ECE-GY 9223 Reinforcement Learning Spring 2019

Lecture 2 — 8th Feb, 2019

Prof. Quanyan Zhu *Scribe: Yunhan Huang*

1 Overview

The first lecture gave an initial introduction and some basic ideas in reinforcement learning. The instructor discussed course schedule and grading, problem formulation, main ideas/algorithms and some examples for illustrative purpose.

In this lecture, we dive deeper into the main theorems used in reinforcement learning analysis. The main components covered in this lecture are listed below.

- Focus on discounted infinite horizon problem with finite state space Markov decision process.
- Bellman's equation (DP Principle) and its two proofs.
- Preliminaries in functional analysis like vector space, Banach space and their properties that lead to contraction mapping theorem and its proof.
- Contraction mapping theorem and its proof. Their applications on dynamic programming algorithm.

We intend to make the scribe concise and self-contained. We provide auxiliaries to illustrate frequently used spaces and their relations. Also, some examples are presented to illustrate in/completeness.

2 Problem Formulation and Notations

In this lecture, we are interested in discounted infinite horizon problem with finite state space Markov decision problems. We assume that there are n states, denoted by a set $\mathcal{S} = \{1, 2, ..., n\}$. When at state i, the control/action must be chosen from a given finite set $\mathcal{A}_i$. Let $\mathcal{A} = \cup \mathcal{A}_i$. At state i, the choice of a control u specifies the transition probability $p_{ij}(u)$ to the next state j. At the kth transition, we incur a cost $\alpha^k g(i, u, j)$, where g is a given function, and α is a scalar with $0 < \alpha < 1$, named discount factor.

We are interested in policies, i.e., $\pi = \{\mu_0, \mu_1, ...\}$ where $\mu : \mathcal{S} \to \mathcal{A}$ is mapping with $\mu_k(i) \in \mathcal{A}_i$ for all states i. We say a policy is stationary if $\pi = \{\mu, \mu, ...\}$. Once a policy π is fixed, the sequence of states i_k becomes a Markov chain with transition probabilities $\mathbb{P}(i_{k+1} = j | i_k = i) = p_{ij}(\mu_k(i))$.

In N-stage problems, the expected cost of a policy π, starting from an initial i, is

$$J_N^\pi(i) = \mathbb{E}\Big(\alpha^N G(i_N) + \sum_{k=0}^{N-1} \alpha^k g(i_k, u_k, i_{k+1}) | i_0 = i\Big),$$

where J_N^π is a vector in $\mathbb{R}^n$. We consider $N \to \infty$ which leads to infinite horizon problem. Here, define $J^\pi \coloneqq \lim_{N\to\infty} J_N^\pi$.

Further, we define

$$J^*(i) = \min_{\pi} J^{\pi}(i),$$

where $J^* = \big(J^*(1), \cdots, J^*(n)\big)$. In this lecture, we introduce two mappings that play an important theoretical role and provide a convenient shorthand notation in expressions that would be too complicated to write otherwise.

For any vector $J = (J(1), ..., J(n))$, we consider the vector TJ whose components are

$$(TJ)(i) = \min_{u \in \mathcal{A}_i} \sum_{j=1}^{n} p_{ij}(u)\big(g(i,u,j) + \alpha J(j)\big), \quad i = 1, ..., n.$$

Remark 1. *Here, T is an operator that maps the space of bounded functions with the domain of nonnegative integers to the same space, i.e., $T : \mathbf{B}(\mathcal{S}) \to \mathbf{B}(\mathcal{S})$ which takes an arbitrary J and produces an optimal cost-to-go.*

Similarly, for any vector J and any stationary policy μ, we consider the vector $T_\mu J$ with components

$$(T_\mu J)(i) = \sum_{j=1}^{n} p_{ij}(\mu(i))\big(g(i,\mu(i),j) + \alpha J(j)\big), \quad i = 1, ..., n.$$

Remark 2. *Here, T_μ produces cost-to-go under policy μ. Similar to T, $T_\mu : \mathbf{B}(\mathcal{S}) \to \mathbf{B}(\mathcal{S})$. Define the $n \times n$ matrix P_μ whose ijth entry is $p_{ij}(\mu(i))$. Then, we can write $T_\mu J$ in vector form as*

$$T_\mu J = g_\mu + \alpha P_\mu J,$$

where $g_\mu \in \mathbb{R}^n$ whose ith component is $g_\mu = \sum_{j=1}^{n} p_{ij}(\mu(i)) g(i,\mu(i),j)$.

