Lecture 20, Feb 26, 2026

Optimal Control With Stationary Policies

- Consider a stationary policy $\{\mu, \mu, \dots\} \in \mathcal{P}$, i.e. the feedback is independent of time, which is simply denoted $\mu \in \mathcal{M}$
- The cost is now $V^\mu(x_0) = \sum_{x=0}^{\infty} \gamma^k r(x(k), \mu(x(k)))$

$$\begin{aligned} &= r(x_0, \mu(x_0)) + \sum_{k=1}^{\infty} \gamma^k r(x(k), \mu(x(k))) \\ &= r(x_0, \mu(x_0)) + \gamma \sum_{k=0}^{\infty} \gamma^k r(x(k+1), \mu(x(k+1))) \\ &= r(x_0, \mu(x_0)) + \gamma V^\mu(f(x_0, \mu(x_0))) \end{aligned}$$

  - This is known as the *Bellman equation* for this stationary policy $\mu(x)$
- Given the optimal cost, the optimal policy is $\mu^*(x) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x,u) + \gamma V^*(f(x,u)) \}$
  - In this case, we require that a minimum exists (hence the arg min instead of inf)
  - This optimal control is not necessarily unique
- Denote $(T^\mu V)(x) = r(x, \mu(x)) + \gamma V(f(x, \mu(x)))$ (we can think of T^μ as an operator)
- Denote $(TV)(x) = \inf_{u \in \mathcal{U}(x)} r(x,u) + \gamma V(f(x,u))$
- Using this notation, the HJB equation is simply $V^* = TV^*$, and the Bellman equation is $V^\mu = T^\mu V^\mu$
  - Note we omit x to denote $V^*(x) = TV^*(x), \forall x \in \mathcal{X}$
  - We can see that the optimal cost is a fixed point of the T operator, which is important since we can use an iterative procedure to solve this – starting with some initial guess, keep applying the operator until convergence
- This gives rise to 2 iterative numerical schemes for solving the HJB equation: *value iteration* and *policy iteration*
- Value iteration algorithm:
  1. Initialization: select $V^0 \geq 0$ (i.e. positive at all values of x)
  2. Value iteration: $V^{j+1} = TV^j \iff V^{j+1}(x) = \inf_{u \in \mathcal{U}(x)} \{ r(x,u) + \gamma V^j(f(x,u)) \}$
     - Practically, to do this we require discretization of the state space $\mathcal{X}$ or some other technique, since $\mathcal{X}$ is often infinite
- Policy iteration algorithm:
  1. Initialization: select any admissible feedback $\mu^0 = \mathcal{M}$
  2. Policy evaluation: compute V^{μ^j} by solving the Bellman equation, $V^{\mu^j}(x) = r(x, \mu^j(x)) + \gamma V^{\mu^j}(f(x, \mu^j(x)))$
     - We have techniques to make this computationally tractable
  3. Policy improvement: $\mu^{j+1}(x) = \underset{u \in \mathcal{U}(x)}{\arg\min} \{ r(x,u) + \gamma V^{\mu^j}(f(x,u)) \}$
     - For each value of x, we select the value of u that would lead to the best cost
- For both algorithms, we iterate until we reach a stationary point, i.e. the value function or policy no longer changes