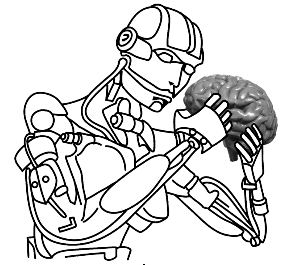
CS545—Contents XVII

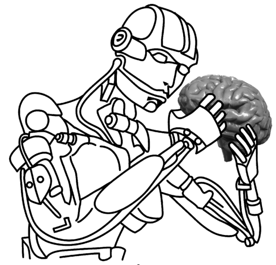
- Optimal Control
  - The optimal control framework
    (from the view of reinforcement learning)
  - Bellman's principle of optimality
  - Linear quadratic regulator control
- Reading Assignment for Next Class
  - See http://www-clmc.usc.edu/~cs545

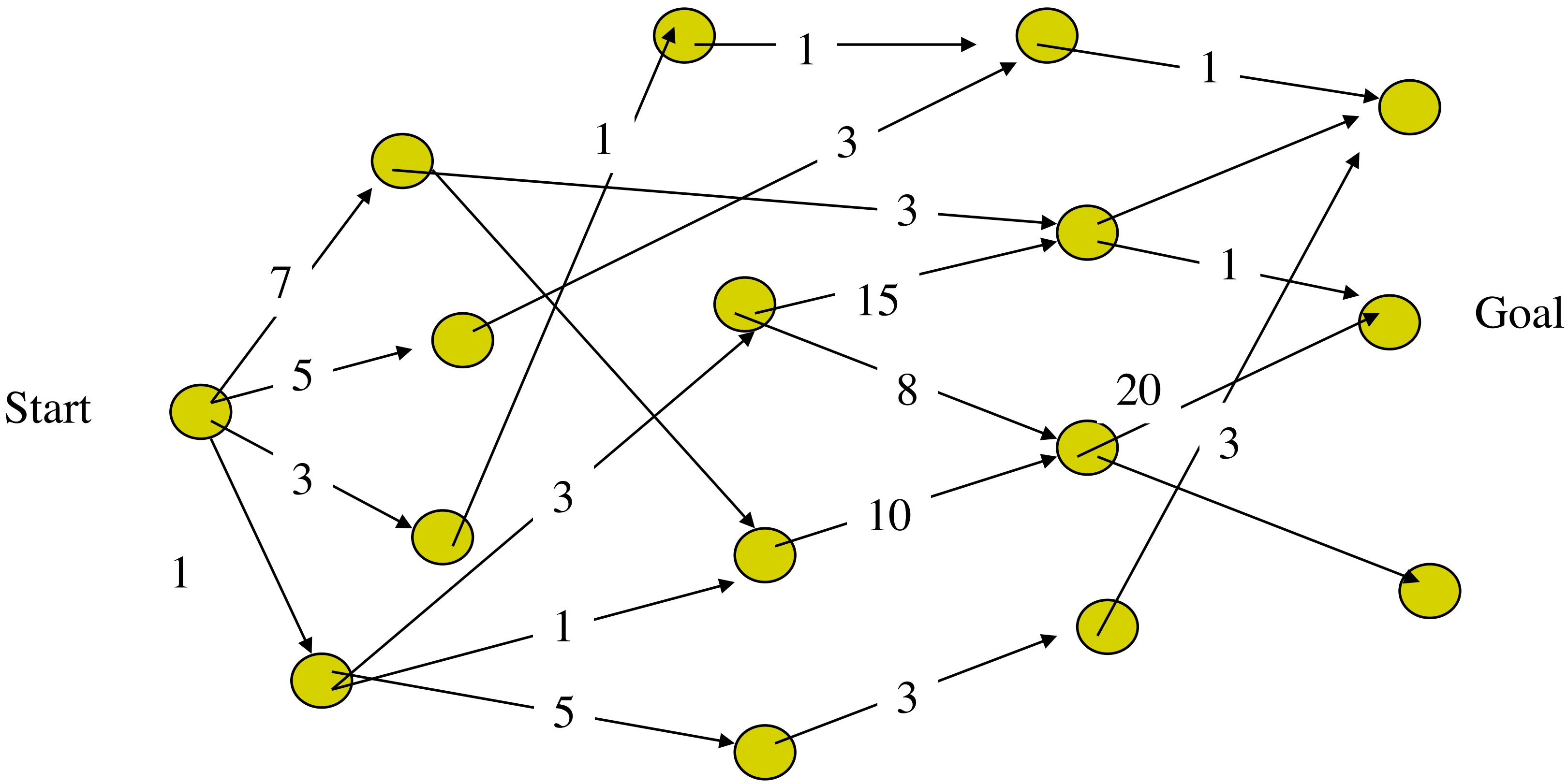
The Optimal Control Framework



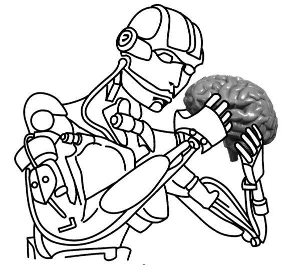
- Given:
  - A controlled dynamical system (discrete time notation, for simplicity) $$\mathbf{x}^{n+1} = f\left(\mathbf{x}^n, \mathbf{u}^n\right)$$
  - A cost function $$J = \phi\left(\mathbf{x}^N, \alpha\right) + \sum_{i=0}^{N-1} L\left(\mathbf{x}^i, \mathbf{u}^i, \alpha\right)$$
  - I.e., there is a cost attached to every action taken in every state
- Goal:
  - Find the sequence of commands that minimizes (maximized) the cost function

