Course Notes On Dynamic Optimization (Fall 2023)
Lecture 7A: Indefinite Horizon Problems

Instructor: Daniel Russo

Email: djr2174@gsb.columbia.edu

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

These notes are partly 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.

Topics:

- Stochastic shortest path problems as a generalization of finite horizon and discounted problems.
- Contraction in a weighted maximum norm.

1 Indefinite horizion problems

1.1 Overview and perspective

So far we have covered two important problem classes: finite horizon (with general additive objective) and infinite horizon problems with discount objectives.

Some situations are not so naturally modeled with either formulation. Imagine you play an Atari game trying to accrue as many points as possible before your player perishes. Fixing a finite horizon and searching over nonstationary policies seems like a clunky way to model the situation. But discounting also seems artificial — the 'financial' interest accrued during a 30 minute game is. . .not large.

Here we cover *stochastic shortest path problems*. These model interactions, like the Atari game or interactions with a customer on a web-service, which continue until some special termination state is reached. The goal is to minimize cumulative expected cost (or maximize cumulative expected reward) accrued throughout this *indefinite*[1] horizon.

I (Dan) love this class of problems because they unify my thinking. Dynamic programming can quickly get cluttered. One can study finite horizon models, infinite horizon discounted models, infinite horizon total cost models (in both the positive and negative cost case), average

[1]Indefinite: "lasting for an unknown or unstated length of time."

cost objectives, and so on. Indefinite horizon models provide a single formalism which generalizes finite horizon and infinite-horizon discounted problems, and extends seamlessly to study a special class of problems with average cost objective. Many other issues in more advanced DP (e.g. costs being infinite under some very bad policy) are previewed in this case, also.

1.2 Transition probability notation

In these notes, I will restrict to finite state spaces (or, later, countable state spaces) and will work the transition probabilities

$$p(x'|x,u) = \mathbb{P}(f(x,u,w_k) = x'),$$

where the probability on the right-hand-side is integrating over the i.i.d disturbance w_k. In words, this is the probability of transitioning to x' when control u was applied in previous state x.

1.3 Problem Formulation

We consider the problem of minimizing expected costs until a special termination state $\emptyset$ is reached.

- The state space is $\mathcal{X} \cup \{\emptyset\}$ where $\mathcal{X}$ where $\mathcal{X}$ is countable.
- The 'terminal state' $\emptyset$ is costless ($g(\emptyset, u) = 0$ and absorbing ($\mathbb{P}(x_{k+1} = \emptyset | x_k = \emptyset, u_k = u) = 1$). This implies that any policy incurs zero expected cost starting from $\emptyset$.
- Single period costs are uniformly bounded, i.e. $\sup_{x,u} |g(x,u)| < \infty$.
- Minima are attained, i.e. for any J and x, $\arg\min_{u \in U(x)} g(x,u) + \sum_{x' \in \mathcal{X}} g(x,u) + p(x'|x,u)J(x')$ is nonempty. (*For instance, the set of feasible controls at any state is finite.*)

Remark (Warning on notation)**.** *We can view the cost-to-go function J either as (1) having domain $\mathcal{X} \cup \{\emptyset\}$ or as (2) having domain $\mathcal{X}$. In the first case, we follow the implicit convention that $J(\emptyset) = 0$ any time we write J, without repeating this. In the second case, you should remember that the transition matrix induced by a policy μ is* sub-stochastic*, in the sense that $\sum_{x' \in \mathcal{X}} p(x'|x,u) \leqslant 1$ for given state/control pair (x,u).*

Define the termination time

$$\tau = \inf\{k \in \{0,1,2,\ldots\} \,:\, x_k = \emptyset\},$$

following the convention that $\tau = \infty$ if the terminal state is never reached. Define the cost-to-go function of a policy

$$J^\pi(x) = \mathbb{E}^\pi \left[\sum_{k=0}^{\tau} g(x_k, u_k) \mid x_0 = x \right]$$

A policy π^* is said to be optimal if it satisfies $J^{\pi^*}(x) = \inf_\pi J^\pi(x)$ for every $x \in \mathcal{X}$.

Assumption 1. *Under any policy and initial state, the terminal state is reached with probability 1. Moreover, the expected termination time is uniformly bounded:*

$$\sup_\pi \sup_{x \in \mathcal{X}} \mathbb{E}^\pi \left[\tau \mid x_0 = x \right] < \infty.$$

1.4 Casting finite horizon problems as a special case

Roughly speaking, finite horizon problems are a special case of our formulation in one views the pair (x, k) indicating both a 'state of some system' x and a 'state of time' k as the overall system state. Each time period is associated with a transition $x_k = (x, k) \to x_{k+1} = (x', k+1)$ until some period $k = N-1$ after which the system transitions to the the terminal state.

