B9120 Dynamic Programming Lecture 4 - 02/17/2020

Discounted Problems in Discrete and Continuous Time

Lecturer: Daniel Russo *Scribe: Cecilia Li, Mike Li*

Outline

- Proof of main theorem from last week
- Uniformization
- Preview of next week

Last Time

- We will think of the Bellman Operators mapping, which map a cost-to-go function to another cost-to-go function. More formally, the Bellman operator $T : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{X})$ is mapping from the space of bounded functionals $\mathcal{B}(\mathcal{X}) = \{J : \mathcal{X} \to \mathbb{R} \mid \|J\|_\infty < \infty\}$ to itself. The Bellman operator (and it's counterpart corresponding to policy μ) are defined as:

$$TJ(x) = \min_{u \in U(x)} \mathbb{E}_w[g(x,u,w) + \alpha J(f(x,u,w))]$$
$$T_\mu J(x) = \mathbb{E}[g(x,\mu(x),w) + \alpha J(f(x,\mu(x),w))]$$

- The DP Algorithms for N-period problems with terminal cost $J(x_N)$ can be expressed concisely as:

$$J_N^\star = TJ_{N-1}^\star = \cdots = T^N J$$
$$J_N^\mu = T_\mu J_{N-1}^\mu = \cdots = T_\mu^N J$$

- The next theorem shows that there exists a unique $J^* = \lim_{N\to\infty} T^N J$ and $J^\mu = \lim_{N\to\infty} T_\mu^N J$, where the limits are taken with respect to $||J||_\infty = \sup_x |J(x)|$.

1 Main Theorem

Recall the main theorem discussed at the end of the last lecture. Recall the notation $G(J) = \{\mu : T_\mu J = TJ\}$ is the set of stationary policies that "greedy" with respect to J. The policies attain the argmin over controls u in the definition of the Bellman operator, so they are derived be performing a one-step-lookahead with respect to J.

Theorem 1. *1. There exists a unique optimal cost function $J^\star$ that satisfies:*

$$J^\star = TJ^\star$$

and for all J,

$$||T^N J - J^\star||_\infty \le \alpha^N ||J - J^\star||_\infty$$

2. For every stationary policy μ, there exists a unique associated cost function $J_\mu : S \to R$ that solves the fixed point equation:

$$J_\mu = T_\mu J_\mu$$

and for all J,

$$||T_\mu^N J - J_\mu||_\infty \le \alpha^N ||J - J_\mu||_\infty$$

3. If $\mu \in G(J^\star)$, then $J_\mu(x) = J^\star(x)$ for all $x \in \mathcal{X}$. For any (possibly non-stationary) policy π,

$$\liminf_{N\to\infty} \mathbb{E}^\pi \left[\sum_{k=0}^{N-1} \alpha^k g_k(x_K, u_k, w_k) \middle| x_0 = x \right] \ge J^\star(x)$$

1.1 Properties of Bellman Operators

We prove some generic properties of the Bellman operator which will be useful throughout. Theorem 1 also follows as a simple consequence of these properties.

1. **Monotonicity:** If $J(x) \leq J'(x)\ \forall\, x \in \mathcal{X}$, then $TJ(x) \leq TJ'(x)\ \forall\, x \in \mathcal{X}$. We can also write $J \precsim J' \Rightarrow TJ \precsim TJ'$.
2. **Constant Shift:** $T(J + ce) = TJ + \alpha ce$ where we take $e(x) = 1\ \forall\, x \in \mathcal{X}$ to denote a vector of all ones.
3. **Contraction:** $||TJ - TJ'||_\infty \leq \alpha ||J - J'||_\infty\ \ \forall\, J, J'$. In words, applying T to two cost-to-go functions brings them geometrically closer.

The exact same statements also hold for T_μ. The monotonicity and constant shift follow by inspecting the definition of the Bellman operators. We prove that these imply the contraction property.

