Problem Set 1

Dynamic Programming

Patrick Altmeyer

09 June, 2021

Contents

1 Policy evaluation

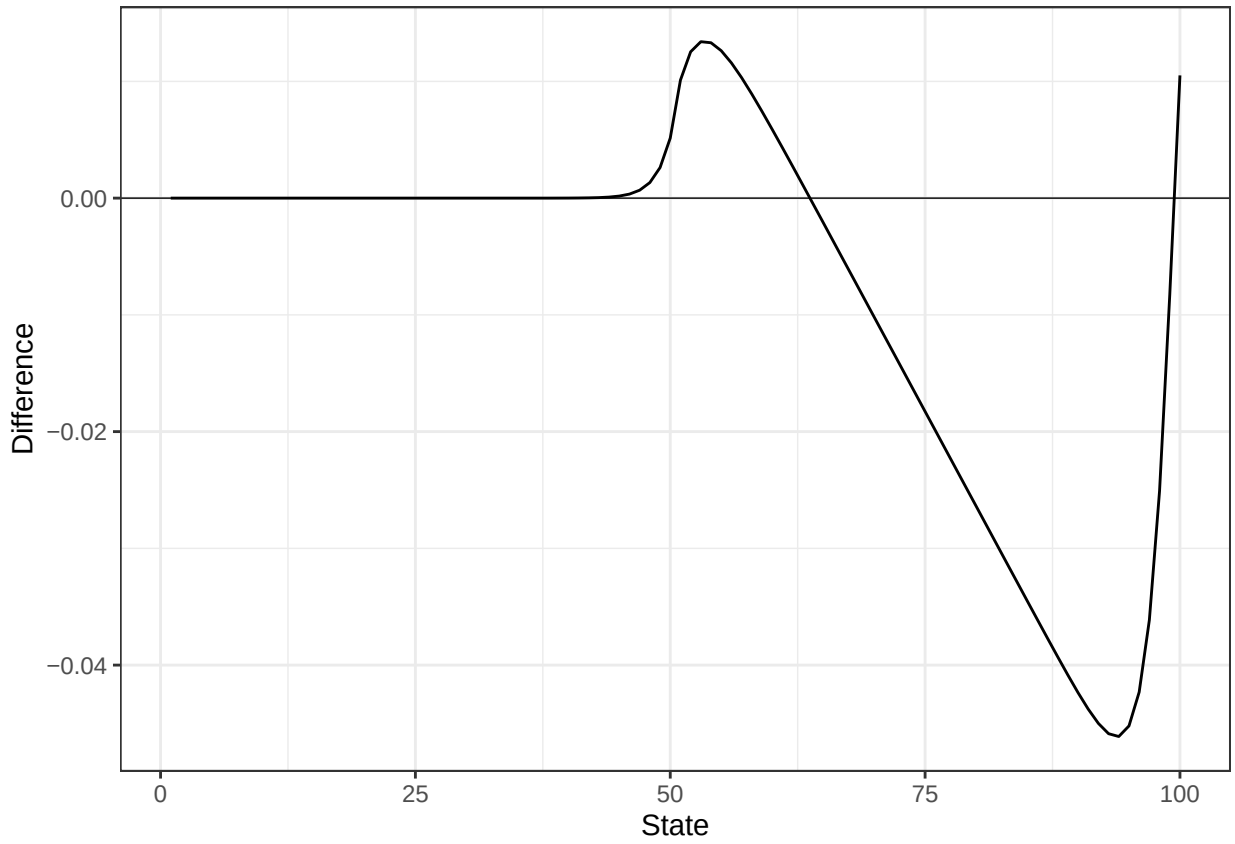

Figure 1: Comparison of lazy and aggressive policy.

Table 1: Number of iterations until convergence or timeout is reached.

