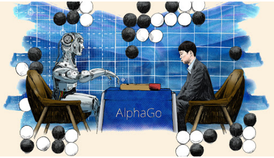
Control theory and Reinforcement Learning - Lecture 1

Carlos Esteve Yagüe

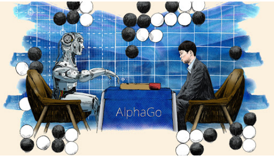
Universidad Autónoma de Madrid - Fundación Deusto

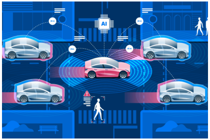
September 2020

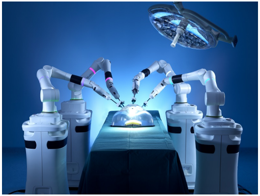
We aim to act on a controlled environment in order to achieve a prescribed goal



We aim to act on a controlled environment in order to achieve a prescribed goal

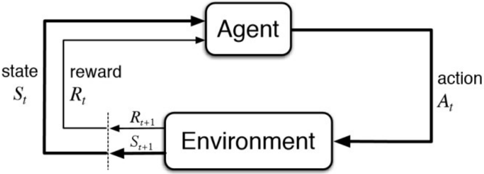






Definition: Reinforcement Learning is the study of how to use past data to enhance the future manipulation of a dynamical system.

Origins of RL: Samuel, Klopf, Werbös, in the 1960's and 70's
Barto, Sutton, Bertsekas from the 1990's-. . .
and many others

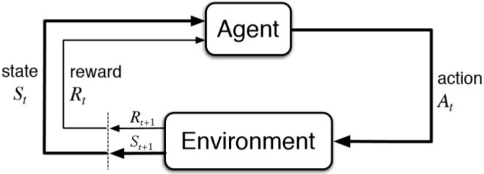


Drawing from Sutton and Barto, Reinforcement Learning: An Introduction, 1998.

Control Theory	**Reinforcement Learning**
Continuous or discrete setting	Discrete setting (Markov Decision Processes)
model + opt. criteria $\Big\}\longrightarrow$ action	data $\longrightarrow$ action

Definition: Reinforcement Learning is the study of how to use past data to enhance the future manipulation of a dynamical system.

Origins of RL: Samuel, Klopf, Werbös, in the 1960's and 70's
Barto, Sutton, Bertsekas from the 1990's-...
and many others



Drawing from Sutton and Barto, Reinforcement Learning: An Introduction, 1998.

Control Theory	**Reinforcement Learning**
Continuous or discrete setting	Discrete setting (Markov Decision Processes)
$\left.\begin{array}{c}\text{model}\\ +\\ \text{opt. criteria}\end{array}\right\} \longrightarrow \text{action}$	data $\longrightarrow$ action

Plan of the lecture:

1. General concepts and mathematical setting.
2. The value function and the Dynamic Programming Principle.
3. Value iteration method.
4. Linear Quadratic Regulator.

Control Theory

Dynamical system (discrete time)
Let $X \subset \mathbb{R}^d, \mathcal{U} \subset \mathbb{R}^p$ and $f : X \times \mathcal{U} \to X$

$$x_{t+1} = f(x_t, u_t)$$

- $x_0, x_1, x_2, \ldots$ are the states of the system. We have $x_t \in X$, for $t \geq 1$.
- $u_0, u_1, u_2, \ldots$ are the actions taken at each time (the policy). We have $u_t \in \mathcal{U}_t \subset \mathcal{U}$, for $t \geq 1$.

The next state depends on the current state and the action taken by the user (plus some random effects).

We define a **policy** π as a function which associates an action to any given history of the process

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

We will be interested on policies that only depend on the current state, i.e.

$$u_t = \pi(x_t)$$

Stochastic dynamical system (discrete time)
Let $X \subset \mathbb{R}^d, \mathcal{U} \subset \mathbb{R}^p$ and $f : X \times \mathcal{U} \times \mathcal{W} \to X$

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

- $x_0, x_1, x_2, \dots$ are the states of the system. We have $x_t \in X$, for $t \geq 1$.
- $u_0, u_1, u_2, \dots$ are the actions taken at each time (the policy). We have $u_t \in \mathcal{U}_t \subset \mathcal{U}$, for $t \geq 1$.
- $w_1, w_2, w_3, \dots$ are inputs that I cannot control (error measurement, noise effects ...). We will assume w_i is a stochastic process.