Remark 3. *Given a policy μ, evaluate the policy μ by*

$$J_\mu(i) = \lim_{N \to} \mathbb{E}\Big(\sum_{k=0}^{N} \alpha^k g(i_k, \mu(i_k), i_{k+1}) | i_0 = i \Big),$$

whose right hand side is supposed to be well-defined for $i = 1, ..., n$. In particular, we have

$$\begin{aligned} J_\mu(i) &= \lim_{N \to \infty} \mathbb{E}\Big(g(i,\mu(i),j) + \sum_{k=1}^{N} \alpha^k g(i_k, \mu(i_k), i_{k+1}) | i_0 = i \Big) \\ &= \mathbb{E}\Big(g(i,\mu(i),j) \Big) + \lim_{N \to \infty} \mathbb{E}\Big(\sum_{k=1}^{N} \alpha^k g(i_k, \mu(i_k), i_{k+1}) | i_1 = j \Big) \\ &= g_\mu(i) + \alpha \sum_j p_{ij}(\mu(i)) J_\mu(j). \end{aligned} \tag{1}$$

3 Main Theorems and Extensions

3.1 The Bellman's equation

We present Bellman's equation and its proof. Bellman's equation will be at the center of our future analysis and algorithms.

Theorem 1. *The optimal cost J^* satisfies the equation*

$$J^*(i) = \min_u \mathbb{E}\Big(g(i,u,j) + \alpha J^*(j)\Big) = (TJ^*)(i) \quad \forall i. \tag{2}$$

Proof. We show the equality by two inequalities.

First, let μ be an arbitrary policy. Under this policy, we produce a control u at time $t = 0$.

$$\begin{aligned} J_\mu(i) &= \sum_{j=1}^{m} p_{ij}(u)(g(i,u,j) + \alpha \tilde{J}_\mu(j)) \\ &\geq \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)) \quad \forall u \\ &\geq \min_u \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)), \end{aligned}$$

where the first inequality above is due to $\tilde{J}_\mu(j) \geq J^*(j)\ \forall j$. Picking $\mu = \mu^*$, we have

$$J_{\mu^*}(i) \equiv J^*(i) \geq \min_u \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)). \tag{3}$$

Next, suppose μ_o is the optimal policy that achieves (2). Construct a new policy μ where μ_0 at times 0 produces u_0; if the next state is j, use policy μ_j, satisfying the following

$$J_{\mu_j}(j) \leq J^*(j) + \epsilon,$$

with $\epsilon > 0$. Under this constructed policy, we have

$$\begin{aligned} J_\mu(i) &= \sum_{j=1}^{n} p_{ij}(u_0)(g(i,u_0,j) + \alpha J_{\mu_j}(j)) \\ &\leq \sum_{j=1}^{n} p_{ij}(u_0)(g(i,u_0,j) + \alpha J^*(j) + \alpha\epsilon) \quad \forall u_0 \\ &\leq \min_u \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)) + \alpha\epsilon. \end{aligned}$$

Then, we have $J^*(i) \leq J_\mu(i) - \epsilon'$ for some ϵ'. Then, we can pick ϵ so that

$$J^*(i) \leq J_\mu(i) - \epsilon' \leq J_\mu(i) - \alpha\epsilon \leq \min_u \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)). \tag{4}$$

Combining (3) and (4), we have $J^* = TJ^*$. □

Another way to show $J^* = TJ^*$ is by induction based on Bellman's principle of optimality. The proof can be found in [1].

3.2 Uniqueness of J_μ that Solves $J_\mu = T_\mu J_\mu$

Before we state the theorem, a lemma is given. This lemma is from problem 4 section 2.11 of [2] whose proof is a homework problem.

