Dynamic Programming Lecture #6

Outline:

- Worst case DP
- Stochastic DP preview

Deterministic DP Review

- System:

$$x_{k+1} = f_k(x_k, u_k)$$

  - State: $x_k \in S_k$
  - Control (decision): $u_k \in U_k(x_k)$

- Policy shorthand: $\pi = \{\mu_0, \mu_1, ..., \mu_{N-1}\}$

$$\mu_k : x_k \to u_k \in U_k(x_k)$$

- Cost of policy π:

$$J_\pi(x_0) = g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k))$$

- Optimization:

$$J^*(x_0) = \min_\pi J_\pi(x_0)$$

- Value iteration:

$$J_N(x_N) = g_N(x_N)$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} g(x_k, u_k) + J_{k+1)}(f_k(x_k, u_k))$$

- Principle of optimality:

$$J^*(x_0) = J_0(x_0)$$

$$J_k(x_k) = \text{optimal cost-to-go at stage } k$$

$$\mu_k^*(x_k) = \arg\min g(x_k, u_k) + J_{k+1}(f_k(x_k, u_k))$$

Minimax/worst-case formulation

- Setup:

$$x_{k+1} = f_k(x_k, u_k, w_k)$$
$$u_k \in U_k(x_k)$$
$$w_k \in W_k(x_k, u_k)$$

- Cost of policy:

$$J_\pi(x_0) = \max_{w_0, w_1, \ldots, w_{N-1}} g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)$$

- Optimization:

$$J^*(x_0) = \min_\pi J_\pi(x_0)$$

- New element: *Adversarial Disturbance*
  - Disturbance seeks to maximize cost
  - Control commits to *policy* before disturbance acts
  - Very different from control commits to *actions*
  - Disturbance can be constrained by state/control
- Motivation:
  - Design guarantees/verifiable performance
  - Strategic interaction

Examples

- Disturbance rejection:

$$x_{k+1} = Ax_k + Bu_k + Lw_k$$
$$|w_k| \leq 1$$

  Objective:

$$\min_{\pi} \max_{w} \max_{k \geq 0} |Cx_k|$$

- Switching systems:

$$x_{k+1} = A(w_k)x_k + Bu_k$$
$$A(w_k) \in \left\{A^1, A^2, ..., A^m\right\}$$

  Objective:

$$\min_{\pi} \max_{w} \max_{k \geq 0} |Cx_k|$$

- More sophisticated disturbance model:

$$|w_{k+1} - w_k| \leq \rho$$

- "Disturbance" need not be adversarial, but worst case formulation provides guarantees

Rational & Adversarial Disturbances

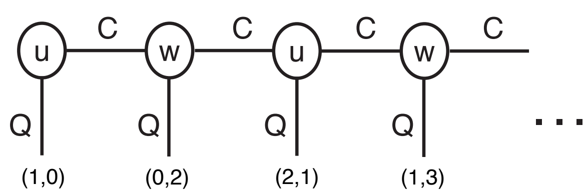
- Pursuit/Evasion:

$$p_{k+1} = A_p p_k + B_p u_k \quad \text{(pursuer)}$$
$$e_{k+1} = A_e e_k + B_e w_k \quad \text{(evader)}$$

  Objective:

$$\min_{\pi} \max_{e} |C(p_k - e_k)|$$

- Strategic games (chess, go, etc)
- Questioning rational models: Centipede game



  - C = Continue, Q = Quit
  - *Reward* to (u, w) is $(v, -v)$
  - Rational model of opponent forces u to immediately quit...even after observing multiple missteps!?

Value Iteration

- Setup:

$$x_{k+1} = f_k(x_k, u_k, w_k)$$

$$u_k \in U_k(x_k)$$

$$w_k \in W_k(x_k, u_k)$$

- Cost of policy:

$$J_\pi(x_0) = \max_{w_0, w_1, \ldots, w_{N-1}} g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)$$

- Optimization:

$$J^*(x_0) = \min_\pi J_\pi(x_0)$$

- Value Iteration:

$$J_N(x_N) = g_N(x_N)$$

$$J_k(x_k) = \min_{u_k \in U(x_k)} \max_{w_k \in W(x_k, u_k)} g(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))$$

  Note: Left to right = order of commitment

- Theorem:
  - $J^*(x_0) = J_0(x_0)$
  - $\mu_k^*(x_k) = \arg\min_{u_k \in U(x_k)} \max_{w_k \in W(x_k, u_k)} g(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))$

