ECE-GY 9223 Reinforcement Learning Spring 2019

Lecture 3 — Feb. 15, 2019

Prof. Quanyan Zhu *Scribe: Tongyu Zong*

1 Overview

In the last lecture we talked about the fixed-point problem of J^*, contraction mapping theorem as well as some concepts including linear vector space, norm, Cauchy sequence, complete space and Banach space.

In this lecture we continued to focus on classical projection theorem. At the beginning, we introduced some definitions including pre-Hilbert space and Hilbert space. Then, we introduced in detail what projection theorem means. We gave some examples to illustrate the application of projection theorem as well as its extensions. Furthermore, we discussed about one of the methods for computations for MDPs: Value Iteration. Next lecture will focus on Policy Iteration.

2 Classical Projection Theorem

2.1 Hilbert Space

- **Pre-Hilbert Space**:

  A **pre-Hilbert** space is a linear vector space X together with an inner product defined on X x X corresponding to each pair of vectors $x, y \in X$. The inner product $(x|y)$ of x and y is a scalar, satisfying

$$(x|y) = \overline{(y|x)}$$

$$(x+y|z) = (x|z) + (y|z)$$

$$(\lambda x|y) = \lambda\,(x|y)$$

$$(x|x) \geq 0 \; and \; (x|x) = 0 \; iff \; x = 0$$

  Also, $\|x\| = \sqrt{(x|x)}$ is a norm.

- **Hilbert Space**:

  A complete pre-Hilbert space is called **Hilbert Space**.

$$\mu_i\,(x_i, x_{-i})$$

- *Remark 1:* In a pre- Hilbert Space, $x, y \in X$.

  Def. x and y are **orthogonal** if $(x|y) = 0$, denoted as $x \perp y$.

  If $x \perp y$, $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

2.2 The Classical Projection Theorem

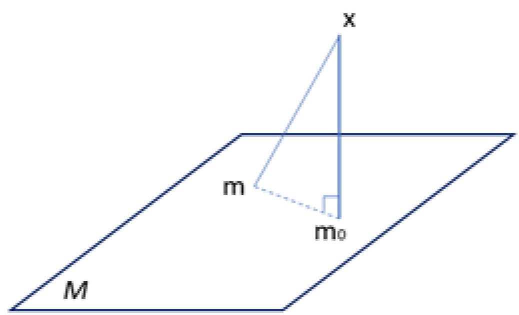
Let H be a Hilbert Space and M is a closed subspace of H.

a) Corresponding to any vector $x \in H$, there is a unique vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$.

b) A necessary and sufficient condition that $m_0 \in M$ be the unique minimizing vector is that $(x - m_0) \perp M$.

The first conclusion needs that the space should be Hilbert and the second one only needs that the space should be pre-Hilbert.

Furthermore, the graphic illustration to this theorem is as below:



- *Proof*

Let's first **prove b)**:

Necessity:

Let's show that if m_0 is a minimizing vector then $x - m_0$ is orthogonal to M.

Suppose this is not true. Then there exists $m \in M$ which is not orthogonal to $x - m_0$ i.e. $(x - m_0|m) = \delta \neq 0$. Without loss of generality, assume $\|m\|^2 = 1$.

Define $m_1 = m_0 + \delta m$. Then because $m_0 \in M$ and $\delta m \in M$, we have $m_1 \in M$.

$$\therefore \; \|x - m_1\|^2 = \|x - m_0 - \delta m\|^2$$

$$= \|x - m_0\|^2 + \delta^2 \|m\|^2 - (x - m_0|\delta m) - (\delta m|x - m_0)$$

$$= \|x - m_0\|^2 + \delta^2 - \delta^2 - \delta^2 = \|x - m_0\|^2 - |\delta|^2 < \|x - m_0\|^2$$

$\Longrightarrow m_1 \in M$ is strictly better than m_0.

$\therefore$ m_0 is not a minimizing vector, which violates the assumption.

Therefore, there cannot exist a $m \in M$ that is not orthogonal to $x - m_0$.

Sufficiency: If $(x - m_0) \perp M$ then m_0 is a **unique** minimizing vector.