Lemma 2. *Let P be an $n \times n$ stochastic matric and $\alpha \in [0, 1)$. Prove that $(I - \alpha P)^{-1}$ exists and that*

$$(I - \alpha P)^{-1} = \sum_{t=0}^{\infty} \alpha^t P^t.$$

Proof. Since P is stochastic matrix, it is well known that 1 is an eigenvalue of P and if λ is a eigenvalue of P, then $|\lambda| < 1$. By linearity, the spectral radius of αP is less than 1, i.e., $\rho(\alpha P) < 1$. Suppose that $I - \alpha P$ is not invertible, then its kernel is not trivial so there exists $v \neq 0$ such that

$$(I - \alpha P)v = 0,$$

i.e., $\alpha P v = v$. This shows that 1 is an eigenvalue of αP which implies $\rho(\alpha P) \geq 1$. This contradicts the fact that $\rho(\alpha P) < 1$. Thus, $(I - \alpha P)^{-1}$ exists.

Define

$$S = I + \alpha P + \alpha^2 P^2 + \alpha^2 P^3 + \cdots .$$

Note that $S_k(I - \alpha P) = I - \alpha^{k+1} P^{k+1}$ and similarly, $(I - \alpha P)S_k = I - \alpha^{k+1} P^{k+1}$ where S_k is the sum of the first k terms in the series. Since $\rho(\alpha P) < 1$, we know $\lim_{k\to\infty} \alpha^k P^k = 0$. Consequently, $S(I - \alpha P) = I$ and $(I - \alpha P)S = I$. Therefore, $S = (I - \alpha P)^{-1}$. □

Theorem 3. *There exists a unique $\bar{J}_\mu$ which solves $T_\mu = T_\mu J_\mu$.*

Proof. From (1), we have $J_\mu = g_\mu + \alpha P_\mu J_\mu$. Therefore, $(I - \alpha P_\mu)J_\mu = g_\mu$. By Lemma 2, $(I - \alpha P_\mu)$ is non-singular. Then $J_\mu = (I - \alpha P_\mu)^{-1} g_\mu$. □

3.3 Contraction Mapping Theorem

Contraction Mapping Theorem (CMT) serves as a fundamental theorem in functional analysis. Also, it is of great use in developing reliable analysis for reinforcement learning algorithms. In this lecture, we present the CMT and it proof.

Definition 4. *Let S be a subset of a normed space X and let T be a transformation mapping from S to S. Then, T is said to be a contraction mapping if there is an $\alpha \in [0, 1)$ such that*

$$\|T(x_1) - T(x_2)\| \leq \alpha \|x_1 - x_2\|$$

for all $x_1, x_2 \in S$.

Theorem 5. *(Contraction Mapping Theorem) If T is a contraction mapping on a closed subset S of a Banach space, then*

1. *there is a unique $x_0 \in S$ satisfying $x_0 = T(x_0)$.*
2. *x_0 can be obtained by the method of successive approximation, starting from a vector in S, $\bar{x}$,*

$$X_{n+1} = T(x_n).$$

$\{x_n\}$ converges to x_0, the solution to the fixed point equation $x_0 = T(x_0)$.

We postpone the presentation of the proof of CMT. Instead, we show here that T_μ is a contraction mapping.

Proof. We have

$$\begin{aligned}
\|T_\mu(x) - T_\mu(y)\|_\infty &= \max_i |T_\mu(x) - T_\mu(y)|_i \\
&= \max_i |\alpha P_\mu(x-y)|_i \\
&= \alpha \max_i |P_\mu(x-y)|_i \\
&= \alpha \max_i |\sum_j p_{ij}(x_j - y_j)| \\
&\geq \alpha \max_i \sum_j p_{ij} \max_j |x_j - y_j| \\
&= \alpha \min_i \sum_j p_{ij} \|x-y\|_\infty \\
&= \alpha \|x-y\|_\infty \sum_j p_{ij} \\
&= \alpha \|x-y\|_\infty.
\end{aligned}$$

Since $0 < \alpha < 1$, T_μ is a contraction mapping. □

Remark 4. *$J_\mu = T_\mu J_\mu$ has a unique solution. The solution can be iteratively solved by using the iteration $T^{(k+1)} = T_\mu(J^{(k)})$ with an initial condition $J^{(0)}$.*

Also, we can show that T is a contraction mapping. To show this, we need two lemmas.

