Control theory and Reinforcement Learning - Lecture 2

Carlos Esteve Yagüe

Universidad Autónoma de Madrid - Fundación Deusto

September 2020

We have a **deterministic** discrete-time dynamical system in $\mathbb{R}^n$:

$$\begin{cases} x_{t+1} = f(x_t, u_t) & t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}^n. \end{cases}$$

1. The sequence $u_0, u_1, \dots \in \mathbb{R}^m$ are the controls (or actions) that we can choose.
2. $x_0, x_1, x_2, \dots \in \mathbb{R}^n$ is the sequence of states of the system.
3. The function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ determine the dynamics (might be known or unknown).

Problem:

$$\underset{u_0, u_1, u_3 \dots}{\text{minimize}} \quad \sum_{t=0}^{\infty} \gamma^t C(x_t, u_t)$$

$$\underset{u_0,u_1,u_3\ldots}{\text{minimize}} \quad \sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \quad \text{subject to} \quad \begin{cases} x_{t+1} = f(x_t, u_t) & t = 0, 1, 2, 3, \ldots \\ x_0 = x \in \mathbb{R}^n. \end{cases}$$

Policy: for $t = 0, 1, 2, \ldots$,

$$u_t = \pi(x_0, x_1, \ldots, x_t, u_0, u_1, \ldots, u_{t-1})$$

The value function

$$V(x) = \min_{\pi(\cdot)} \sum_{t=0}^{\infty} \gamma^t C(x_t, u_t)$$

Bellman equation:

$$V(x) = \min_u \{C(x, u) + \gamma V(f(x, u))\}$$

Value iteration: Set an initial guess $V_0(x)$, and then improve the approximation

$$V_{k+1}(x) := \min_u \{C(x, u) + \gamma V_k(f(x, u))\}$$

Plan of the Lecture

- Bellman equation for continuous-time models with deterministic dynamics.
- Bellman equation for continuous-time models with stochastic dynamics.
- Value iteration method for continuous-time models.

Linear deterministic dynamics in $\mathbb{R}$: let $h > 0$ be fixed

$$\begin{cases} x_{t+1} = x_t + h u_t & t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}. \end{cases} \qquad \left(\begin{array}{c} \text{we can make it more general} \\ x_{t+1} = x_t + h\, f(x_t,\, u_t) \end{array} \right)$$

Value function: quadratic cost and discount factor $\gamma = e^{-h}$

$$V(x) = \min_{\pi(\cdot)} \left\{ \sum_{t=0}^{\infty} h\, e^{-ht} (x_t^2 + u_t^2) \right\}$$

Observation

If we rewrite the dynamics as

$$\frac{x_{t+1} - x_t}{h} = u_t, \qquad t = 0, 1, 2, 3, \dots$$

the limit of this dynamical system as $h \to 0^+$ corresponds to the continuous dynamics

$$x'(t) = u(t), \qquad t \in (0, \infty).$$

Linear deterministic dynamics in $\mathbb{R}$: let $h > 0$ be fixed

$$\begin{cases} x_{t+1} = x_t + hu_t & t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}. \end{cases} \qquad \left(\begin{array}{c} \text{we can make it more general} \\ x_{t+1} = x_t + h\, f(x_t,\, u_t) \end{array} \right)$$

Value function: quadratic cost and discount factor $\gamma = e^{-h}$

$$V(x) = \min_{\pi(\cdot)} \left\{ \sum_{t=0}^{\infty} h\, e^{-ht} (x_t^2 + u_t^2) \right\}$$

Observation

The cost functional associated to the continuous-time control $u(t)$ and its corresponding state $x(t)$ is given by

$$\int_0^{\infty} e^{-t} (x(t)^2 + u(t)^2)$$

Bellman equation:

$$V(x) = \min_u \left\{ h(x^2 + u^2) + e^{-h} V(x + hu) \right\}$$

We can rewrite it as

$$0 = \min_u \left\{ x^2 + u^2 + \frac{e^{-h} V(x + hu) - V(x)}{h} \right\}$$

By letting $h \to 0$ we obtain

$$0 = x^2 + \min_u \left\{ u^2 - V(x) + uV'(x) \right\}$$

which is equivalent to

Hamilton-Jacobi equation

$$V(x) + \frac{V'(x)^2}{4} = x^2$$

Stochastic dynamics in $\mathbb{R}$: let $h > 0$ be fixed

$$\begin{cases} x_{t+1} = x_t + hu_t + \sqrt{h}w_t & t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}, \end{cases}$$