From the Pythagorean Theorem, $\|x - m\|^2 = \|x - m_0 + m_0 - m\|^2 = \|x - m_0\|^2 + \|m_0 - m\|^2$, where $m \neq m_0$. Therefore, $\|x - m\|^2 > \|x - m_0\|^2$ for $m \neq m_0 \Longrightarrow m_0$ is a unique minimizer.

Let's first **prove a)**:

We require Hilbert space's completeness.

If $x \in M$ then $m_0 = x$.

If not, let $\delta = \|x - m\|$. We want to produce a $m_0 \in M$ with $\|x - m_0\| = \delta$.

Let $\{m_i\}$ be a sequence of vectors in M such that $\|x - m_0\| \to \delta$, which means that $m \to m_0$. Now let's prove $\{m_i\}$ is Cauchy.

Based on the equation $2\|a\|^2 + 2\|b\|^2 = \|a+b\|^2 + \|a-b\|^2$, just let $a = m_j - x$ and $b = x - m_i$. Then we have

$$2\|m_j - x\|^2 + 2\|x - m_i\|^2 = \|m_j - m_i\|^2 + \|m_j + m_i - 2x\|^2$$

$$= \|m_j - m_i\|^2 + 4\|x - 0.5(m_j + m_i)\|^2$$

Due to the fact that $m_i, m_j \in M$, we can conclude $0.5(m_i + m_j) \in M$. Just let $\|x - 0.5(m_j + m_i)\| \geq \delta$, then we have

$$\|m_j - m_i\| \leq 2\|m_j - x\|^2 + 2\|x - m_i\|^2 - 4\delta^2$$

Because $\|x - m_i\| \to \delta$ and $\|x - m_i\| \to \delta$ as $i, j \to \infty$, we can conclude

$$\|m_j - m_i\|^2 \to 0 \; as \; i, j \to \infty$$

So $\{m_i\}$ is a Cauchy sequence.

Due to the fact that M is a closed subspace of a complete space, we have $m_0 \in M$. By continuity, we can conclude $\|x - m_0\| = \delta$.

Example1: Approximation

Suppose $y_1, y_2, \ldots, y_n$ are elements of a Hilbert Space H which generate a closed finite-dimension subspace M. Given $x \in H$, find $\widehat{x} \in M$ ($\widehat{x} = \sum_i \alpha_i y_i$) that minimizes $\|x - \widehat{x}\|$.

For this problem, we can use projection theorem: $\widehat{x}$ is the orthogonal projection of x on M:

$(x - \widehat{x}|y_i) = 0$ for $i = 1, 2, \ldots, n$.

Therefore,

$$(y_1|y_1)\,\alpha_1 + (y_2|y_1)\,\alpha_1 + \ldots + (y_n|y_1)\,\alpha_n = (x|y_1)$$

$$(y_1|y_2)\,\alpha_1 + (y_2|y_2)\,\alpha_1 + \ldots + (y_n|y_2)\,\alpha_n = (x|y_2)$$

$$\ldots\ldots$$

$$(y_1|y_n)\,\alpha_1 + (y_2|y_n)\,\alpha_1 + \ldots + (y_n|y_n)\,\alpha_n = (x|y_n)$$

In this way we turn this problem into matrix eqution:

$$G\alpha = b$$

in which the symmetric matrix G is called GRAM matrix of $y_1, y_2, \ldots, y_n$. To guarantee the non-singularity of G, $y_1, y_2, \ldots, y_n$ should be linearly independent.

- Theorem:

Let $y_1, y_2, \ldots, y_n$ be linearly independent, then let δ be the minimal distance from a vector x to subspace M generated by $\{y_i\}$. Then we have:

$$\delta = min \left\| x - \sum_i \alpha_i y_i \right\| = min \left\| x - \widehat{x} \right\|$$

$$\delta^2 = \frac{g\left(y_1, y_2, \ldots, y_n, x\right)}{g\left(y_1, y_2, \ldots, y_n\right)}$$

in which g is the determination of the GRAM matrix.