The next state depends on the current state and the action taken by the user (plus some random effects).

We define a **policy** π as a function which associates an action to any given history of the process

$$u_t = \pi_t(x_0, \dots, x_t, u_0, \dots, u_{t-1})$$

We will be interested on policies that only depend on the current state, i.e.

$$u_t = \pi(x_t)$$

Stochastic dynamical system (discrete time)
Let $X \subset \mathbb{R}^d, \mathcal{U} \subset \mathbb{R}^p$ and $f : X \times \mathcal{U} \times \mathcal{W} \to X$

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

- $x_0, x_1, x_2, \ldots$ are the states of the system. We have $x_t \in X$, for $t \geq 1$.
- $u_0, u_1, u_2, \ldots$ are the actions taken at each time (the policy). We have $u_t \in \mathcal{U}_t \subset \mathcal{U}$, for $t \geq 1$.
- $w_1, w_2, w_3, \ldots$ are inputs that I cannot control (error measurement, noise effects ...). We will assume w_i is a stochastic process.

The next state depends on the current state and the action taken by the user (plus some random effects).

We define a **policy** π as a function which associates an action to any given history of the process

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

We will be interested on policies that only depend on the current state, i.e.

$$u_t = \pi(x_t)$$

Stochastic dynamical system (discrete time)
Let $X \subset \mathbb{R}^d$, $\mathcal{U} \subset \mathbb{R}^p$ and $f : X \times \mathcal{U} \times \mathcal{W} \to X$

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

- $x_0, x_1, x_2, \ldots$ are the states of the system. We have $x_t \in X$, for $t \geq 1$.
- $u_0, u_1, u_2, \ldots$ are the actions taken at each time (the policy). We have $u_t \in \mathcal{U}_t \subset \mathcal{U}$, for $t \geq 1$.
- $w_1, w_2, w_3, \ldots$ are inputs that I cannot control (error measurement, noise effects ...). We will assume w_i is a stochastic process.

The next state depends on the current state and the action taken by the user (plus some random effects).

We define a **policy** π as a function which associates an action to any given history of the process

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

We will be interested on policies that only depend on the current state, i.e.

$$u_t = \pi(x_t)$$

Stochastic dynamical system (discrete time)
Let $X \subset \mathbb{R}^d, \mathcal{U} \subset \mathbb{R}^p$ and $f : X \times \mathcal{U} \times \mathcal{W} \to X$

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

- $x_0, x_1, x_2, \ldots$ are the states of the system. We have $x_t \in X$, for $t \geq 1$.
- $u_0, u_1, u_2, \ldots$ are the actions taken at each time (the policy). We have $u_t \in \mathcal{U}_t \subset \mathcal{U}$, for $t \geq 1$.
- $w_1, w_2, w_3, \ldots$ are inputs that I cannot control (error measurement, noise effects ...). We will assume w_i is a stochastic process.

The next state depends on the current state and the action taken by the user (plus some random effects).

We define a **policy** π as a function which associates an action to any given history of the process

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

We will be interested on policies that only depend on the current state, i.e.

$$u_t = \pi(x_t)$$

Markov Decision Process (MDP)
Let X and $\mathcal{U}$ be finite sets:

$$x_{t+1} \sim p(\cdot \mid x_t, u_t)$$

where for all $x, x' \in X$ and $u' \in \mathcal{U}$,

$$p(x \mid x', u') := \Pr\{X_{t+1} = x \mid X_t = x',\, U_t = u'\}.$$

For each $x', u' \in X \times \mathcal{U}$, the function

$$\begin{array}{rccc} p(\cdot|x', u') : & X & \longrightarrow & [0,1] \\ & x & \longmapsto & \Pr\{X_{t+1} = x \mid X_t = x',\, U_t = u'\} \end{array}$$

defines a probability distribution over the finite set X that determines the dynamics of the MDP.

The probability of the next state is a function of the current state and the action.

Main feature: The set of states X and of actions $\mathcal{U}$ are finite, so everything can be done using tables, rather than continuous functions as in the continuous setting.