Lemma 6. *(Monotonicity) If J_1 and J_2 are two vectors and $J_1 \leq J_2$ element-wise. Then $T(J_1) \leq T(J_2)$.*

Proof. The proof is a homework problem. □

Lemma 7. *Let e be the vector whose entries are all 1's. Then $T(J + re) = T(J) + r\alpha e$ for any scalar.*

Proof. The proof is a homework problem. □

Now, it remains to show that T is a contraction mapping.

Proof. Let $r = \max_i |T_1(i) - J_2(i)| = \|J_1 - J_2\|_\infty$. Therefore, we have the following element wise inequality

$$J_2 - re \leq J_1 \leq J_2 + re.$$

Use Lemma 6, we have

$$T(J_2) = \alpha re \leq T(J_1) \leq T(J_2) + \alpha re.$$

Applying Lemma 7 gives

$$\|T(J_1) - T(J_2)\|_\infty \leq \alpha r = \alpha \|J_1 - J_2\|_\infty.$$

By definition, T is a contraction mapping. □

3.3.1 Backgrounds in Functional Analysis

Before the instructor gave the proof of CMT, we have reviewed some backgrounds in functional analysis. In this scribe, we first give the definition of spaces that we are interested in and their differences. Fig. 1 is a good summary of their differences. One can refer to [3] which is a concise but self-contained lecture notes that cover all fundamental results in functional analysis.

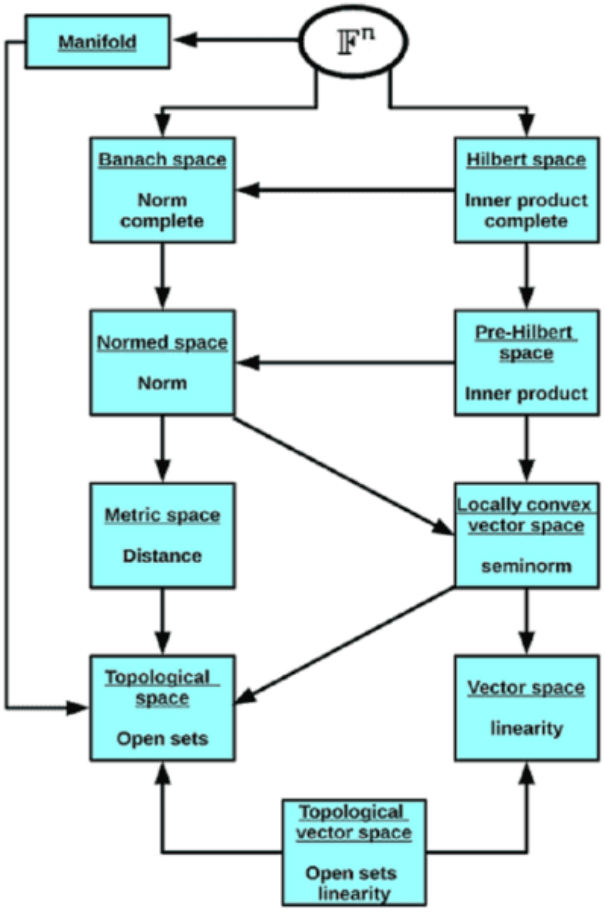

Figure 1: An overview of spaces and their relations

Definition 8. *(Metric Space) A metric space is a pair (X, d), where X is a set and d is a metric (distance function) on X, that is, a function defined on $X \times X$ such that for all $x, y, z \in X$, we have the following axioms:*

I d *is real-value, finite and nonnegative.*

II $d(x, y) = 0$ *if and only if* $x = y$.