where w_t is a random variable such that

$$\Pr[w_t = 1] = 0.5 \qquad \text{and} \qquad \Pr[w_t = -1] = 0.5$$

Value function: quadratic cost and discount factor $\gamma = e^{-h}$

$$V(x) = \min_{\pi(\cdot)} \mathbb{E}_w \left\{ \sum_{t=0}^{\infty} he^{-ht}(x_t^2 + u_t^2) \right\}$$

Bellman equation:

$$V(x) = \min_u \mathbb{E}_w \left\{ h(x^2 + u^2) + e^{-h} V(x + hu + \sqrt{h}w) \right\}$$

Using the definition of expectation, we obtain

$$V(x) = \min_u \left\{ h(x^2 + u^2) + \frac{1}{2} e^{-h} V(x + hu + \sqrt{h}) + \frac{1}{2} e^{-h} V(x + hu - \sqrt{h}) \right\}$$

that we can rewrite as

$$0 = \min_u \left\{ x^2 + u^2 + \frac{1}{2} \left(\frac{e^{-h} V(x + hu + \sqrt{h}) + e^{-h} V(x + hu - \sqrt{h}) - 2V(x)}{h} \right) \right\}$$

and also as

$$0 = \min_u \left\{ x^2 + u^2 + \frac{e^{-h}}{2} \left(\frac{V(x + hu + \sqrt{h}) + V(x + hu - \sqrt{h}) - 2V(x)}{h} \right) + \frac{e^{-h} V(x) - V(x)}{h} \right\}$$

Bellman equation:

$$V(x) = \min_u \mathbb{E}_w \left\{ h(x^2 + u^2) + e^{-h} V(x + hu + \sqrt{h}w) \right\}$$

Using the definition of expectation, we obtain

$$V(x) = \min_u \left\{ h(x^2 + u^2) + \frac{1}{2} e^{-h} V(x + hu + \sqrt{h}) + \frac{1}{2} e^{-h} V(x + hu - \sqrt{h}) \right\}$$

that we can rewrite as

$$0 = \min_u \left\{ x^2 + u^2 + \frac{1}{2} \left(\frac{e^{-h} V(x + hu + \sqrt{h}) + e^{-h} V(x + hu - \sqrt{h}) - 2V(x)}{h} \right) \right\}$$

and also as

$$0 = \min_u \left\{ x^2 + u^2 + \frac{e^{-h}}{2} \left(\frac{V(x + hu + \sqrt{h}) + V(x + hu - \sqrt{h}) - 2V(x)}{h} \right) + \frac{e^{-h} V(x) - V(x)}{h} \right\}$$

$$\lim_{h\to 0^+} \frac{V(x+hu+\sqrt{h})+V(x+hu-\sqrt{h})-2V(x)}{h} = ??$$

Let us assume that V is sufficiently smooth:

$$\begin{aligned} V(x+hu+\sqrt{h}) &= V(x)+(hu+\sqrt{h})V'(x)+\frac{(hu+\sqrt{h})^2}{2}V''(x)+o(h) \\ V(x+hu-\sqrt{h}) &= V(x)+(hu-\sqrt{h})V'(x)+\frac{(hu-\sqrt{h})^2}{2}V''(x)+o(h) \end{aligned}$$

Then,

$$\frac{V(x+hu+\sqrt{h})+V(x+hu-\sqrt{h})-2V(x)}{h} = 2uV'(x)+V''(x)+\frac{o(h)}{h}$$

Example 2

We plug it in the Bellman equation to obtain

$$0 = \min_u \left\{ x^2 + u^2 + \frac{e^{-h}}{2} \left(2uV'(x) + V''(x) + \frac{o(h)}{h} \right) + \frac{e^{-h}V(x) - V(x)}{h} \right\}.$$

By letting $h \to 0$ we obtain

$$0 = x^2 - V(x) + \frac{1}{2}V''(x) + \min_u \left\{ u^2 + uV'(x) \right\}$$

which is equivalent to

Viscous Hamilton-Jacobi equation

$$V(x) - \frac{1}{2}V''(x) + \frac{V'(x)^2}{4} = x^2$$

Remark

The diffusion coefficient depends on the variance of the stochastic process w_t.

