6.231 DYNAMIC PROGRAMMING

LECTURE 13

LECTURE OUTLINE

- Control of continuous-time Markov chains – Semi-Markov problems
- Problem formulation – Equivalence to discrete-time problems
- Discounted problems
- Average cost problems

CONTINUOUS-TIME MARKOV CHAINS

- Stationary system with finite number of states and controls
- State transitions occur at discrete times
- Control applied at these discrete times and stays constant between transitions
- Time between transitions is random
- Cost accumulates in continuous time (may also be incurred at the time of transition)
- Example: Admission control in a system with restricted capacity (e.g., a communication link)
  - Customer arrivals: a Poisson process
  - Customers entering the system, depart after exponentially distributed time
  - Upon arrival we must decide whether to admit or to block a customer
  - There is a cost for blocking a customer
  - For each customer that is in the system, there is a customer-dependent reward per unit time
  - Minimize time-discounted or average cost

PROBLEM FORMULATION

- $x(t)$ and $u(t)$: State and control at time t
- t_k: Time of kth transition ($t_0 = 0$)
- $x_k = x(t_k)$; $\quad x(t) = x_k$ for $t_k \leq t < t_{k+1}$.
- $u_k = u(t_k)$; $\quad u(t) = u_k$ for $t_k \leq t < t_{k+1}$.
- No transition probabilities; instead transition distributions (quantify the uncertainty about both transition time and next state)

$$Q_{ij}(\tau, u) = P\{t_{k+1} - t_k \leq \tau,\ x_{k+1} = j \mid x_k = i,\ u_k = u\}$$

- Two important formulas:

(1) Transition probabilities are specified by

$$p_{ij}(u) = P\{x_{k+1} = j \mid x_k = i,\ u_k = u\} = \lim_{\tau \to \infty} Q_{ij}(\tau, u)$$

(2) The Cumulative Distribution Function (CDF) of τ given i, j, u is (assuming $p_{ij}(u) > 0$)

$$P\{t_{k+1} - t_k \leq \tau \mid x_k = i,\ x_{k+1} = j,\ u_k = u\} = \frac{Q_{ij}(\tau, u)}{p_{ij}(u)}$$

Thus, $Q_{ij}(\tau, u)$ can be viewed as a "scaled CDF"

EXPONENTIAL TRANSITION DISTRIBUTIONS

- Important example of transition distributions:

$$Q_{ij}(\tau, u) = p_{ij}(u)\big(1 - e^{-\nu_i(u)\tau}\big),$$

where $p_{ij}(u)$ are transition probabilities, and $\nu_i(u)$ is called the transition rate at state i.

- Interpretation: If the system is in state i and control u is applied
  - the next state will be j with probability $p_{ij}(u)$
  - the time between the transition to state i and the transition to the next state j is exponentially distributed with parameter $\nu_i(u)$ (independently of j):

$$P\{\text{transition time interval } > \tau \mid i, u\} = e^{-\nu_i(u)\tau}$$

- The exponential distribution is memoryless. This implies that for a given policy, the system is a continuous-time Markov chain (the future depends on the past through the current state).

- Without the memoryless property, the Markov property holds only at the times of transition.

COST STRUCTURES

- There is cost $g(i,u)$ per unit time, i.e.

$$g(i,u)dt = \text{ the cost incurred in time } dt$$

- There may be an extra "instantaneous" cost $\hat{g}(i,u)$ at the time of a transition (let's ignore this for the moment)

- Total discounted cost of $\pi = \{\mu_0, \mu_1, \ldots\}$ starting from state i (with discount factor $\beta > 0$)

$$\lim_{N\to\infty} E\left\{ \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} e^{-\beta t} g\big(x_k, \mu_k(x_k)\big) dt \;\Big|\; x_0 = i \right\}$$

- Average cost per unit time

$$\lim_{N\to\infty} \frac{1}{E\{t_N\}} E\left\{ \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} g\big(x_k, \mu_k(x_k)\big) dt \;\Big|\; x_0 = i \right\}$$

- We will see that both problems have equivalent discrete-time versions.