Proof of Contraction. Let $||\cdot||$ denote the infinity-norm $||\cdot||_\infty$. Consider:

$$J' - ||J - J'||e \precsim J \precsim J' + ||J - J'||e$$

Applying monotonicity gives:

$$T(J' - ||J - J'||e) \precsim TJ \precsim T(J' + ||J - J'||e)$$

Applying constant shift gives:

$$TJ' - \alpha||J - J'||e \precsim TJ \precsim TJ' + \alpha||J - J'||e$$

Subtracting TJ' yelds our desired result. □

1.2 Contraction Mapping Theorem

We will prove this for any contraction operator $F : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{X})$, where $\mathcal{B}(\mathcal{X}) = \{J : \mathcal{X} \to R : ||J||_\infty < \infty\}$. An essentially identical proof to general complete metric spaces (a Cauchy sequence converges to a limit in a complete metric space).

Theorem 2. *If $||FJ - FJ'|| \leq \alpha||J - J'||$ for all $J, J' \in \mathcal{B}(\mathcal{X})$, then:*

1. *There exists a unique $J^* \in \mathcal{B}(\mathcal{X})$ such that $FJ^* = J^*$.*
2. *$\forall J \in \mathcal{B}(\mathcal{X}),\ ||F^N J - J^*|| \leq \alpha^N ||J - J'||$.*

Proof. We first prove that $\{J_k\}$ is Cauchy, and by the completeness of the space the limit J_∞ exists. For fixed J, let $J_0 = J, J_k = FJ_{k-1}$ for $k = 1, 2, \cdots$. We have:

$$\begin{aligned} ||J_{k+1} - J_k|| &= ||FJ_k - FJ_{k-1}|| \\ &\leq \alpha||J_k - J_{k-1}|| \quad \text{(contraction operator)} \\ &\leq \cdots \leq \alpha^k||J_1 - J_0|| \quad \text{(induction)} \end{aligned}$$

For all $m \geq 1$, we have:

$$\begin{aligned} ||J_{k+m} - J_k|| &\leq \sum_{l=1}^{m} ||J_{k+l} - J_{k+l-1}|| \quad \text{(triangle inequality)} \\ &\leq \alpha^k \sum_{l=0}^{m-1} \alpha^l ||J_1 - J_0|| \\ &\leq \alpha^k \frac{1}{1-\alpha} ||J_1 - J_0|| \to 0 \text{ as } k \to \infty \end{aligned}$$

Therefore, $\{J_k\}$ is Cauchy and by the completeness of the space, $J_\infty = \lim_{N\to\infty} F^N J$ exists.

- We now show the existence of a fixed point. The natural candidate for this fixed point is J_∞. We have:

$$\begin{aligned} 0 &\leq ||FJ_\infty - J_\infty|| \\ &\leq ||FJ_\infty - J_k|| + ||J_k - J_\infty|| \quad \text{(separation and triangle inequality)} \\ &\leq \alpha||J_\infty - J_{k-1}|| + ||J_k - J_\infty|| \to 0 \text{ as } k \to \infty \end{aligned}$$

  We can also show existence by using the fact that contractivity gives continuity; $F(\lim_{k\to\infty} J_k) = \lim_{k\to\infty} FJ_k = J^*$.

- Now we look at the geometric convergence rate. Recall that $FJ_\infty = J_\infty$. Then,

$$||J_k - J_\infty|| = ||F^k J - F^k J_\infty|| \leq \alpha^k ||J_0 - J_\infty||$$

- For the last step, we look at uniqueness. Suppose that $J = FJ$ and $J' = FJ'$ are two fixed points. In order to show that $J = J'$, we can equivalently consider the distance between them:

$$||J - J'|| = ||FJ - FJ'|| \leq \alpha||J - J'||$$

  Therefore, $||J - J'|| = 0$, and $J = J'$.

□

Note: Some simple functions don't have a fixed point, consider $f(x) = e^x$. Intuitively, the issue is that e^x magnifies differences, whereas contraction mappings "dampen" them.

