Algorithm 2 Policy Iteration

1: Initialize π randomly.
2: **for** until convergence **do**
3: Let $V := V^{\pi}$. ▷ typically by linear system solver
4: For each state s, let

$$\pi(s) := \arg\max_{a \in A} \sum_{s'} P_{sa}(s') V(s').$$