Minimax Lemmas

- FACT: (minimax inequality)

$$\min_{x \in X} \max_{y \in Y} G(x, y) \geq \max_{y \in Y} \min_{x \in X} G(x, y)$$

  Inspect: For any (x, y)

$$\max_{y} G(x, y) \geq G(x, y) \geq \min_{x} G(x, y)$$

  LHS depends only on x & RHS depends only on y

$$\min_{x} \mathrm{LHS}(x) \geq \max_{y} \mathrm{RHS}(y)$$

- FACT: (minimax exchange)

$$\min_{\mu(\cdot)} \max_{w} G(\mu(w), w) = \max_{w} \min_{u} G(u, w)$$

  Know

$$\min_{\mu} \max_{w} G(\mu(w), w) \geq \max_{w} \min_{\mu} G(\mu(w), w) = \max_{w} \min_{u} G(u, w)$$

  Use

$$\mu^*(w) = \arg \min_{u} G(u, w)$$

  to show equality

Proof of Value Iteration

- Special case: $N = 2$
- $J_2(x_2) = g_2(x_2)$
- $J_1(x_1) = \min_{u_1} \max_{w_1} g_1(x_1, u_1, w_1) + J_2(f_1(x_1, u_1, w_1))$
- $J_0^*(x_0) =$
$$\min_{\mu_0} \min_{\mu_1} \max_{w_0} \max_{w_1} G'(\mu_0, \mu_1, w_0, w_1; x_0)$$
- Define
$$G(\mu_0, \mu_1, w_0; x_0) = \max_{w_1} G'(\mu_0, \mu_1, w_0, w_1; x_0)$$
- Now faced with
$$\min_{\mu_0} \min_{\mu_1} \max_{w_0} G(\mu_0, \mu_1, w_0; x_0)$$
Note that for x_0 and μ_0 specified, x_1 is effectively a function of w_0
- Apply minimax exchange and restate G'
$$\min_{\mu_0} \max_{w_0} \min_{u_1} \max_{w_1} G'(\cdot)$$
- Expand definitions to show that
$$J_0^*(x_0) = J_0(x_0)$$

Example

- Scalar linear system:
$$x_{k+1} = x_k + u_k + w_k, \quad |w_k| \le 1$$
$$g(x,u,w) = |x| + (3/2)\,|u|$$
$$g_N(x) = |x|$$
$$N = 2$$
- $J_2(x) = |x|$
- J_1:
  - $J_1(x) = \min_u \max_{|w|\le 1} |x| + (3/2)\,|u| + |x+u+w|$
  - Worst case w aligned with $x+u$
  - Best case u at vertex, either $u=0$ or $u=-x$
  - Compare:
$$u = 0 : |x| + 0 + |x| + 1$$
$$u = -x : |x| + (3/2)\,|x| + 1$$
  - Therefore,
$$J_1(x) = 2\,|x| + 1 \quad \& \quad \mu_1^*(x_1) = 0$$
- J_0:
$$J_0(x) = \min_u \max_{|w|\le 1} |x| + (3/2)\,|u| + 2\,|x+u+w| + 1$$
$$u = 0 : J_0(x) = |x| + 0 + 2\,|x| + 2 + 1$$
$$u = -x : J_0(x) = |x| + (3/2)\,|x| + 2 + 1$$
- Therefore,
$$J_0(x) = (5/2)\,|x| + 3 \quad \& \quad \mu_0^*(x) = -x$$

Stochastic DP Preview

- Objective: Study systems with random phenomena.
- Example: Inventory control

$$x_{k+1} = x_k + u_k - w_k$$

  - $x = \begin{cases} \text{inventory} & x > 0 \\ \text{backlog} & x < 0 \end{cases}$
  - $u =$ production
  - $w =$ demand
  - Total cost:

$$R(x_N) + \sum_{k=0}^{N-1} r(x_k) + cu_k$$

  terminal cost + sum of stage cost & production cost

- How to model w?
  - Worst case:

$$\min_{\mu_0,\dots,\mu_{N-1}} \max_{w \in \mathcal{W}} \text{etc}$$

  - "Random":

$$\min_{\mu_0,\dots,\mu_{N-1}} E(\text{etc})$$

  - $E \stackrel{\text{def}}{=}$ expected value = average cost over lots of experiments
- Random formulation is a MODEL (not necessarily reality) that expresses unwillingness/futility of pursuing a more detailed model.