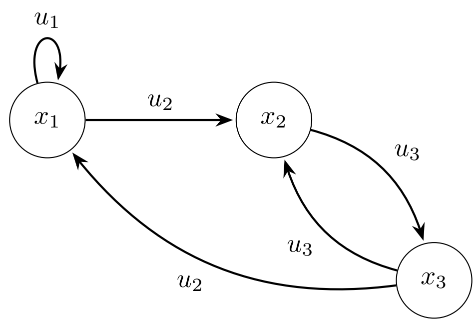
Tutorial 7, Mar 2, 2026



- Example: consider the finite state discrete time system above, with the terminal cost being zero for all states, and all other costs are 1, except $r(x_1, u_1) = 0$; solve for the optimal control $\mu^*(\cdot)$ using policy iteration
  - Initialize with some policy: $\mu^0(x_1) = u_2, \mu^0(x_2) = u_3, \mu^0(x_3) = u_3$
  - First iteration:
    * Policy evaluation: $\begin{cases} V^{\mu^0}(x_1) = r(x_1, u_2) + \gamma V^{\mu^0}(x_2) = 1 + \gamma V^{\mu^0}(x_2) \\ V^{\mu^0}(x_2) = r(x_2, u_3) + \gamma V^{\mu^0}(x_3) = 1 + \gamma V^{\mu^0}(x_3) \\ V^{\mu^0}(x_3) = r(x_3, u_3) + \gamma V^{\mu^0}(x_2) = 1 + \gamma V^{\mu^0}(x_2) \end{cases}$
      - Using this we can solve for the values of V^{μ^0} by substituting $V^{\mu^0}(x)$ and treating it like a system of equations
        - $V^{\mu^0}(x_3) = \dfrac{1+\gamma}{1-\gamma^2} = \dfrac{1}{1-\gamma}$
        - This turns out to be the same for all other states
    * Policy improvement:
      - $\mu^1(x_1) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x_1, u_1) + \gamma V^{\mu^0}(f(x_1, u_1)), r(x_1, u_2) + \gamma V^{\mu^0}(f(x_1, u_2)) \}$

        $= \underset{u \in \mathcal{U}(x)}{\arg\min} \left\{ \dfrac{\gamma}{1-\gamma}, 1 + \dfrac{\gamma}{1-\gamma} \right\}$

        $= u_1$
      - $\mu^1(x_2) = u_3$ since it is the only input in that state
      - $\mu^1(x_3) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x_3, u_3) + \gamma V^{\mu^0}(f(x_3, u_3)), r(x_3, u_2) + \gamma V^{\mu^0}(f(x_3, u_2)) \}$

        $= \underset{u \in \mathcal{U}(x)}{\arg\min} \{ 1 + \dfrac{\gamma}{1-\gamma}, 1 + \dfrac{\gamma}{1-\gamma} \}$

        $= u_2$
    * Since $\mu^1(x_1) \neq \mu^0(x_1)$ and $\mu^1(x_3) \neq \mu^0(x_3)$, we must iterate again until we converge
  - Second iteration:
    * Policy evaluation: $\begin{cases} V^{\mu^1}(x_1) = r(x_1, u_1) + \gamma V^{\mu^1}(x_1) = \gamma V^{\mu^1}(x_1) \\ V^{\mu^1}(x_2) = r(x_2, u_3) + \gamma V^{\mu^1}(x_3) = 1 + \gamma V^{\mu^1}(x_3) \\ V^{\mu^1}(x_3) = r(x_3, u_2) + \gamma V^{\mu^1}(x_1) = 1 + \gamma V^{\mu^1}(x_1) \end{cases}$
      - Solve to get $V^{\mu^1}(x_1) = 0, V^{\mu^1}(x_2) = 1 + \gamma, V^{\mu^1}(x_3) = 1$
    * Policy improvement:
      - $\mu^2(x_1) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x_1, u_1) + \gamma V^{\mu^1}(x_1), r(x_1, u_2) + \gamma V^{\mu^1}(x_2) \}$

        $= \underset{u \in \mathcal{U}(x)}{\arg\min} \{ 0, 1 + \gamma(1+\gamma) \}$

        $= u_1$
      - $\mu^2(x_2) = u_3$ again since it's the only option

- $\mu^2(x_3) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x_3, u_3) + \gamma V^{\mu^2}(x_2), r(x_3, u_2) + \gamma V^{\mu^2}(x_1) \}$

  $= \underset{u \in \mathcal{U}(x)}{\arg\min} \{ 1 + \gamma(1+\gamma), 1 \}$

  $= u_2$

  * After the second iteration, $\mu^1 = \mu^2$ for all values of x, so now we can stop since we've reached a stationary point

– To double check, we can do policy evaluation again to ensure that nothing changes