DISCOUNTED CASE - COST CALCULATION

- For a policy $\pi = \{\mu_0, \mu_1, \ldots\}$, write

$$J_\pi(i) = E\{\text{1st transition cost}\} + E\{e^{-\beta\tau} J_{\pi_1}(j) \mid i, \mu_0(i)\}$$

where $E\{\text{1st transition cost}\} = E\left\{\int_0^\tau e^{-\beta t} g(i, \mu_0(i))dt\right\}$ and $J_{\pi_1}(j)$ is the cost-to-go of $\pi_1 = \{\mu_1, \mu_2, \ldots\}$

- We calculate the two costs in the RHS. The $E\{\text{1st transition cost}\}$, if u is applied at state i, is

$$G(i,u) = E_j\left\{E_\tau\{\text{1st transition cost} \mid j\}\right\}$$
$$= \sum_{j=1}^{n} p_{ij}(u) \int_0^\infty \left(\int_0^\tau e^{-\beta t} g(i,u)dt\right) \frac{dQ_{ij}(\tau,u)}{p_{ij}(u)}$$
$$= g(i,u) \sum_{j=1}^{n} \int_0^\infty \frac{1 - e^{-\beta\tau}}{\beta} dQ_{ij}(\tau,u)$$

- Thus the $E\{\text{1st transition cost}\}$ is

$$G\big(i, \mu_0(i)\big) = g\big(i, \mu_0(i)\big) \sum_{j=1}^{n} \int_0^\infty \frac{1 - e^{-\beta\tau}}{\beta} dQ_{ij}\big(\tau, \mu_0(i)\big)$$

(The summation term can be viewed as a "discounted length of the transition interval $t_1 - t_0$".)

COST CALCULATION (CONTINUED)

- Also the expected (discounted) cost from the next state j is

$$
\begin{aligned}
&E\big\{e^{-\beta\tau}J_{\pi_1}(j) \mid i,\mu_0(i)\big\} \\
&\quad = E_j\big\{E\{e^{-\beta\tau} \mid i,\mu_0(i),j\}J_{\pi_1}(j) \mid i,\mu_0(i)\big\} \\
&\quad = \sum_{j=1}^{n} p_{ij}\big(\mu_0(i)\big)\left(\int_0^\infty e^{-\beta\tau}\frac{dQ_{ij}(\tau,\mu_0(i))}{p_{ij}(\mu_0(i))}\right)J_{\pi_1}(j) \\
&\quad = \sum_{j=1}^{n} m_{ij}\big(\mu_0(i)\big)J_{\pi_1}(j)
\end{aligned}
$$

where $m_{ij}(u)$ is given by

$$
m_{ij}(u) = \int_0^\infty e^{-\beta\tau}dQ_{ij}(\tau,u)\ \left(< \int_0^\infty dQ_{ij}(\tau,u) = p_{ij}(u)\right)
$$

and can be viewed as the "effective discount factor" [the analog of $\alpha p_{ij}(u)$ in discrete-time case].

- So $J_\pi(i)$ can be written as

$$
J_\pi(i) = G\big(i,\mu_0(i)\big) + \sum_{j=1}^{n} m_{ij}\big(\mu_0(i)\big)J_{\pi_1}(j)
$$

i.e., the (continuous-time discounted) cost of 1st period, plus the (continuous-time discounted) cost-to-go from the next state.

COST CALCULATION (CONTINUED)

- Also the expected (discounted) cost from the next state j is

$$
\begin{aligned}
&E\Big\{e^{-\beta\tau} J_{\pi_1}(j) \mid i, \mu_0(i)\Big\} \\
&\quad = E_j\Big\{E\{e^{-\beta\tau} \mid i, \mu_0(i), j\} J_{\pi_1}(j) \mid i, \mu_0(i)\Big\} \\
&\quad = \sum_{j=1}^{n} p_{ij}(\mu_0(i)) \left(\int_0^\infty e^{-\beta\tau} \frac{dQ_{ij}(\tau, \mu_0(i))}{p_{ij}(\mu_0(i))}\right) J_{\pi_1}(j) \\
&\quad = \sum_{j=1}^{n} m_{ij}\big(\mu_0(i)\big) J_{\pi_1}(j)
\end{aligned}
$$

where $m_{ij}(u)$ is given by

$$
m_{ij}(u) = \int_0^\infty e^{-\beta\tau} dQ_{ij}(\tau, u) \ \left(< \int_0^\infty dQ_{ij}(\tau, u) = p_{ij}(u)\right)
$$

and can be viewed as the "effective discount factor" [the analog of $\alpha p_{ij}(u)$ in discrete-time case].

- So $J_\pi(i)$ can be written as

$$
J_\pi(i) = G\big(i, \mu_0(i)\big) + \sum_{j=1}^{n} m_{ij}\big(\mu_0(i)\big) J_{\pi_1}(j)
$$

i.e., the (continuous-time discounted) cost of 1st period, plus the (continuous-time discounted) cost-to-go from the next state.

EQUIVALENCE TO AN SSP

- Similar to the discrete-time case, introduce an "equivalent" stochastic shortest path problem with an artificial termination state t

- Under control u, from state i the system moves to state j with probability $m_{ij}(u)$ and to the termination state t with probability $1 - \sum_{j=1}^{n} m_{ij}(u)$

- Bellman's equation: For $i = 1, \ldots, n$,

$$J^*(i) = \min_{u \in U(i)} \left[G(i,u) + \sum_{j=1}^{n} m_{ij}(u) J^*(j) \right]$$

- Analogs of value iteration, policy iteration, and linear programming.