1.3 Back to our Main Theorem

We now have the necessary tools to prove the main theorem from last week.

Proof. The unique existence of J^* and the geometric convergence of $T^N J$ to it are both guaranteed by the Contraction Mapping Theorem. It remains to prove item 3.

Suppose that $\mu \in G(J^*)$. By definition, $T_\mu J^* = TJ^* = J^*$. Note that J^* is the unique fixed point of T_μ, so $J^\mu = J^*$.

Warmup: Optimality among stationary policies: Show that $J^* \precsim J^\mu$ for all μ. We observe that

$$J_\mu = T_\mu J_\mu \succsim TJ_\mu.$$

Applying T to both sides and recognising monotonicity gives $TJ_\mu \succsim T^2 J_\mu$. Repeating this inductively gives

$$J_\mu \succsim TJ_\mu \succsim T^2 J_\mu \succsim \cdots \succsim T^k J_\mu \succsim \cdots \succsim J^*.$$

*O*ptimality among non-stationary policies: $M = ||g||_\infty$. For any non-stationary policy $\pi = (\mu_0, \mu_1, \cdots)$,

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}^\pi \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k) \middle| x_0 = x \right] &\geq \lim_{N\to\infty} \mathbb{E}^\pi \left[\sum_{k=0}^{m-1} \alpha^k g(x_k, u_k, w_k) \middle| x_0 = x \right] - \sum_{k=m}^{N-1} \alpha^k M \\ &\geq \inf_{\pi'} \mathbb{E}^{\pi'} \left[\sum_{k=0}^{m-1} \alpha^k g(x_k, u_k, w_k) \middle| x_0 = x \right] - \sum_{k=m}^{\infty} \alpha^k M \\ &= (T^m \vec{0})(x) - \frac{\alpha^m}{1-\alpha} M \\ &\to J^*(x) \text{ as } m \to \infty. \end{aligned}$$

The second inequality uses that $T^m \vec{0}$ is the optimal cost-to-go function *among possibly non-stationary policies* for an m period problem with zero costs incurred at the terminal sate. □

Remark: Note that the error term above is uniform over all policies. In this way, we can argue that any near-optimal policy for a long but finite horizon problem is near-optimal for an infinite horizon problem, and vice-versa.

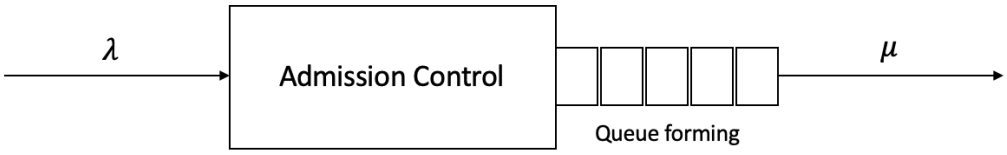
2 Discounted Continuous Time Problems

Some of the techniques we have seen so far also apply to continuous time problems, where decisions are still made a discrete but possible random time instances. The results go through for general finite state problems. For countable state problems, we need some regularity conditions ensuring that the time between decision epochs does not become infinitesimally small.

2.1 Prototypical Example: Queuing Admission Control Problem

We begin with a motivating example. Consider a queuing admission control problem with the following characteristics:

1. Single server queue.
2. Poisson arrivals with rate parameter λ.
3. Service times follow an exponential distribution with rate parameter μ.
4. Control decision (binary): whether to admit or reject.



While this problem is formulated in continuous time, decisions are made at the discrete (random) times when there is a change in the system state. There are two types of costs:

1. cost c per customer in queue per time instant.
2. cost D per job dropped (cost for turning people away).