Value Iteration	Policy Iteration
10	3
20	3
50	4
90	3

2 Value Iteration and Policy Iteration

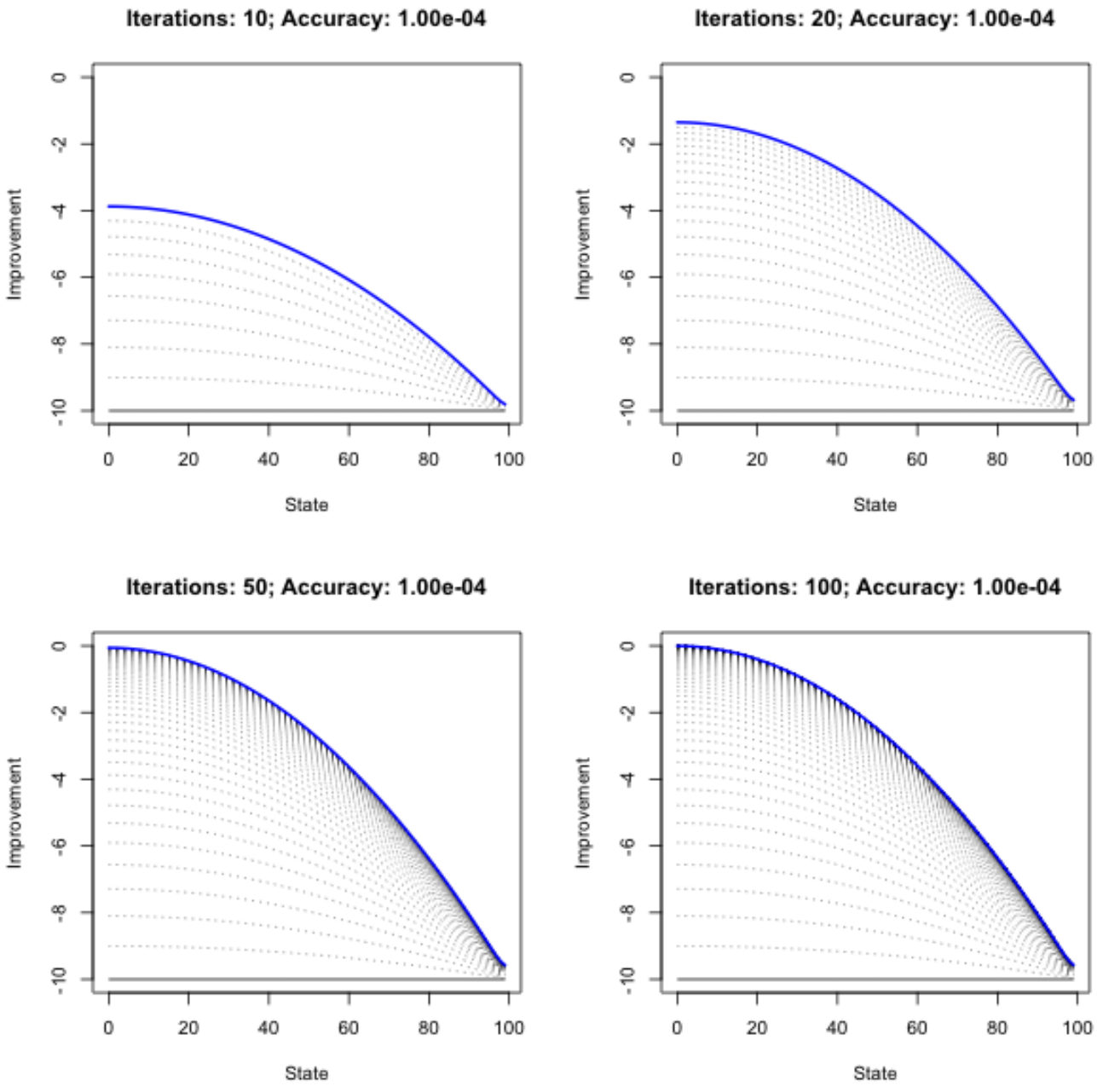

Figure 2: Value Iteration with 10, 20, 50 and 100 iterations. The blue line represents the value function corresponding to the final estimate of the optimal value function.