Markov Decision Process (MDP)
Let X and $\mathcal{U}$ be finite sets:

$$x_{t+1} \sim p(\cdot \,|\, x_t, u_t)$$

where for all $x, x' \in X$ and $u' \in \mathcal{U}$,

$$p(x \,|\, x', u') := \Pr\{X_{t+1} = x \,|\, X_t = x',\, U_t = u'\}.$$

For each $x', u' \in X \times \mathcal{U}$, the function

$$\begin{array}{rccc} p(\cdot|x', u') : & X & \longrightarrow & [0,1] \\ & x & \longmapsto & \Pr\{X_{t+1} = x \,|\, X_t = x',\, U_t = u'\} \end{array}$$

defines a probability distribution over the finite set X that determines the dynamics of the MDP.

The probability of the next state is a function of the current state and the action.

Main feature: The set of states X and of actions $\mathcal{U}$ are finite, so everything can be done using tables, rather than continuous functions as in the continuous setting.

Markov Decision Process (MDP)
Let X and $\mathcal{U}$ be finite sets:

$$x_{t+1} \sim p(\cdot \mid x_t, u_t)$$

where for all $x, x' \in X$ and $u' \in \mathcal{U}$,

$$p(x \mid x', u') := \Pr\{X_{t+1} = x \mid X_t = x',\ U_t = u'\}.$$

For each $x', u' \in X \times \mathcal{U}$, the function

$$\begin{array}{rccc} p(\cdot|x', u') : & X & \longrightarrow & [0, 1] \\ & x & \longmapsto & \Pr\{X_{t+1} = x \mid X_t = x',\ U_t = u'\} \end{array}$$

defines a probability distribution over the finite set X that determines the dynamics of the MDP.

The probability of the next state is a function of the current state and the action.

Main feature: The set of states X and of actions $\mathcal{U}$ are finite, so everything can be done using tables, rather than continuous functions as in the continuous setting.

The time-horizon

- $T \in (0, \infty)$ is given (finite horizon), possibly with a terminal cost $g(x(T))$.
- T is a random stopping time, probably depending on x_t.
- T is infinite with $\gamma < 1$ (discounted cost).
- T is infinite with $\gamma \to 1^-$ (average cost).

$$\begin{aligned} \underset{\pi(\cdot)}{\text{minimize}}\ & \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t, u_t) + C_f(x_T) \right] \\ \text{s.t. } & x_{t+1} = f(x_t, u_t, w_t) \\ & x_0 = x \\ & u_t = \pi(\tau_t) \end{aligned}$$

$$\begin{aligned} \underset{\pi(\cdot)}{\text{minimize}}\ & \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right] \\ \text{s.t. } & x_{t+1} = f(x_t, u_t, w_t) \\ & x_0 = x \\ & u_t = \pi(\tau_t) \end{aligned}$$

Here x is the given initial state and $\tau_t = (x_0, \ldots, x_t, u_0, \ldots, u_{t-1})$ is the history of the process until time t.

The Optimal Control Problem

The time-horizon

- $T \in (0, \infty)$ is given (finite horizon), possibly with a terminal cost $C_f(\cdot)$.
- T is a random stopping time, probably depending on x_t.
- T is infinite with $\gamma < 1$ (discounted cost).
- T is infinite with $\gamma \to 1^-$ (average cost).
  (sometimes we can consider $\gamma = 1$)

$$\begin{aligned} \underset{\pi(\cdot)}{\text{minimize}}\ \mathbb{E}_w &\left[\sum_{t=0}^{T-1} C(x_t, u_t) + C_f(x_T) \right] \\ \text{s.t. } x_{t+1} &= f(x_t, u_t, w_t) \\ x_0 &= x \\ u_t &= \pi(\tau_t) \end{aligned}$$

$$\begin{aligned} \underset{\pi(\cdot)}{\text{minimize}}\ \mathbb{E}_w &\left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right] \\ \text{s.t. } x_{t+1} &= f(x_t, u_t, w_t) \\ x_0 &= x \\ u_t &= \pi(\tau_t) \end{aligned}$$

Here x is the given initial state and $\tau_t = (x_0, \dots, x_t, u_0, \dots, u_{t-1})$ is the history of the process until time t.

The value function

$$V^*(x,T) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t, u_t) + C_f(x_T) \right], \quad V^*(x) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right].$$

Bellman's Dynamic Programming (Bellman equation)

$$V^*(x,T) = \min_{u \in \mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + V^*(f(x,u,w_0), T-1) \right\}$$

$$V^*(x) = \min_{u \in \mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + \gamma V^*(f(x,u,w_0)) \right\}$$

Why is it good to have the value function?

