Course Notes On Dynamic Optimization (Fall 2023)
Lecture 4: Infinite horizon discounted objectives

Instructor: Daniel Russo

Email: djr2174@gsb.columbia.edu

Graduate Instructor: David Cheikhi
Email: d.cheikhi@columbia.edu

These notes are based of scribed notes from a previous edition of the class. I have done some follow up light editing, but there may be typos or errors.

This class introduces

1 Finite horizon discounted objectives

Here we specify our problem where the cost function only depends on the time through the discounted factor a^k, while the new state x_{k+1} now is decided by f instead of f_k, that is:

$$g_k(x_k, u_k, w_k) = \alpha^k g(x_k, u_k, w_k) \quad \alpha \in (0,1)$$

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

where $\{w_k\}$ are now iid instead of simply independent, which means they do not depends on time as well. The problem

$$\inf_{\pi} \mathbb{E}[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k)] \tag{1}$$

is nevertheless solved by a nonstationary policies $\pi^* = (\mu_0^*, \ldots, \mu_{N-1}^*)$.

The DP algorithm solves for the optimal cost-to-go functions:

$$\begin{aligned} J_N(x) &= \alpha^N J(x) \\ J_{N-k}(x) &= \min_u \mathbb{E}[\alpha^{N-k} g(x,u,w) + J_{N-k+1}(f(x,u,w))], \end{aligned}$$

Rather than write it in this form, we can write $\tilde{J_k}^\star(x) = \frac{J^\star_{N-k}(x)}{\alpha^{N-k}}$ and the DP algorithm becomes:

$$\begin{aligned} \tilde{J_0}(x) &= J(x) \\ \tilde{J_k}(x) &= \min_{u \in U(x)} \mathbb{E}[g(x,u,w) + \alpha \tilde{J}_{k-1}(f(x,u,w))] \end{aligned}$$

From now on, we'll drop the $\tilde{}$.

Random horizon interpretation Introduce the random variable $\tau \sim \text{Geometric}(1-\alpha)$, assumed to be independent of all else. Then, since $\mathbb{P}(\tau \geqslant k) = \alpha^k$,

$$\mathbb{E}^\pi \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k) \right] = \mathbb{E}^\pi \left[\sum_{k=0}^{(N-1)\wedge\tau} g(x_k, u_k, w_k) \right],$$

where $\tau \wedge (N-1) = \min\{\tau, N-1\}$. Discounting can be understood as a reflection of horizon uncertainty.

The source of nonstationarity We've seen that, despite i.i.d disturbances, stationary costs and transitions, a nonstationary policy is generally optimal for the problem (1). Why? the main factor distinguishing the sub problems beginning in period k from those beginning at some later period $k' > k$ (at the same state) is the number of periods remaining. As k nears N, the decision-maker becomes increasingly myopic. This may be desirable feature of the problem if the end of the horizon N is a true constraint. In other settings, we want the decision-maker to prioritize the near-term without imposing the definitive end-time N. Infinite horizon discounted objectives are an elegant way to formalize this goal.

2 Infinite horizon discounted objectives

Consider now the infinite horizon analogue of (1), where the goal is to solve

$$\inf_\pi \limsup_{N\to\infty} \quad \mathbb{E}^\pi [\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k)]. \tag{2}$$

Don't pay much attention to the difference between 'limsup' and 'lim' here. We'll impose conditions under which limits exist. Under such regularity conditions, the random horizon interpretation extends to this case and one can understand the objective to be $\inf^\pi \mathbb{E}^\pi[\sum_{k=0}^{\tau} g(x_k, u_k, w_k)]$.

Proposition (Informal)**.** *Under appropriate regularity conditions,* (2) *admits an optimal policy* $\pi^* = (\mu_0^*, \mu_1^*, \ldots)$ *which is stationary. That is* $\mu_1^* = \mu_2^* = \mu_3^* = \cdots$.