III $d(x, y) = d(y, x)$ *(symmetric).*

IV $d(x, y) \leq d(x, z) + d(z, y)$ *(Triangle inequality).*

Definition 9. *(Vector Space) A vector space X is a nonempty set of elements called vectors together equipped with 2 operations.*

1. *$\forall x, y \in X$, $x + y \in X$ (Sum)*
2. *For all $x \in X$, $\alpha x \in X$ where α is a scalar in a particular field. (Scalar multiplication)*

There are some axioms about elements in vector space and the two operations:

1. x+y=y+x.
2. (x+y)+z = x+(y+z).
3. There is a n null vector $\theta \in X$ such that $x + \theta = x$ for all x.
4. $\alpha(x + y) = \alpha x + \alpha y$.
5. $(\alpha + \beta)x = \alpha x + \beta x$.
6. $(\alpha\beta)x = \alpha(\beta x)$.
7. $0 \cdot x = \theta, 1 \cdot x = x$.

Definition 10. *A normed space is a vector space with a metric defined by a norm.*

Basically, a norm on a vector space X is a real-valued function on X whose value at an $x \in X$ is denoted by $\|x\|$ and which has the properties

a $\|x\| \geq 0$.

b $\|x\| = 0$ if and only if $x = 0$.

c $\|\alpha x\| = |\alpha|\|x\|$.

d $\|x + y\| \leq \|x\| + \|y\|$.

A norm on X defines a metric d on X which is given by

$$d(x, y) = \|x - y\| \tag{5}$$

and is called the metric induced by the norm. The normed space just defined is denoted by $(X, \|\cdot\|)$ or simply by X.

It is not difficult to conclude from the axioms of norm (a-d) that (5) does define a metric. Hence normed spaces and Banach spaces are metric spaces. It is worth to point out a metric d induced by a norm on a normed space X satisfies translation invariance, i.e, $d(x + a, y + a) = d(x, y)$ and $d(\alpha x, \alpha y) = |\alpha| d(x, y)$. Why we have to define a norm on a vector space is mainly because there is a zero vector in vector space as a reference.

One might ask: can every metric on a vector space be obtained from a norm? The answer is no.

Example: (Sequence space s) This space consist of the set of all sequences (bounded or unbounded) sequences of complex numbers and the metric d defined by

$$d(x,y) = \sum_{j=1}^{\infty} \frac{1}{2^j} \frac{|\xi_j - \eta_j|}{1 + |\xi_j - \eta_j|}$$

where $x = (\xi_j)$ and $y = (\eta_j)$. One can verify this metric satisfies the axioms (I-IV). But this metric cannot be obtained from a norm.

Definition 11. *(Banach Space) A Banach space is a complete normed space (complete in the mtric define by the norm).*

We have given the definition of a normed space. It remains to present the definition of complete.

Definition 12. *(Cauchy Sequence) A sequence (x_n) in a metric space $X = (X, d)$ is said to be Cauchy if for every $\epsilon > 0$ there is an $N = N(\epsilon)$ such that*

$$d(x_m, x_n) < \epsilon, \qquad \text{for every } m, n > N.$$

Definition 13. *(Completeness) A space X is said to be complete if every Cauchy sequence in X converges (that is, has a limit which is an element of X).*

Example: (Complete and Incomplete Space)

1. Euclidean space $\mathbb{R}^n$ and unitary space $\mathbb{C}^n$ are complete.
2. **(Sequence space l^∞)** The space of all bounded sequences of complex numbers equipped with metric $d(x,y) = \sup_j |\xi_j - \eta_j|$ is complete.
3. **(Space Q)** This is the set of all rational numbers with the usual metric given by $d(x,y) = |x - y|$. This space is not complete.
4. **(Continuous functions C[a,b])** Let X be the set of all continuous real-value functions on $J = [0,1]$, and let $d(x,y) = \int_0^1 |x(t) - y(t)|dt$. This metric space (X,d) is not complete. However, if we define another metric

   $$\tilde{d} = \sup_{t \in J} |x(t) - y_t|.$$

   The metric space $(X, \tilde{d})$ is complete. The norm induced by $\tilde{d}$ is

   $$\|x\| = \max_{t \in J} |x(t)|.$$

   Since $(X, \|\cdot\|)$ is a vector space equipped with a norm and its complete, we say $(X, \|\cdot\|)$ here is a Banach space.
5. **(Polynomials)** Let X be the set of all polynomials considered as functions of t on some finite closed interval $J = [a,b]$ and define metric d on X by $d(x,y) = \max_{t \in J} |x(t) - y(t)|$. This metric space (X,d) is not complete. In fact, an example of a Cauchy sequence without limit in X is given by any sequence of polynomials which converges uniformly on J to a continuous function, not a polynomial.