Example2: Hilbert space of random variable:

Let $\mathcal{L}^2\left(\Omega, \mathcal{F}; C\left[0,1\right]\right)$ be space of all parameterized random variables, where $\mathcal{F}$ is σ-algebra and $C\left[0,1\right]$ consists of all continuous functions in $[0,1]$. Then $X\left(t;\omega\right)$ is a random process, for which $t \in [0,1]$ and $\omega \in \Omega$. For every fixed time t, $\mathrm{X}\left(t;\ \bullet\right)$ is a second-order random variable defined on $\left(\Omega, \mathcal{F}, \mathbb{R}\ \right)$ and $X\left(\bullet\ ;\omega\right)$ is a continuous function in $C\left[0,1\right]$.

Now define inner product:

$$\left(X, Z\right) = \mathbb{E}\left[\int_0^1 X\left(t;\omega\right) Z\left(t;\omega\right) w\left(t\right) \mathrm{dt}\right]$$

in which $w\left(t\right) > 0$ and $w\left(t\right) \in C\left[0,1\right]$, which is a weighted factor.

Determine a stochastic process $\widehat{X}\left(t;\omega\right) \in \mathcal{L}^2\left(\Omega, \mathcal{F}; C\left[0,1\right]\right)$, minimizing the norm $\left\| X - \widehat{X} \right\|$ and satisfying $\mathbb{E}\left[\int_0^1 \widehat{X}\left(t;\omega\right) K_i\left(t\right) y_i\left(\omega\right) \mathrm{dt}\right] = C_i,\ i = 1, 2,\ K_1, K_1 \in C\left[0,1\right]$ and C_1, C_1 are constants. In this case, we can know that $\widehat{X}$ should be related with K and Y. And we can use projection theorem to find the random process in a Hilbert space.

Extension of classical projection theorem:

Let x be a vector in Hilbert space H and K be a closed convex subset of H. Then,

a) There's a unique vector $k_0 \in K$ such that $\|x - k_0\| \leq \|x - k\|$ for all $k \in K$.

b) The necessary and sufficient condition that $k_0 \in K$ be the minimizing vector is $(x - k_0 | k - k_0) \leq 0$ for all $k \in K$.

Example3: Least-square estimation:

Suppose y is an $m\ x\ 1$ vector and W is an m x n matrix with linearly independent columns. Then, there is a unique minimizer β^* which minimizes:

$$min \left\| y - W\beta \right\|$$

This problem is similar to Example 1. Let $W = (\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n})$. Then $\widehat{y} = W\beta = \sum_i \beta_i \mathbf{w_i}$. Using projection theorem, the solution β^* should satisfy the orthogonality:

$(y - \widehat{y} | \mathbf{w_i}) = 0$ for $i = 1, 2, \ldots, n$.

We can obtain

$$W^T W \beta^* = W^T y$$

Finally,

$$\beta^* = \left(W^T W\right)^{-1} W^T y$$

- Extension:

  Let y_1, y_2 be given vectors, X_1, X_2 be given matrices, and ${X_1}^T X_1$ be positive definite.

  Problem 1: $\varphi_1 = argmin \left\|y_1 - X_1 r\right\|^2$

  Problem 2: $\varphi_2 = argmin \left\|y_1 - X_1 r\right\|^2 + \left\|y_2 - X_2 r\right\|^2$

  Let's begin:

  Problem 1:let $f\left(r\right) = \left\|y_1 - X_1 r\right\|^2 = {y_1}^T y_1 - r^T {X_1}^T y_1 - {y_1}^T X_1 r + r^T {X_1}^T X_1 r$.
  $\frac{\partial f}{\partial r} = -2{X_1}^T y_1 + 2{X_1}^T X_1 r = 0$, (FOC)

  $$r^* = \varphi_1 = \left({X_1}^T X_1\right)^{-1} {X_1}^T y_1$$

  Let φ_0 be any given initial condition, then we rewrite φ_1 as below to construct a recursion:

  $$\varphi_1 = \ \varphi_0 + \left({X_1}^T X_1\right)^{-1} {X_1}^T \left(y_1 - X_1\ \varphi_0\right)$$

  Then we can prove that φ_2 can also be written in this form using $\varphi_1,\ X_2,\ y_2$.