Intuitive rationale. Imagine N is enormous, but finite. For k that is much much smaller than N, we expect that $J_k(x) \approx J_0(x)$; The difference between having N periods remaining and $N-k$ periods remaining is effectively irrelevant, discount factor downweights the far away future *at an expontential rate*. Hence, a policy μ_0^* with $\mu^*(x) \in \arg\min_u \mathbb{E}[g(x,u,w) + \alpha J_1(f(x,u,w))]$ will also nearly solve $\min_u \mathbb{E}[g(x,u,w) + \alpha J_k(f(x,u,w))]$ for k that are very small relative to N. □

2.1 Technical assumptions

- The state space is finite or countable.

  - This assumption sidesteps subtle measure theoretic issues that arise in dynamic programming problems with general state space. For many specific models, like linear quadratic control, these issues clearly do not arise. But when developing our generic theory, our state spaces are defined over the rationals rather than the reals.

- The control space is finite.

  - – This is used only to ensure that all minima are attained. Even when minima are not attained, most of these arguments carry through in terms of infima (i.e. sequences of policies whose performances converges to the infimum in (2)).

- The cost functions are uniformly bounded, i.e. $\sup_{(x,u,w)} |g(x,u,w)| \leqslant M < \infty$.

  - Typically, cost functions are written as some function $g(x_k, u_k, x_{k+1})$ of the state, action and next state. In this case, the above assumption is satisfied when the state space is finite or when the state space is compact and g is continuous. For problems where the assumption is violated, the proofs given below will not work because they are based on the max-norm of cost-to-go functions, which would be infinite. Arguments are then based on other weighted max-norms.

2.2 Bellman operators

Each iteration of the DP algorithm can be viewed as an operation that takes a function J_{k-1} and gets a new function J_k, we introduce the bellman operators. For bounded $J : \mathcal{X} \to R$, we define $(TJ) : \mathcal{X} \to R$ by:

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

Suppose our goal is to evaluate

$$J_n^*(x) = \mathbb{E}^{\mu}[\sum_{k=n}^{N-1} \alpha^k g(x_k, \mu(x_k), w_k) + \alpha^N J(x_N) | x_n = x]$$

The DP algorithm from previous classes tell us how. WARNING. Here we have indexed time backward. J_k^* is the optimal cost-to-go function for a problem with k periods remaining, whereas previously it referred to a problem with $N - K$ periods remaining.

The DP algorithm can be written concisely in terms of Bellman operators as

$$\begin{aligned} J_0^\star &= J \\ J_1^\star &= TJ \\ &\cdots \\ J_N^\star &= TJ_{N-1}^\star = \cdots = T^N J \end{aligned}$$

Bellman operator for a stationary policy For a fixed policy $\mu : \mathcal{X} \to \cup_x U(x)$ where $\mu(x) \in U(x)$, we define:

$$(T_\mu J)(x) = \mathbb{E}[g(x, \mu(x), w) + \alpha J(f(x, \mu(x), w))]$$

Suppose our goal is to evaluate

$$J_n^\mu(x) = \mathbb{E}^\mu[\sum_{k=n}^{N-1} \alpha^k g(x_k, \mu(x_k), w_k) + \alpha^N J(x_N)|x_n = x]$$

Then we can write DP algorithm for policy evaluation:

$$\begin{aligned} J_0^\mu &= J \\ J_1^\mu &= T_\mu J \\ &\cdots \\ J_N^\mu &= T_\mu J_{N-1}^\mu = \cdots = T_\mu^N J \end{aligned}$$

Greedy policies
When we apply bellman operator on cost-to-go function J, we take the minimum. We now define the set of policies that attain those minimum. We call them the greedy policies of J : $G(J) = \{\mu | T_\mu J = TJ\}$, i.e.

$$\mu(x) \in \arg\min_{u \in U(x)} \mathbb{E}[g(x,u,w) + \alpha J(f(x,u,w))] \quad \forall x$$

.