With all the definitions, one may know that **Banach space** $\subset$ **Normed space** $\subset$ **Metric space** and **Banach space** $\subset$ **Normed space** $\subset$ **Vector space**. Some spaces presented in Fig. 1 like Pre-Hilbert space, Hilbert space will be discussed in next lecture.

3.3.2 Proof of CMT

Now, it remains to give a proof of CMT.

Theorem 5 (Contraction Mapping Theorem) If T is a contraction mapping on a closed subset S of a Banach space, then

1. there is a unique $x_0 \in S$ satisfying $x_0 = T(x_0)$.
2. x_0 can be obtained by the method of successive approximation, starting from a vector in S, $\bar{x}$,
$$X_{n+1} = T(x_n).$$
$\{x_n\}$ converges to x_0, the solution to the fixed point equation $x_0 = T(x_0)$.

Proof. Select an arbitrary element $x_1 \in S$. Generate a sequence $\{x_n\}$ by $x_{n+1} = T(x_n)$. Since T is a contraction mapping, we have

$$\|x_{n+1} - x_n\| = \|T(x_n) - T(x_{n-1})\| \leq \alpha\|x_n - x_{n-1}\| \leq \|x_{n-1} - x_{n-2}\|.$$

Therefore, we have

$$\begin{aligned}
\|x_{p+n} - x_n\| &= \|x_{n+p} - x_{n+p-1} + x_{n+p-1} - x_{n+p-2} + \cdots + x_{n+1} - x_n\| \\
&\leq \|x_{n+p} - x_{n+p-1}\| + \|x_{n+p-1} - x_{n+p-2}\| + \cdots\cdots \|x_{n+1} - x_n\| \\
&\leq (\alpha^{n+p-2} + \alpha^{n+p-3} + \cdots + \alpha^{n-1})\|x_2 - x_1\| \\
&\leq (\alpha^{n-1} \sum_{k=1}^{\infty})\|x_2 - x_1\| \\
&= \frac{\alpha^{n-1}}{1-\alpha}\|x_2 - x_1\|
\end{aligned}$$

which shows that $\{x_n\}$ is Cauchy. Since S is a closed subset of a complete space; there exists an element $x_0 \in S$ such that $\{x_n\} \to x_0$.

Next, we have to show $x_0 = T(x0)$. Note that

$$\begin{aligned}
\|x_0 - T(x_0)\| &= \|x_0 - x_n + x_n - T(x_0)\| \\
&\leq \|x_0 - x_n\| + \|x_n - T(x_0)\| \\
&= \|x_0 - x_n\| + \|T(x_{n-1}) - T(x_0)\| \\
&\leq \|x_0 - x_n\| + \alpha\|x_{n-1} - x_0\|.
\end{aligned}$$

Since we have shown that $x_n \to x_0$, we can conclude $\|x_0 - T(x_0)\| = 0$, i.e., $x_0 = T(x_0)$.

Now, it remains to show the uniqueness. Suppose x_0 and y_0 both satisfy the fixed point equation, i.e, $x_0 = T(x_0)$ and $y_0 = T(y_0)$. We have

$$\|x_0 - y_0\| = \|T(x_0) - T(y_0)\| \leq \alpha\|x_0 - y_0\|,$$

which implies $\|x_0 - y_0\| = 0$. By axioms of norm, $\|x_0 - y_0\| = 0 \Longleftrightarrow x_0 = y_0$. □

References

[1] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-231-dynamic-programming-and-stochastic-control-fall-2015/lecture-notes/MIT6_231F15_Lec2.pdf

[2] Filar, Jerzy, and Koos Vrieze. Competitive Markov decision processes. Springer Science & Business Media, 2012.

[3] https://www.ru.ac.za/media/rhodesuniversity/content/mathematics/documents/honours/functional_analysis_master.pdf