Repeat the computations with $\Pr\left[w_t = \pm\frac{1}{4}\right] = 0.5$.

Stochastic dynamics in $\mathbb{R}^n$: Let $h > 0$ be fixed

$$\begin{cases} x_{t+1} = x_t + h\, f(x_t, u_t) + \sqrt{h} w_t \quad t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}^n, \end{cases}$$

where w_t is a Gaussian noise in $\mathbb{R}^n$ with 0 mean and covariance matrix $\Sigma_w > 0$.

Value function: running cost $h\, C(x_t, u_t)$ and discount factor $\gamma = e^{-h}$

$$V(x) = \min_{\pi(\cdot)} \mathbb{E}_w \left\{ \sum_{t=0}^{\infty} h e^{-ht} C(x_t, u_t) \right\}$$

Bellman equation:

$$V(x) = \min_u \mathbb{E}_w \left\{ h\, C(x,u) + e^{-h} V(x + h\, f(x,u) + \sqrt{h} w) \right\}$$

Using the definition of expectation, we can write

$$V(x) = \min_u \left\{ h\, C(x,u) + \frac{e^{-h}}{\sqrt{(2\pi)^n |\Sigma_w|}} \int_{\mathbb{R}^n} e^{-\frac{y' \Sigma_w^{-1} y}{2}} V(x + h\, f(x,u) + \sqrt{h} y) dy \right\}$$

Recall: the probability density for a n-dimensional normal distribution $\mathcal{N}(0, \Sigma_w)$:

$$p(y) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_w|}} e^{-\frac{y' \Sigma_w^{-1} y}{2}}$$

Letting $h \to 0^+$ (after some computations), we obtain the **viscous Hamilton-Jacobi-Bellman equation:**

$$V(x) - \frac{1}{2}\sum_{i,j} \sigma_{ij}\partial_{ij} V(x) - H(x, \nabla V(x)) = 0$$

where the Hamiltonian H is defined as

$$H(x,p) := \min_u \{p \cdot f(x,u) + C(x,u)\},$$

and σ_{ij} are the coefficients of the covariance matrix Σ_w.

Observation: The assumption $\Sigma_w > 0$ ensures the uniform ellipticity of the differential operator $\sum_{i,j} \sigma_{ij}\partial_{ij}$.

- **Linear dynamics:** $f(x,u) = Ax + Bu$
- **Quadratic cost for the control:** $C(x,u) = \|u\|^2 + \phi(x)$.
- **Isotropic Gaussian noise:** $\Sigma_w = \sigma\, I_n$.

How does HJB equation look like?

We can compute the Hamiltonian explicitly:

$$\begin{aligned} H(x,p) &= \min_u \{p \cdot f(x,u) + C(x,u)\} \\ &= \min_u \left\{p \cdot (Bu) + \|u\|^2\right\} + p \cdot (Ax) + \phi(x) \\ &= -\frac{\|B'p\|^2}{4} + p \cdot (Ax) + \phi(x). \end{aligned}$$

In this case, the second-order operator $\sum_{i,j} \sigma_{ij}\partial_{ij}$ is just the Laplacian.

Hamilton-Jacobi-Bellman equation

$$V - \sigma \Delta V + \frac{\|B'\nabla V\|^2}{4} - \nabla V \cdot (Ax) = \phi(x).$$

Discrete deterministic dynamics in $\mathbb{R}^n$: let $h > 0$ be fixed

$$\begin{cases} x_{t+1} = x_t + h\, f(x_t, u_t) & t = 0, 1, 2, 3, \dots \\ x_0 = x \in \mathbb{R}^n. \end{cases}$$

The value function: discount factor $\gamma = e^{-h}$.

$$V(x) = \min_{\pi(\cdot)} \left\{ \sum_{t=0}^{\infty} h\, e^{-ht} C(x_t, u_t) \right\}.$$

Value iteration: initialize with $V_0(x)$ arbitrarily chosen, and the improve it as follows

$$V_{k+1}(x) = \min_{u} \left\{ h\, C(x, u) + e^{-h} V_k(x + h\, f(x, u)) \right\}.$$

As we have seen before, the continuous model is obtained by letting $h \to 0^+$.