2.3 Main result

Proposition 1. *1. There exists a unique $J^\star : \mathcal{X} \to \mathbb{R}$ (called the optimal cost-to-go function) that satisfies the fixed point equation:*

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

. For any J,

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

2. For every stationary policy μ, there exists a unique cost-to-go function $J^\mu : \mathcal{X} \to R$ that solves the fixed point equation:

$$J^\mu = T_\mu J^\mu.$$

For any J,

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

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

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

2.4 Remarks on interpretation

1. The first part of the theorem states that for a given policy, its cost-to-go function is the unique fixed point of the Bellman operator T_μ. Recalling the definition of T_μ, this can be interpreted as a temporal consistency condition: the expected cost to go must equal the expected instantaneous cost plus the expected cost-to-go from the next state. Recall that $T_\mu^N J$

is the cost-to-go function for an N period problem; the second part of item (1) identifies J^μ as its infinite horizon limit.

2. The second part of the algorithm shows that the optimal cost function $T^N J$ for a N period problem converges at a geometric rate to an infinite horizon limit: $J^* = \lim_{N\to\infty} T^N J$. Moreover, the optimal cost-to-go function J^* is the unique solution to the Bellman fixed point equation $J^* = J^*$.

3. The third part of the result shows that stationary policy μ is optimal if it attains the argmin (implicitly expressed via T) in the Bellman optimality equation $J^* = TJ^*$.

2.5 Properties of the Bellman operator

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

1. **Monotonicity:** If $J(x) \leqslant J'(x) \quad \forall x \in X$, then $TJ(x) \leqslant TJ'(x) \quad \forall x \in X$. We can also write $J \preceq J' \Rightarrow TJ \preceq TJ'$.

2. **Constant Shift:** $T(J + c \cdot e) = TJ + \alpha c \cdot e$ where we take $e(x) = 1 \quad \forall x \in X$ to denote a vector of all ones.

3. **Contraction:** $\|TJ - TJ'\|_\infty \leqslant \alpha \|J - J'\|_\infty \quad \forall J, J'$. In words, applying T to two cost-to-go functions brings them geometrically closer.

The exact same statements also hold for $T\mu$. 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:

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

Applying monotonicity gives:

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

Applying constant shift gives:

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

Subtracting TJ' yields our desired result. □

2.6 Contraction mapping theorem

We will prove this for $F : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{X})$, where $\mathcal{B}(\mathcal{X}) = \{J : \mathcal{X} \to \mathbb{R} : \|J\|_\infty < \infty\}$. Note that we need the completeness of the space for the theorem to work (a Cauchy sequence converges to a limit in a complete metric space).

Proposition 2. *If $||FJ - FJ'|| \leqslant \alpha\|J - J'\|$ 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. *For all $J \in \mathcal{B}(\mathcal{X})$, $\|F^N J - J^*\| \leqslant \alpha^N \|J - J'\|$.*

Proof. We first prove that $\{J_k\}$ is Cauchy, and by the completeness of the space the limit J_∞ exists.

For fixed J, let $J_0 = J$, $J_k = FJ_{k-1}$. We have:

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

$\forall m \geqslant 1$, we have:

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

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:

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

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. Recalling that $FJ_\infty = J_\infty$, we have:

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

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:

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

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 the map $x \mapsto e^x$ magnifies differences in the input, whereas contraction mappings "dampen" them.

2.7 Back to our main result

We now have the necessary tools to prove the main proposition.

Proof. The unique existence of J^* is guaranteed by the Contraction Mapping Theorem, thus 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 $J^\mu = J^*$.

We now show no policy can attain lower cost We first show this for stationary policies, i.e. we 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 applying the monotonicity property gives:

$$J^\mu \succsim TJ^\mu \succsim T^2 J^\mu \succsim ... \succsim T^k J^\mu \succsim ... \succsim J^*$$

Now, for non-stationary π, denote $M = ||g||_\infty$, and

$$\begin{aligned} J^\pi(x) = \lim_{N\to\infty} \mathbb{E}^\pi\Big[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k)\Big| x_0 = x\Big] &\geqslant \lim_{N\to\infty} \mathbb{E}^\pi\Big[\sum_{k=0}^{m-1} \alpha^k g(x_k, u_k, w_k)\Big| x_0 = x\Big] - \sum_{k=m}^{N-1} \alpha^k M \\ &\geqslant (T^m \vec{0})(x) - \frac{\alpha^m}{1-\alpha} M \\ &\to J^*(x) \text{ as } m \to \infty. \end{aligned}$$

□