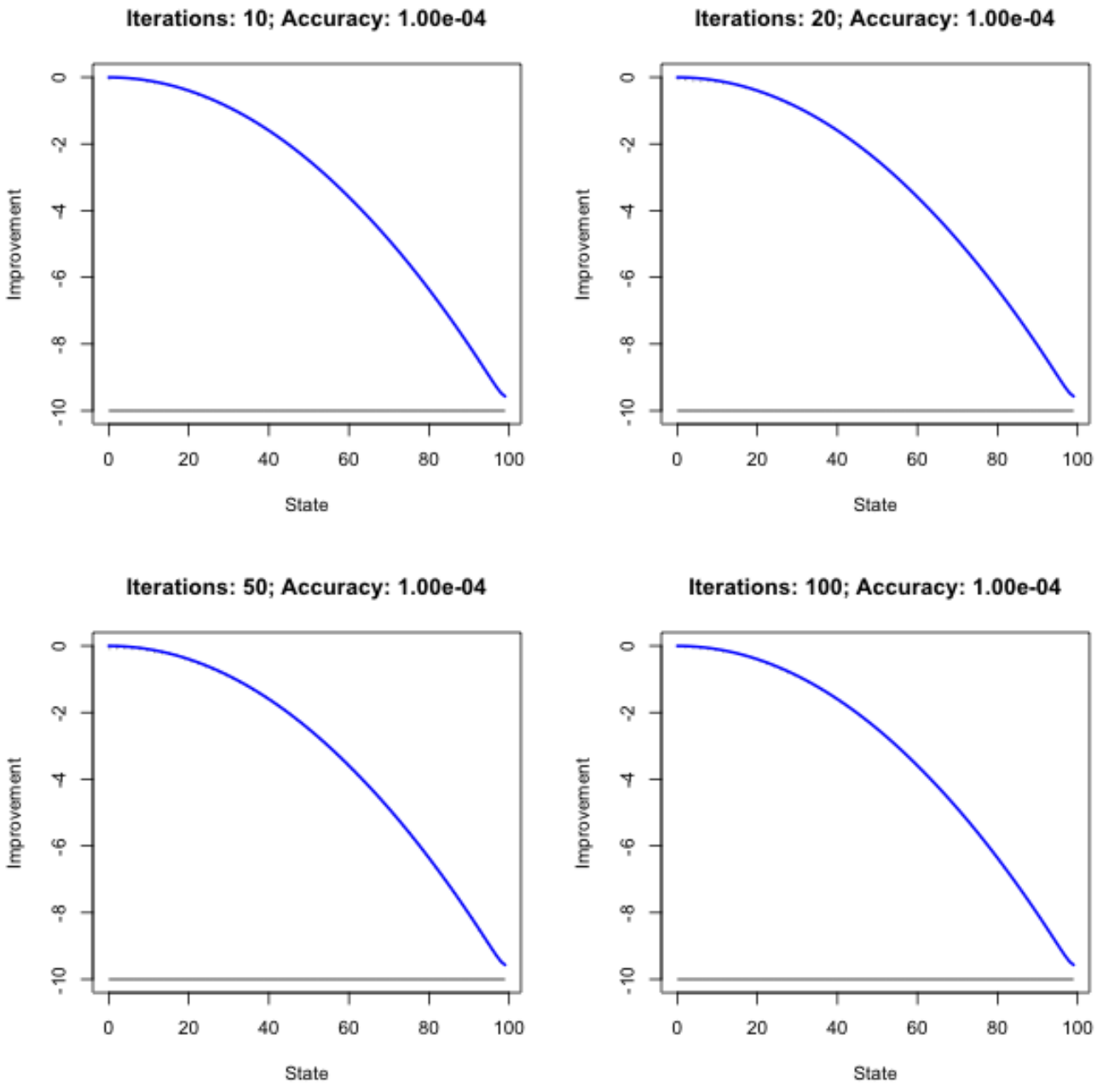

Figure 3: Policy Iteration with 10, 20, 50 and 100 iterations. The blue line represents the value function corresponding to the final estimate of the optimal value function.