Bellman's Optimality Principle: An Example





Bellman's Principle of Optimality



"**An optimal sequence of controls in a multistage optimization problem has the property that whatever the initial stage, state and controls are, the remaining controls must constitute an optimal sequence of decisions for the remaining problem with stage and state resulting from previous controls considered as initial conditions.**"

- Formally this means:

$$J_i^*(\mathbf{x}_i) = \arg\min_{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)} \left\{ L(\mathbf{x}_i, \mathbf{u}_i, \alpha) + J_{i+1}^*(\mathbf{x}_{i+1}) \right\}$$

Possible Cost Functions

- Minimum Time Control (Bang-bang control)

$$L\left(\mathbf{x}^i,\mathbf{u}^i,\alpha\right)=1$$

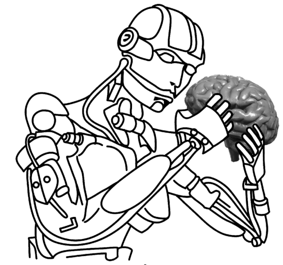
- Minimum Motor Command Control

$$L\left(\mathbf{x}^i,\mathbf{u}^i,\alpha\right)=\mathbf{u}^{iT}\mathbf{R}\mathbf{u}^i$$

- Quadratic Cost Criterion

$$L\left(\mathbf{x}^i,\mathbf{u}^i,\alpha\right)=\mathbf{x}^{iT}\mathbf{Q}\mathbf{x}^i+\mathbf{u}^{iT}\mathbf{R}\mathbf{u}^i$$

Linear Quadratic Regulator Control (LQR)



- Special Assumption: Linear System Dynamics

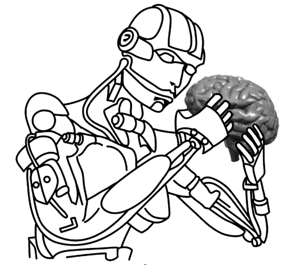
$$\mathbf{x}^{n+1} = \mathbf{A}\mathbf{x}^{n} + \mathbf{B}\mathbf{u}^{n}$$

- Quadratic cost function

$$L\left(\mathbf{x}^{i}, \mathbf{u}^{i}, \alpha\right) = \mathbf{x}^{iT}\mathbf{Q}\mathbf{x}^{i} + \mathbf{u}^{iT}\mathbf{R}\mathbf{u}^{i}$$

- Goal:
  - Bring the system to a setpoint and keep it there
  - Note: this can also be done with a nonlinear system by a local linearization

Deriving the LQR Control Law



- Start with Bellman's principle of optimality

$$J_i^*(\mathbf{x}_i) = \underset{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)}{\arg\min} \left\{ L(\mathbf{x}_i, \mathbf{u}_i, \alpha) + J_{i+1}^*(\mathbf{x}_{i+1}) \right\}$$

$$= \underset{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)}{\arg\min} \left\{ \mathbf{x}^{i^T} \mathbf{Q} \mathbf{x}^i + \mathbf{u}^{i^T} \mathbf{R} \mathbf{u}^i + J_{i+1}^*(\mathbf{x}_{i+1}) \right\}$$

$$= \underset{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)}{\arg\min} \left\{ \mathbf{x}^{i^T} \mathbf{Q} \mathbf{x}^i + \mathbf{u}^{i^T} \mathbf{R} \mathbf{u}^i + J_{i+1}^*(\mathbf{A}\mathbf{x}_i + \mathbf{B}\mathbf{u}^i) \right\}$$

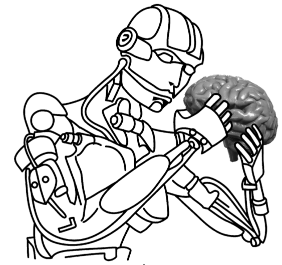
- Since we are going to design a linear control law

$$\mathbf{u}^{i^*} = -\mathbf{K}^i \mathbf{x}^i$$

- it is correct to write the optimal cost at stage i as a quadratic form

$$J^{i^*} = \mathbf{x}^{i^T} \mathbf{P}^i \mathbf{x}^i$$

Deriving the LQR Control Law (cont'd)