Problem Formulation

1. Decisions are made at discrete (random) times $t_1 < t_2 < ... < t_k < ....$
   - It will also be useful to introduce the following notation: denote $\tau_k = t_k - t_{k-1}$ for $k = \{1, 2, \ldots\}$
2. At time t_k, the decision maker observes the state $i(t_k)$ and selects a (feasible) control $u(t_k) \in U(i(t_k))$ which is applied throughout $[t_k, t_{k+1})$.
3. $(\tau_{k+1} \,|\, i(t_k) = i, u(t_k) = u) \sim \exp(\nu_i(u))$, where $\nu_i(u) \leq \nu \; \forall \, (i, u)$.
   - We can consider ν to be a uniform control on rate at which events occur, for example we can consider $\nu = \max_{i,u} \nu_i(u)$.
4. The next state is j with probability $p_{ij}(u)$, independent of τ_k.

Objective: Minimize the expected cost

$$\limsup_{T\to\infty} \mathbb{E}^\pi \left[\int_{t=0}^{T} e^{-\beta t} g(i(t),u(t))dt\right] = \limsup_{N\to\infty} \mathbb{E}^\pi \left[\int_{t=0}^{t_N} e^{-\beta t} g(i(t),u(t))dt\right]$$
$$= \limsup_{N\to\infty} \sum_{k=0}^{N} \mathbb{E}^\pi \left[\int_{t=t_k}^{t_{k+1}} e^{-\beta t} g(i_k,u_k)dt\right]$$

where $i_k = i(t)$ for $t \in [t_k, t_{k+1})$ and $u_k = u(t)$ for $t \in [t_k, t_{k+1})$. Note, here we denote π to be an admissible policy, i.e. $u(t_k) \in U(i(t_k))$ and the expectation is taken with respect to the state transitions as well as the random transition times.

2.2 Warm-up Case: $\nu_i(u) = \nu,\ \forall\,(i,u)$

In the simple case when the time between decision epochs is distributed i.i.d. the objective becomes:

$$\limsup_{N\to\infty} \sum_{k=0}^{N} \mathbb{E}\left[\int_{t_k}^{t_{k+1}} e^{-\beta t} dt\right] \mathbb{E}[g(i_k,u_k)] = \limsup_{N\to\infty} \frac{1}{\beta+\nu} \sum_{k=0}^{N} \alpha^k \mathbb{E}[g(i_k,u_k)]$$

where $\alpha = \mathbb{E}[e^{-\beta\tau}] = \int_0^\infty e^{-\beta t}\nu e^{-\nu t}dt = \frac{\nu}{\beta+\nu}$. This follows from the calculation,

$$\mathbb{E}\left[\int_{t_k}^{t_{k+1}} e^{-\beta t dt}\right] = \mathbb{E}\left[e^{-\beta t_k}\int_0^{\tau_{k+1}} e^{-\beta t}dt\right] = \mathbb{E}\left[e^{-(\beta\tau_1+\ldots+\beta\tau_k)}\right]\frac{1}{\beta}(1-\mathbb{E}[e^{-\beta\tau}]) = \frac{1}{\beta+\nu}\prod_{i=1}^{k}\mathbb{E}[e^{-\beta\tau_i}] = \alpha^k,$$

where we use the fact that the random variables $\{\tau_k\}$ are distributed i.i.d with rate ν. Note how this mimics the infinite horizon discounted objective which we are now familiar with.

2.3 Non-Uniform Transition Rates: $\nu_i(u) \le \nu$

The main idea for the non-uniform case is to reduce to the uniform case by introducing fictitious events (where the event occurs but nothing happens, can be viewed as a transition to self-state). This can be viewed as an analogue of *thinning*, a commonly used technique in the study of Poisson processes. Like before, we assume the inter-event times to be i.i.d $\tau_1,\ldots,\tau_k \sim \exp(\nu)$, but to account for non-uniform transition rates we introduce some "fictitious" events where the system state does not change (that is the system transitions to the same state).

New Transition Probabilities under Uniform Version

$$\widetilde{p_{ij}}(u) = \begin{cases} \frac{\nu_i(u)}{\nu} p_{ij}(u) & \text{if } i \neq j \\ \frac{\nu_i(u)}{\nu} p_{ii}(u) + \frac{\nu-\nu_i(u)}{\nu} & \text{if } i = j \end{cases}$$

The interpretation is that we have added a fictitious events that occur at rate $\frac{\nu-\nu_i(u)}{\nu}$. A fraction $\nu_i(u)\nu$ of events are "genuine" and drawn according to the probabilities $p_{ij}(u)$.