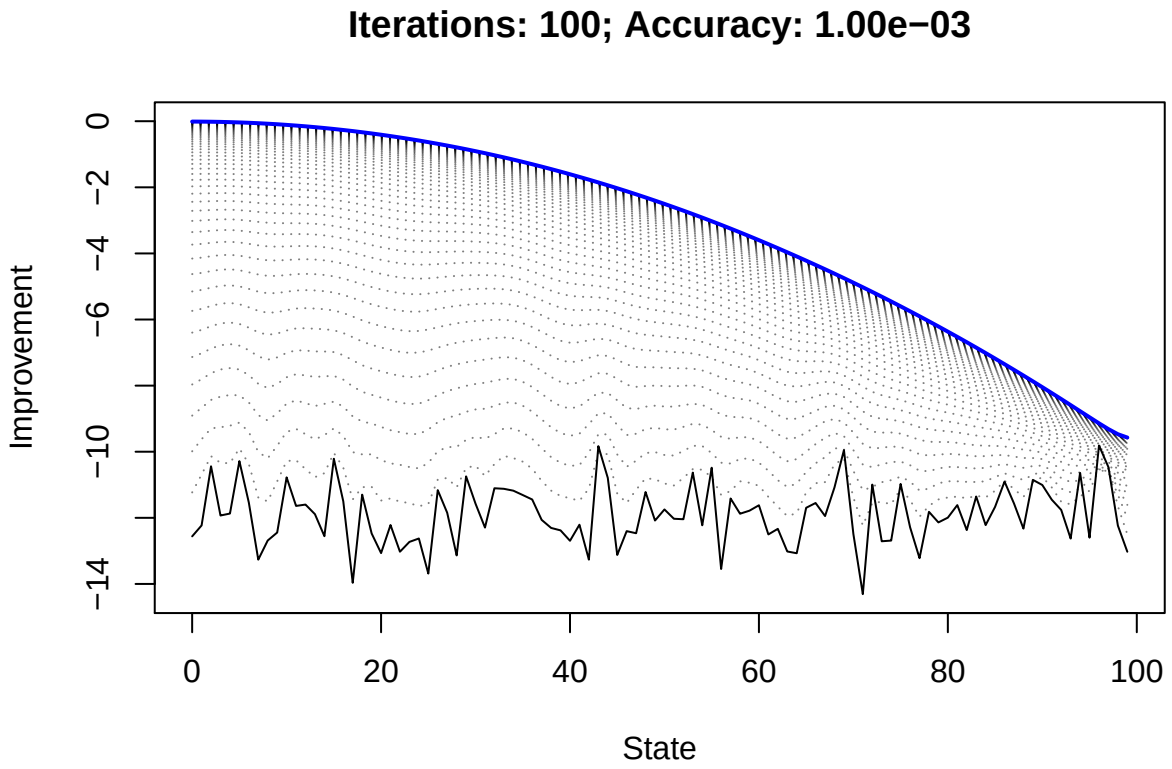

Figure 4: Initializing with random values.