- If in addition to the cost per unit time g, there is an extra (instantaneous) one-stage cost $\hat{g}(i,u)$, Bellman's equation becomes

$$J^*(i) = \min_{u \in U(i)} \left[\hat{g}(i,u) + G(i,u) + \sum_{j=1}^{n} m_{ij}(u) J^*(j) \right]$$

MANUFACTURER'S EXAMPLE REVISITED

- A manufacturer receives orders with interarrival times uniformly distributed in $[0, \tau_{\max}]$.
- He may process all unfilled orders at cost $K > 0$, or process none. The cost per unit time of an unfilled order is c. Max number of unfilled orders is n.
- The nonzero transition distributions are

$$Q_{i1}(\tau, \text{Fill}) = Q_{i(i+1)}(\tau, \text{Not Fill}) = \min\left[1, \frac{\tau}{\tau_{\max}}\right]$$

- The one-stage expected cost G is

$$G(i, \text{Fill}) = 0, \qquad G(i, \text{Not Fill}) = \gamma\, c\, i,$$

where

$$\gamma = \sum_{j=1}^{n} \int_0^\infty \frac{1 - e^{-\beta\tau}}{\beta} dQ_{ij}(\tau, u) = \int_0^{\tau_{\max}} \frac{1 - e^{-\beta\tau}}{\beta\tau_{\max}} d\tau$$

- There is an "instantaneous" cost

$$\hat{g}(i, \text{Fill}) = K, \qquad \hat{g}(i, \text{Not Fill}) = 0$$

MANUFACTURER'S EXAMPLE CONTINUED

- The "effective discount factors" $m_{ij}(u)$ in Bellman's Equation are

$$m_{i1}(\text{Fill}) = m_{i(i+1)}(\text{Not Fill}) = \alpha,$$

where

$$\alpha = \int_0^\infty e^{-\beta\tau} dQ_{ij}(\tau, u) = \int_0^{\tau_{\max}} \frac{e^{-\beta\tau}}{\tau_{\max}} d\tau = \frac{1 - e^{-\beta\tau_{\max}}}{\beta\tau_{\max}}$$

- Bellman's equation has the form

$$J^*(i) = \min\big[K + \alpha J^*(1),\ \gamma c i + \alpha J^*(i+1)\big], \quad i = 1, 2, \ldots$$

- As in the discrete-time case, we can conclude that there exists an optimal threshold i^*:

fill the orders <==> their number i exceeds i^*

AVERAGE COST

- Minimize $\lim_{N\to\infty} \frac{1}{E\{t_N\}} E\left\{\int_0^{t_N} g\big(x(t),u(t)\big)dt\right\}$ assuming there is a special state that is "recurrent under all policies"

- Total expected cost of a transition

$$G(i,u) = g(i,u)\overline{\tau}_i(u),$$

where $\overline{\tau}_i(u)$: Expected transition time.

- We apply the SSP argument used for the discrete-time case.
  - Divide trajectory into cycles marked by successive visits to n.
  - The cost at (i,u) is $G(i,u) - \lambda^* \overline{\tau}_i(u)$, where λ^* is the optimal expected cost per unit time.
  - Each cycle is viewed as a state trajectory of a corresponding SSP problem with the termination state being essentially n.

- So Bellman's Eq. for the average cost problem:

$$h^*(i) = \min_{u\in U(i)} \left[G(i,u) - \lambda^* \overline{\tau}_i(u) + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right]$$

MANUFACTURER EXAMPLE/AVERAGE COST

- The expected transition times are

$$\overline{\tau}_i(\text{Fill}) = \overline{\tau}_i(\text{Not Fill}) = \frac{\tau_{\max}}{2}$$

the expected transition cost is

$$G(i, \text{Fill}) = 0, \qquad G(i, \text{Not Fill}) = \frac{c\, i\, \tau_{\max}}{2}$$

and there is also the “instantaneous” cost

$$\hat{g}(i, \text{Fill}) = K, \qquad \hat{g}(i, \text{Not Fill}) = 0$$

- Bellman’s equation:

$$h^*(i) = \min\Big[K - \lambda^* \frac{\tau_{\max}}{2} + h^*(1), \\ ci \frac{\tau_{\max}}{2} - \lambda^* \frac{\tau_{\max}}{2} + h^*(i+1)\Big]$$

- Again it can be shown that a threshold policy is optimal.

6.231 Dynamic Programming and Stochastic Control
Fall 2015