More formally, finite horizon problems are a special case of our formulation satisfying the following condition.

- $\mathcal{X}$ factors into N disjoint sets as $\mathcal{X} = \mathcal{X}_0 \cup \cdots \cup \mathcal{X}_{N-1}$ where

$$p(\emptyset|x,u) = 1 \text{ for all } x \in \mathcal{X}_{N-1}\,,\ u \in U(x)$$

  and, for each $k < N-1$,

$$\sum_{x' \in \mathcal{X}_{k+1}} p(x'|x,u) = 1 \text{ for all } x \in \mathcal{X}_k\,,\ u \in U(x).$$

1.5 Casting infinite horizon discounted problems as a special case

Infinite horizon discounted problems are *essentially* a special case of our formulation in which

$$p(\emptyset|x,u) = 1 - \alpha \in (0,1)$$

for all x, u. In this case, $\tau \mid \pi, x_0 \sim \text{Geometric}(1-\alpha)$ and

$$\begin{aligned} J^\pi(x) = \mathbb{E}^\pi\left[\sum_{k=0}^{\tau} g(x_k,u_k) \mid x_0 = x\right] &= \lim_{N\to\infty} \mathbb{E}^\pi\left[\sum_{k=0}^{N} \mathbb{1}(\tau \geqslant k) g(x_k,u_k) \mid x_0 = x\right] \\ &= \lim_{N\to\infty} \mathbb{E}^\pi\left[\sum_{k=0}^{N} \mathbb{P}(\tau \geqslant k) g(x_k,u_k) \mid x_0 = x\right] \\ &= \lim_{N\to\infty} \mathbb{E}^\pi\left[\sum_{k=0}^{N} \alpha^k g(x_k,u_k) \mid x_0 = x\right]. \end{aligned}$$

Every policy incurs the same expected cost in this indefinite horizon problem as it does in the analogous infinite horizon discounted problem.

2 Theory of indefinite horizon problems

2.1 Bellman operators

Following the convention that $J \in \mathbb{R}^{\mathcal{X}}$ (omitting the terminal state). As before, Bellman operators can be undertood as mapping the space of cost-to-go function into itself; here the correct space is the set of bounded functions $\mathcal{J} = \{J \in \mathbb{R}^{\mathcal{X}} : \|J\|_\infty < \infty\}$. Define the Bellman operator $T_\mu : \mathcal{J} \to \mathcal{J}$ by

$$T_\mu J(x) = g(x, \mu(x)) + \sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) J(x')$$

and the the Bellman optimality operators $T : \mathcal{J} \to \mathcal{J}$ by

$$TJ(x) = \min_{u \in U(x)} \; g(x,u) + \sum_{x' \in \mathcal{X}} p(x'|x,u) J(x').$$

These have the following properties:

- Monotonicity: If $J \succeq J'$ then $TJ \succeq TJ'$ (same for T_μ).
- Sub-constant shift: for a scalar $r > 0$ and a vector of all ones $e \in \mathbb{R}^{\mathcal{X}}$, $T(J + re) \preceq T(J) + re$. (The same holds for T_μ and the inequality is reversed if r is negative.)
- Contraction: uh oh, these are not contractions in the maximum norm. Are they contraction operators in some other norm?

2.2 Contraction

We aim to construct a norm in which the Bellman operator is a contraction, searching in the space of weighted maximum-norms. For a strictly positive weighting $w \in \mathbb{R}^{\mathcal{X}}_{>0}$, define the weighted maximum-norm

$$\|J\|_{\infty,w} = \max_{x \in \mathcal{X}} w(x)|J(x)|.$$

Which weighting do we pick? We eventually choose $w(x) = 1/V(x)$ for some "Lyapunov" function V. A Lyapunov function is some kind of "generalized energy function" which is chosen such that energy is expected to dissipate (on average) from any initial state. For our purposes we define it formally as follows.

Definition 1. We say $V : \mathcal{X} \to \mathbb{R}_{>0}$ is a Lyapunov function if there exists $\beta < 1$ such that

$$\max_{u \in U(x)} \sum_{x' \in \mathcal{X}} p(x'|x,u) V(x') \leqslant \beta V(x) \text{ for all } x \in \mathcal{X}.$$

Define β_V to be the smallest scalar β satisfying this inequality.

The next result shows that the Bellman operators are contractions in the weighted maximum norm induced by a Lypunov function.