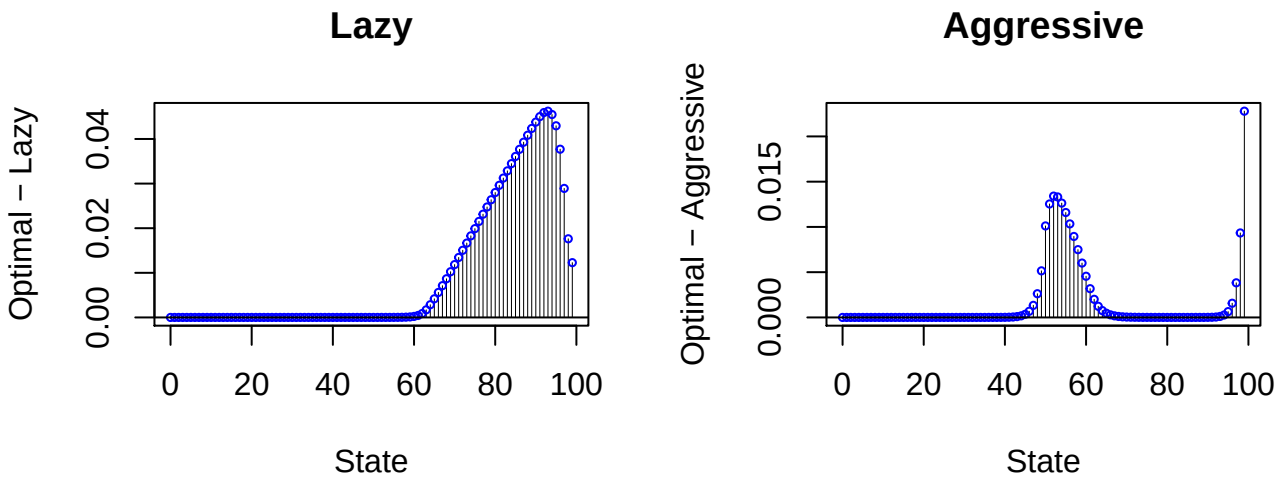

Figure 5: Improvement of estimated optimal policy compared to lazy and aggressive policy.

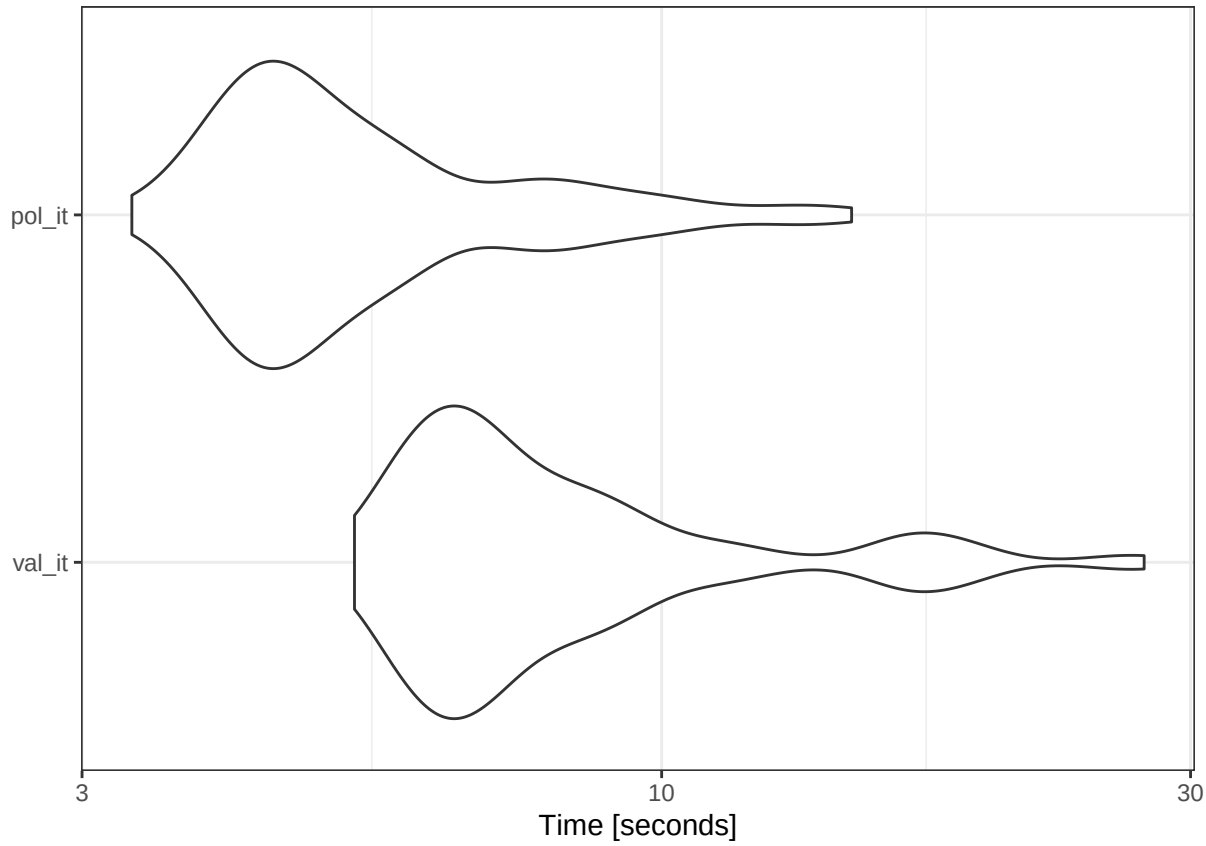

Figure 6: Microbenchmark of computational times for Value Iteration and Policy Iteration.