Implementation: Truncated Policy Iteration

In the previous concept, you learned about **truncated policy evaluation**. Whereas (iterative) policy evaluation applies as many Bellman updates as needed to attain convergence, truncated policy evaluation only performs a fixed number of sweeps through the state space.

The pseudocode can be found below.

Truncated Policy Evaluation

Input: MDP, policy π, value function V, positive integer $max\_iterations$
Output: $V \approx v_\pi$ (if $max\_iterations$ is large enough)
$counter \leftarrow 0$
while $counter < max\_iterations$ **do**
  for $s \in \mathcal{S}$ **do**
    $V(s) \leftarrow \sum_{a \in \mathcal{A}(s)} \pi(a|s) \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p(s', r|s, a)(r + \gamma V(s'))$
  end
  $counter \leftarrow counter + 1$
end
return V

We can incorporate this amended policy evaluation algorithm into an algorithm similar to policy iteration, called **truncated policy iteration**.

The pseudocode can be found below.