Proposition 1. *If V is Lyapunov function, then T and T_μ are contractions with respect to the weighted maximum norm $\|\cdot\|_{\infty,1/V}$ with modulus β_V.*

The previous proposition is quite general, but identifying a Lyapunov function must be done on a case-by-case basis. The next lemma provides a Lyapunov function for our problem.

Lemma 1. *The function $V(x) = \sup_\pi \mathbb{E}^\pi [\tau \mid x_0 = x]$ is a Lypunov function with $\beta_V \leqslant \frac{\|V\|_\infty - 1}{\|V\|_\infty}$.*

Proof. Recognize that V is the cost-to-go function in an alternative problem with "costs" $g(x,u) = -1$ for all x and u; one can show that it satisfies the Bellman equation

$$V(x) = 1 + \max_{u \in U(x)} \sum_{x' \in \mathcal{X}} p(x'|x,u) V(x') \qquad \forall x \in \mathcal{X}. \tag{1}$$

Re-arranging terms yields

$$\sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) V(x') \leqslant V(x) - 1.$$

Dividing both sides by $V(x)$ and taking the maximum over x yields

$$\max_{x \in \mathcal{X}} \frac{\sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) V(x')}{V(x)} \leqslant \max_{x \in \mathcal{X}} \frac{V(x) - 1}{V(x)} = \frac{\|V\|_\infty - 1}{\|V\|_\infty}.$$

□

2.3 Understanding the weighted maximum norm

It is immediate from the definition that our bound on β_V is close to 1 when the expected termination time could be large, under some policy and some initial state. A somewhat unsatisfying feature of this theory is that you might expect 'reasonably good' policies to rarely visit certain states and to terminate quickly, but nevertheless β_V would depend on the worst policy and state you could choose.

Discounted problems are a special case of our formulation in which termination time has distribution $\tau \mid x_0, \pi \sim \text{Geometric}(1 - \alpha)$. In this case, $\|V\|_\infty = 1/(1 - \gamma)$ so

$$\beta_V \leqslant 1 - \frac{1}{1/(1 - \alpha)} = \alpha.$$

Therefore, we recover our previous result about contractivity of the Bellman operator in the discounted case.

In the finite horizon case, $\|V\|_\infty = N$ and so $\beta_V \leqslant (N - 1)/N = 1 - 1/N$.

Finally, observe that a small weighted-max norm implies the unweighted max-norm is small, since

$$\|J\|_{\infty, 1/V} = \max_{x \in \mathcal{X}} \frac{J(x)}{V(x)} \geqslant \max_{x \in \mathcal{X}} \frac{J(x)}{\|V\|_\infty} = \frac{\|J\|_\infty}{\|V\|_\infty},$$

or $\|J\|_\infty \leqslant \|J\|_{\infty, 1/V} \|V\|_\infty$.

2.4 Proof of Proposition 1

Proof. First, we establish the result for T_μ:

$$\begin{aligned}
\|T_\mu J - T_\mu J'\|_{\infty, 1/V} &= \max_{x \in \mathcal{X}} \frac{1}{V(x)} \left| \sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) (J(x') - J'(x')) \right| \\
&= \max_{x \in \mathcal{X}} \frac{1}{V(x)} \left| \sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) V(x') \left(\frac{J(x') - J'(x')}{V(x')} \right) \right| \\
&\leqslant \max_{x \in \mathcal{X}} \left(\frac{\sum_{x' \in \mathcal{X}} p(x'|x, \mu(x)) V(x')}{V(x)} \right) \|J - J'\|_{\infty, 1/V} \\
&\leqslant \alpha \|J - J'\|_{\infty, 1/V}
\end{aligned}$$

Now, since T_μ is a contraction,

$$\frac{T_\mu J(x)}{V(x)} \leqslant \frac{T_\mu J'(x)}{V(x)} + \beta_V \|J - J'\|_{\infty,1/V} \qquad \forall \mu.$$

Then

$$\frac{TJ(x)}{V(x)} = \min_\mu \frac{T_\mu J(x)}{V(x)} \leqslant \min_\mu \frac{T_\mu J'(x)}{V(x)} + \beta_V \|J - J'\|_{\infty,1/V} = \frac{TJ'(x)}{V(x)} + \beta_V \|J - J'\|_{\infty,1/V}.$$

Reversing the role of J and J' gives

$$\frac{|TJ(x) - TJ'(x)|}{V(x)} \leqslant \alpha \|J - J'\|_{\infty,1/V} \qquad \forall x \in \mathcal{X}.$$

□