Tutorial 9, Mar 16, 2026

- Consider the infinite horizon optimal control problem: $\min_{u \in \mathcal{U}} \{ \gamma^k r(x(k), u(k)) \}$ such that $x(k+1) = ax(k) + bu(k), a, b \in \mathbb{R}$ and unknown
  - Consider an arbitrary policy μ and suppose $Q^\mu(x,u) = \psi_\mu^T w(x,u)$ where $w(x,u) = \begin{bmatrix} x^2 \\ xu \\ ux \\ u^2 \end{bmatrix}$; express the Bellman equation for Q^μ as a function of $\psi_\mu^T w$
    * $\psi_\mu^T w(x(k), u(k)) = r(x(k), u(k)) + \gamma Q^\mu(ax(k) + bu(k), \mu(ax(k) + bu(k)))$
  - Show that for any $u(k)$, $w(k) = w(x(k), u(k))$ is not PE
    * To show that the regressor is not PE, we just need to show that there exists y such that $y^T w(k) w^T(k) y = 0$ for all k, since then the sum will also be zero
    * $w^T(k) y = \begin{bmatrix} x^2 & xu & ux & u^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$
    * Take $y = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$, so $w^T(k) y = 0$ and therefore $y^T w(k) w^T(k) y = 0$
    * Therefore $\frac{1}{N} \sum w(k) w^T(k)$ is not positive definite, so the regressor is not PE
  - Now we want to learn the Q-function
    * Let the TD error $e(k) = r(x,u) + \gamma Q^\mu(f(x,u), \mu(f(x,u))) - Q^\mu(x,u) = 0$
    * Express the stage cost as a product of an unknown parameter and a regressor
      - $$\begin{aligned} r(x,u) &= -(\gamma Q^\mu(f(x,u), \mu(f(x,u))) - Q^\mu(x,u)) \\ &= -(\gamma \psi_\mu^T w(x(k+1), u(k+1)) - \psi_\mu^T w(x(k), u(k)) \\ &= -\psi_\mu^T \left(\gamma w(x(k+1), u(k+1)) - w(x(k), u(k)) \right) \\ &= -\psi_\mu^T v(k) \end{aligned}$$
    * Suppose we have an estimate $\hat{\psi}_\mu$ of ψ_μ and let $\hat{Q}^\mu(x,u) = \hat{\psi}_\mu^T w(k)$; find the TD error $e(k)$ in terms of $\tilde{\psi} = \hat{\psi}_\mu - \psi_\mu$. What is the error model?
      - $$\begin{aligned} e(k) &= \gamma \hat{Q}^\mu(x(k+1), u(k+1)) - \hat{Q}^\mu(x(k), u(k)) + r(k) \\ &= \hat{\psi}_\mu^T v(k) - \psi_\mu^T v(k) \\ &= \tilde{\psi}_\mu^T v(k) \end{aligned}$$
      - This is a static error model
  - Show that for any u, $v(k) = \gamma w(k+1) - w(k)$ is not PE
    * This goes the exact same as proving $w(k)$ is not PE, with the same y
  - Suppose $\hat{\psi}_{ss} = \lim_{k \to \infty} \hat{\psi}_\mu$ satisfies $Q^\mu(x,u) = \hat{\psi}_{ss} w(x,u)$; express S_μ in terms of $\hat{\psi}_{ss}$ such that $Q^\mu(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^T S_\mu \begin{bmatrix} x \\ u \end{bmatrix}$
    * $\hat{Q}^\mu = \hat{\psi}_{ss}^T w(k) = \hat{\psi}_1 x^2 + (\hat{\psi}_2 + \hat{\psi}_3) ux + \hat{\psi}_4 u^2$
    * $\begin{bmatrix} x & u \end{bmatrix} \begin{bmatrix} s_1 & s_2 \\ s_3 & s_4 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = s_1 x^2 + 2 s_2 ux + s_3 u^2$
    * Therefore $S_\mu = \begin{bmatrix} \hat{\psi}_1 & \frac{1}{2}(\hat{\psi}_2 + \hat{\psi}_3) \\ \frac{1}{2}(\hat{\psi}_2 + \hat{\psi}_3) & \hat{\psi}_4 \end{bmatrix}$