  Problem 2: The second problem is also a convex problem, we can get the result by **first-order condition** as well:

  $$\varphi_2 = \ \left({X_1}^T X_1 + {X_2}^T X_2\right)^{-1} \left({X_1}^T y_1 + {X_2}^T y_2\right)$$

  $$= \ \left({X_1}^T X_1 + {X_2}^T X_2\right)^{-1} \left({X_1}^T X_1 \varphi_1 + {X_2}^T X_2 \varphi_1 - {X_2}^T X_2 \varphi_1 + {X_2}^T y_2\right)$$

  $$= \ \left({X_1}^T X_1 + {X_2}^T X_2\right)^{-1} \left(\left({X_1}^T X_1 + {X_2}^T X_2\right)\varphi_1 + {X_2}^T \left(y_2 - X_2 \varphi_1\right)\right)$$

  $$= \varphi_1 + \ \left({X_1}^T X_1 + {X_2}^T X_2\right)^{-1} {X_2}^T \left(y_2 - X_2 \varphi_1\right)$$

  Therefore, φ_2 can also be written in the same form as φ_1. In this way, we construct a recursion process for the problems.

Example4: Kalman Filtering: Assume that ${X_1}^T X_1$ is positive definite.

$$\varphi_i = argmin \sum_{j=1}^{i} \lambda^{i-j} \left\|y_j - X_j r\right\|^2, \quad i = 1, 2, \ldots, m$$

$$\varphi_i = \varphi_{i-1} + {H_i}^{-1} {X_i}^T \left(y_i - X_i \varphi_{i-1}\right)$$

$$H_i = \lambda H_{i-1} + {X_i}^T X_i, \quad i = 1, 2, \ldots, m$$

$$H_0 = 0$$

Extended Kalman Filtering: generalize$\left\|y_j - X_j r\right\|^2$ as some $\left\|g_j\left(r\right)\right\|^2$.

3 Computations for MDPs (Discounted Infinite-Horizon Problem)

3.1 Value Iteration (VI)

T is a contraction mapping, we can use iteration methods:

$$\widehat{J}_{k+1} = T\left(\widehat{J}_k\right) \text{ with a } \widehat{J}_0$$

By contraction mapping theorem,

$$\widehat{J}_k \to J^*, \ \textit{asymptotically}$$

1. VI convergence can be slow.
2. When it iterates for k steps, we need to evaluate how close the solution is.

   Measure $\left\|\widehat{J}_k - J^*\right\|_\infty$:

   $$\left\|\widehat{J}_k - J^*\right\|_\infty = \left\|T\left(\widehat{J}_{k-1}\right) - T\left(J^*\right)\right\|_\infty \leq \alpha \left\|\widehat{J}_{k-1} - J^*\right\|_\infty \leq \alpha^k \left\|\widehat{J}_0 - J^*\right\|_\infty$$

   However, we don't know J^*, but we know:

   $$\left\|\widehat{J}_{k+1} - \widehat{J}_k\right\|_\infty \leq \alpha \left\|\widehat{J}_k - \widehat{J}_{k-1}\right\|_\infty \leq \alpha^k \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty$$

   Therefore,

   $$\left\|\widehat{J}_k - J^*\right\|_\infty \leq \left\|\widehat{J}_k - \widehat{J}_{k+1}\right\|_\infty + \left\|\widehat{J}_{k+1} - \widehat{J}_{k+2}\right\|_\infty + \ldots + \left\|\widehat{J}_{k+l-1} - \widehat{J}_{k+l}\right\|_\infty + \left\|\widehat{J}_{k+l} - J^*\right\|_\infty$$

   $$\leq \alpha^k \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty + \alpha^{k+1} \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty + \ldots + \alpha^{k+l-1} \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty + \left\|\widehat{J}_{k+l} - J^*\right\|_\infty$$

   $$\leq \frac{\alpha^k}{1-\alpha} \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty + \left\|\widehat{J}_{k+l} - J^*\right\|_\infty$$

   Let $l \to \infty$, then $\widehat{J}_{k+l} \to J^*$. Therefore,

   $$\left\|\widehat{J}_k - J^*\right\|_\infty \leq \frac{\alpha^k}{1-\alpha} \left\|\widehat{J}_1 - \widehat{J}_0\right\|_\infty$$