- thus

$$J_i^*(\mathbf{x}_i) = \underset{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)}{\arg\min} \left\{ \mathbf{x}^{iT} \mathbf{Q} \mathbf{x}^i + \mathbf{u}^{iT} \mathbf{R} \mathbf{u}^i + J_{i+1}^*\left(\mathbf{A}\mathbf{x}_i + \mathbf{B}\mathbf{u}^i\right) \right\}$$

$$= \underset{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)}{\arg\min} \left\{ \mathbf{x}^{iT} \mathbf{Q} \mathbf{x}^i + \mathbf{u}^{iT} \mathbf{R} \mathbf{u}^i + \left(\mathbf{A}\mathbf{x}^i + \mathbf{B}\mathbf{u}^i\right)^T \mathbf{P}^{i+1} \left(\mathbf{A}\mathbf{x}^i + \mathbf{B}\mathbf{u}^i\right) \right\}$$

- To optimize, we need the first derivatives to be zero:

$$\frac{\partial J^{i^*}}{\partial \mathbf{u}^i} = 0$$

- thus

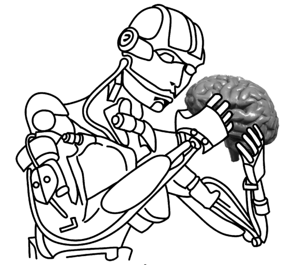
$$2\mathbf{u}^{iT}\mathbf{R} + 2\left(\mathbf{A}\mathbf{x}^i + \mathbf{B}\mathbf{u}^i\right)^T \mathbf{P}^{i+1}\mathbf{B} = 0$$

$$\mathbf{u}^{iT}\mathbf{R} + \mathbf{x}^{iT}\mathbf{A}^T\mathbf{P}^{i+1}\mathbf{B} + \mathbf{u}^{iT}\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{B} = 0$$

$$\mathbf{u}^i = -\left(\mathbf{R}^T + \mathbf{B}^T\mathbf{P}^{i+1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{A}\mathbf{x}^i$$

$$\mathbf{u}^i = -\mathbf{K}^i\mathbf{x}^i$$

Deriving the LQR Control Law (cont' d)



- What is the update of the matrix **P**?

$$J_i^*(\mathbf{x}_i) = \arg\min_{\mathbf{u}_i \in \mathbf{u}(\mathbf{x}_i)} \left\{ \mathbf{x}^{i^T}\mathbf{Q}\mathbf{x}^i + \mathbf{u}^{i^T}\mathbf{R}\mathbf{u}^i + \left(\mathbf{A}\mathbf{x}^i + \mathbf{B}\mathbf{u}^i\right)^T \mathbf{P}^{i+1}\left(\mathbf{A}\mathbf{x}^i + \mathbf{B}\mathbf{u}^i\right) \right\}$$

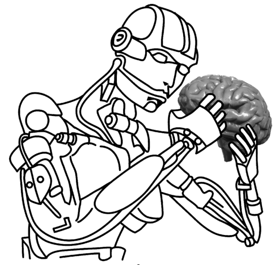
$$= \mathbf{x}^{i^T}\mathbf{Q}\mathbf{x}^i + \mathbf{x}^{i^T}\mathbf{K}^T\mathbf{R}\mathbf{K}\mathbf{x}^i + \left(\mathbf{A}\mathbf{x}^i - \mathbf{B}\mathbf{K}\mathbf{x}^i\right)^T \mathbf{P}^{i+1}\left(\mathbf{A}\mathbf{x}^i - \mathbf{B}\mathbf{K}\mathbf{x}^i\right)$$

$$= \mathbf{x}^{i^T}\left(\mathbf{Q} + \mathbf{K}^T\mathbf{R}\mathbf{K} + \mathbf{A}^T\mathbf{P}^{i+1}\mathbf{A} + \mathbf{K}^T\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{B}\mathbf{K} - \mathbf{A}^T\mathbf{P}^{i+1}\mathbf{B}\mathbf{K} - \mathbf{K}^T\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{A}\right)\mathbf{x}^i$$

$$\mathbf{P}^i = \mathbf{Q} + \mathbf{K}^T\mathbf{R}\mathbf{K} + \mathbf{A}^T\mathbf{P}^{i+1}\mathbf{A} + \mathbf{K}^T\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{B}\mathbf{K} - \mathbf{A}^T\mathbf{P}^{i+1}\mathbf{B}\mathbf{K} - \mathbf{K}^T\mathbf{B}^T\mathbf{P}^{i+1}\mathbf{A}$$

- How to use this:
  - Assume the cost matrix P at the last stage
  - Iterate backward to obtain control gains for all previous stages
- IMPORTANT:
  - **K** converges to a fixed value after some time backward: thus, it is possible to use just one fixed K for all time (assuming that we have a multistage optimization problem with large horizon)

For Continuous Time Systems



$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$$

$$J = h\left(\mathbf{x}\left(t_f\right)\right) + \int_0^{t_f} l\left(\mathbf{x}(t), \mathbf{u}(t), \alpha\right) dt$$

$$J^*(t) = \max_{\mathbf{u}} \left(l\left(\mathbf{x}(t), \mathbf{u}(t), \alpha\right) + \frac{\partial J^*}{\partial \mathbf{x}} \frac{\partial f(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} \right)$$