Remark: by adding fictitious events, we add new periods at the control can be re-chosen. But since the system state does not actually change the original control would still be optimal. This last fact uses crucially that the time between events is exponentially distributed (and hence memory-less).

Bellman Operator

$$TJ(i) = \min_{u\in U(i)} \widetilde{g}(i,u) + \alpha \sum_j \widetilde{p_{ij}}(u) J(j)$$

where $\alpha = \frac{v}{\beta+v}$, $\widetilde{g}(i,u) = \frac{g(i,u)}{\beta+v}$. Again, we are essentially back in the infinite horizon discounted regime with modified cost and transition probabilities. Analyzing both the above cases heavily relied on the memoryless property given that the inter-event times were exponentially distributed, essentially $\{\tau_k\}$ were not dependent on the history. We relax this below.

2.4 General Semi-Markov Problems

We now consider the more general class of Semi-Markov problems, in which the time between events may not be exponentially distributed. We will see that again, most of the core theory we have developed for discrete time discounted dynamic programming applies to these problems.

Now, conditioned on the current sate i and control u, the next state and the time for a transition to occur are drawn jointly from some distribution Q_{iu}:

$$P(\tau_k \leq \tau, i_{k+1} = j | i_k = i, u_k = u) = Q_{iu}(\tau, j).$$

In addition, we assume that, conditioned on i_k and u_k, the next outcome (τ_k, i_{k+1}) is independent of the history $\{(t_\ell, u_\ell, i_\ell)\}_{\ell<k}$ prior to time k.

For such problems, we define a new Bellman operator as:

$$TJ(i) = \min_{u \in U(i)} \mathbb{E}\left[\int_{t_0}^{t_1} e^{-\beta t} g(i(t), u(t))dt + e^{-\beta t_1} J(i(t_1)) \middle| i(t_0) = i, u(t_0) = u\right],$$

where $t_0 = 0$ and the expectation integrates over the event time t_1 and the next state $i(t_1)$. Applying the tower property of conditional expectations, we can rewrite this as:

$$TJ(i) = \min_u \left[\widetilde{g}(i,u) + \sum_j p_{ij}(u)\alpha_{ij}(u)J(j)\right]$$

where $\alpha_{ij}(u) = \mathbb{E}[e^{-\beta t_1} \mid i(t_1) = j, i(t_0) = i, u(t_0) = u]$ acts as a discount factor, down-weighting how much future events matter. The function $\widetilde{g}(i,u) = \mathbb{E}\left[\int_{t_0}^{t_1} e^{-\beta t} g(i(t), u(t))dt \mid i(t_0) = i, u(t_0) = u\right]$ captures the expected discounted flow costs incurred by applying control u in state i until the next event occurs.

While this Bellman operator differs from the discounted case, it is still a monotone contraction operator obeying a variant of the constant-shift property. As a result, our main theoretical results extend to Semi-Markov problems. Precisely, suppose the effective discount factors are uniformly bounded above as $\alpha_{ij}(u) \leq \bar{\alpha} < 1 \ \forall (i,j,u)$. This is always true if the state space and action space is finite, assuming that $\mathbb{P}(\tau_k = 0) = 0$. Then T satisfies the following properties.

- Monotonicity: $J \precsim J' \Rightarrow TJ \precsim TJ'$.
- Strictly sub-constant shift: For $c \geq 0$, $T(J + ce) \precsim TJ + \overline{\alpha}ce$ and $T(J - ce) \succsim TJ - \overline{\alpha}ce$.
- Contraction: $\|TJ - TJ'\|_\infty \leq \overline{\alpha}\|TJ - TJ'\|_\infty$.

The contraction property follows from the other two, which can be seen by repeating the proof we gave in the discounted case.