Value iteration for continuous-time problems

Value iteration: initialize with $V_0(x)$ arbitrarily chose, and then improve it as follows

$$V_{k+1}(x) = \min_u \left\{ h\, C(x,u) + e^{-h} V_k(x + h\, f(x,u)) \right\}.$$

The value function $V(x)$ is obtained as **the limit as k goes to infinity**:

$$V(x) = \lim_{k \to \infty} V_{k+1}(x)$$

Recall: The convergence is ensured by the discount factor $\gamma = e^{-h} \in (0,1)$.

What about the continuous model? i.e. take the limit as $h \to 0$.

Value iteration: initialize with $V_0(x)$ arbitrarily chose, and then improve it as follows

$$V_{k+1}(x) = \min_u \left\{ h\, C(x,u) + e^{-h} V_k(x + h\, f(x,u)) \right\}.$$

The value function $V(x)$ is obtained as **the limit as k goes to infinity**:

$$V(x) = \lim_{k \to \infty} V_{k+1}(x)$$

Recall: The convergence is ensured by the discount factor $\gamma = e^{-h} \in (0,1)$.

What about the continuous model? i.e. take the limit as $h \to 0$.

By subtracting $V_k(x)$ in either side of the iteration formula and dividing by h, we obtain

$$\frac{V_{k+1}(x) - V_k(x)}{h} = \min_u \left\{ C(x,u) + \frac{e^{-h} V_k(x + h\, f(x,u)) - V_k(x)}{h} \right\}$$

We can interpret k as the time-variable. That is, we make the following change of variable

$$k \mapsto t_k := k\, h, \qquad \text{for } k = 0, 1, \ldots$$

and we then identify

$$V_k(x) = V(x,\, k\, h)$$

For all $t_k = k\, h$, we can write the iteration formula as follows

$$\frac{V(x,\, t_k + h) - V(x,\, t_k)}{h} = \min_u \left\{ C(x,u) + \frac{e^{-h} V(x + h\, f(x,u),\, t_k) - V(x,\, t_k)}{h} \right\}$$

By taking the limit as $h \to 0^+$, we obtain the value iteration algorithm for the time-continuous problem as follows:

Value iteration for time-continuous problem

Set an initial guess

$$V(x, 0) = V_0(x)$$

The value function is obtained as the limit

$$V(x) = \lim_{t \to \infty} V(x, t)$$

where $V(x, t)$ is the solution to the **time-evolution Hamilton-Jacobi equation**

$$\begin{cases} \partial_t V(x, t) = H(x, \nabla V(x, t)) - V(x) & (x, t) \in \mathbb{R}^n \times (0, \infty) \\ V(x, 0) = V_0(x), \end{cases}$$

where the Hamiltonian H is defined as

$$H(x, p) := \min_u \{p \cdot f(x, u) + C(x, u)\}$$

Value iteration for continuous-time problems

If we add Gaussian noise w_t in the dynamics with covariance matrix $\Sigma_w = I_n$, i.e.

$$x_{t+1} = x_t + h f(x_t, u_t) + \sqrt{h} w_t,$$

we obtain a **viscous Hamilton-Jacobi equation.**

Value iteration for time-continuous problem with noise

Set an initial guess

$$V(x, 0) = V_0(x)$$

The value function is obtained as the limit

$$V(x) = \lim_{t \to \infty} V(x, t)$$

where $V(x, t)$ is the solution to the time-evolution problem

$$\begin{cases} \partial_t V(x,t) - \Delta V(x,t) = H(x, \nabla V(x,t)) - V(x) & (x,t) \in \mathbb{R}^n \times (0, \infty) \\ V(x,0) = V_0(x), \end{cases}$$

where the Hamiltonian H is defined as

$$H(x, p) := \min_u \{p \cdot f(x, u) + C(x, u)\}$$