $N = \gamma U(P_d^\pi)$

The iterative procedure can then be written as

$$v^{t+1} = (I - \gamma L(P_d^\pi))^{-1} \gamma U(P_d^\pi) v^t + (I - \gamma L(P_d^\pi))^{-1} b$$

Once again we can study the convergence rate of this algorithm by studying $\rho(M^{-1}N)$