3. How about the policy μ_k?

   $\mu_k(i)$ is a solution to

   $$\left(T\widehat{J}_k\right)(i) = min \sum_j P_{\text{ij}}(u) \left(g(i,u,j) + \alpha \widehat{J}_k(j)\right)$$

   We also know that

   $$\left(T_\mu \widehat{J}_k\right)(i) = \sum_j P_{\text{ij}}(\mu(i)) \left(g(i,\mu(i),j) + \alpha \widehat{J}_k(j)\right)$$

   So if $\mu = \mu_k$, then

$$\left(T_{\mu_k}\widehat{J}_k\right)(i)=\sum_j P_{\text{ij}}\left(\mu_k\left(i\right)\right)\left(g\left(i,\mu_k\left(i\right),j\right)+\alpha\widehat{J}_k\left(j\right)\right)=\left(T\widehat{J}_k\right)(i)=\widehat{J}_{k+l}\left(i\right)$$

for any i.

Let's consider how to evaluate μ_k:

We know that J_{μ_k} is the solution of the fixed-point problem $x=T\left(x\right)$, i.e.

$$J_{\mu_k}=T_{\mu_k}\ J_{\mu_k}$$

Now evaluate J_{μ_k} by $\|J_{\mu_k}-J^*\|_\infty$:

$$\|J_{\mu_k}-J^*\|_\infty\leq\left\|J_{\mu_k}-\widehat{J}_k\right\|_\infty+\left\|\widehat{J}_k-J^*\right\|_\infty$$

Let $A=\left\|J_{\mu_k}-\widehat{J}_k\right\|_\infty$ and $B=\left\|\widehat{J}_k-J^*\right\|_\infty$.

Part A:

$$\left\|J_{\mu_k}-\widehat{J}_k\right\|_\infty=\left\|J_{\mu_k}-T_{\mu_k}\left(\widehat{J}_k\right)+T_{\mu_k}\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty$$

$$\leq\left\|T_{\mu_k}\left(J_{\mu_k}\right)-T_{\mu_k}\left(\widehat{J}_k\right)\right\|_\infty+\left\|T_{\mu_k}\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty\leq\alpha\left\|J_{\mu_k}-\widehat{J}_k\right\|_\infty+\left\|T_{\mu_k}\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty$$

Therefore,

$$\left\|J_{\mu_k}-\widehat{J}_k\right\|_\infty\leq\frac{1}{1-\alpha}\left\|T_{\mu_k}\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty=\frac{1}{1-\alpha}\left\|T\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty$$

Part B:

$$\left\|\widehat{J}_k-J^*\right\|_\infty\leq\left\|\widehat{J}_k-T\left(\widehat{J}_k\right)\right\|_\infty+\left\|T\left(\widehat{J}_k\right)-J^*\right\|_\infty$$

$$\leq\left\|\widehat{J}_k-T\left(\widehat{J}_k\right)\right\|_\infty+\left\|T\left(\widehat{J}_k\right)-T\left(J^*\right)\right\|_\infty$$

$$\leq\left\|\widehat{J}_k-T\left(\widehat{J}_k\right)\right\|_\infty+\alpha\left\|\widehat{J}_k-J^*\right\|_\infty$$

Therefore,

$$\left\|\widehat{J}_k-J^*\right\|_\infty\leq\frac{1}{1-\alpha}\left\|T\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty$$

According to Part A and Part B, we can obtain:

$$\|J_{\mu_k}-J^*\|_\infty\leq\frac{2}{1-\alpha}\left\|T\left(\widehat{J}_k\right)-\widehat{J}_k\right\|_\infty$$

$$=\frac{2}{1-\alpha}\left\|\widehat{J}_{k+1}-\widehat{J}_k\right\|_\infty\leq\frac{2\alpha^k}{1-\alpha}\left\|\widehat{J}_1-\widehat{J}_0\right\|_\infty$$

Using this inequation, we can obtain the "distance" between J_{μ_k} and J^* as the evaluation of the policy μ_k.

References

[1] D G. Luenberger, "Optimization by vector space methods", John Wiley & Sons, pp. 46-102, 1997.

[2] D P. Bertsekas, J N. Tsitsiklis, "Neuro-dynamic programming", Belmont, MA: Athena Scientific, pp. 11-28, 1996.