Optimal feedback policy:

$$\pi_t(\tau_t) = \text{argmin}_{u \in \mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0), T-t) \right\}$$

$$\pi(\tau_t) = \text{argmin}_{u \in \mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0)) \right\}$$

The value function

$$V^*(x,T) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t,u_t) + C_f(x_T) \right], \quad V^*(x) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t,u_t) \right].$$

Bellman's Dynamic Programming (Bellman equation)

$$V^*(x,T) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + V^*(f(x,u,w_0), T-1) \right\}$$

$$V^*(x) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + \gamma V^*(f(x,u,w_0)) \right\}$$

Why is it good to have the value function?

Optimal feedback policy:

$$\pi_t(\tau_t) = \mathrm{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0), T-t) \right\}$$

$$\pi(\tau_t) = \mathrm{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0)) \right\}$$

The value function

$$V^*(x,T) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t,u_t) + C_f(x_T) \right], \quad V^*(x) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t,u_t) \right].$$

Bellman's Dynamic Programming (Bellman equation)

$$V^*(x,T) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + V^*(f(x,u,w_0), T-1) \right\}$$

$$V^*(x) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + \gamma V^*(f(x,u,w_0)) \right\}$$

Why is it good to have the value function?

Optimal feedback policy:

$$\pi_t(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0), T-t) \right\}$$

$$\pi(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0)) \right\}$$

The value function

$$V^*(x,T) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t,u_t) + C_f(x_T) \right], \quad V^*(x) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t,u_t) \right].$$

Bellman's Dynamic Programming (Bellman equation)

$$V^*(x,T) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + V^*(f(x,u,w_0), T-1) \right\}$$

$$V^*(x) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + \gamma V^*(f(x,u,w_0)) \right\}$$

Why is it good to have the value function?

Optimal feedback policy:

$$\pi_t(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0), T-t) \right\}$$

$$\pi(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0)) \right\}$$

The value function

$$V^*(x,T) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{T-1} C(x_t,u_t) + C_f(x_T) \right], \quad V^*(x) := \min_{\pi(\cdot)} \mathbb{E}_w \left[\sum_{t=0}^{\infty} \gamma^t C(x_t,u_t) \right].$$

Bellman's Dynamic Programming (Bellman equation)

$$V^*(x,T) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + V^*(f(x,u,w_0), T-1) \right\}$$

$$V^*(x) = \min_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x,u) + \gamma V^*(f(x,u,w_0)) \right\}$$

Why is it good to have the value function?

Optimal feedback policy:

$$\pi_t(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0), T-t) \right\}$$

$$\pi(\tau_t) = \text{argmin}_{u\in\mathcal{U}} \mathbb{E}_{w_0} \left\{ C(x_t,u) + \gamma V^*(f(x_t,u,w_0)) \right\}$$

Value iteration

Let us consider the **finite-horizon** problem with terminal cost. We recall the definition of the value function with $t \in [0, T]$ time-steps to go.

$$V^*(x,t) := \min_{\pi(\cdot)} \left[\sum_{s=T-t}^{T-1} C(x_s, u_s) + C_f(x_T) \right]$$

Recursive formula for the value function:

$$V^*(x,0) = C_f(x),$$

and for all $0 \leq t \leq T-1$

$$V^*(x,t) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), t-1) \right]$$

$$V^*(x,1) = \min_{u \in \mathcal{U}} \left[C(x,u) + C_f(f(x,u)) \right]$$

$$V^*(x,2) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), 1) \right]$$

$$\cdots$$

$$V^*(x,T) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), T-1) \right]$$

Let us consider the **finite-horizon** problem with terminal cost. We recall the definition of the value function with $t \in [0, T]$ time-steps to go.

$$V^*(x,t) := \min_{\pi(\cdot)} \left[\sum_{s=T-t}^{T-1} C(x_s, u_s) + C_f(x_T) \right]$$

Recursive formula for the value function:

$$V^*(x,0) = C_f(x),$$

and for all $0 \leq t \leq T-1$

$$V^*(x,t) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), t-1) \right]$$

$$V^*(x,1) = \min_{u \in \mathcal{U}} \left[C(x,u) + C_f(f(x,u)) \right]$$

$$V^*(x,2) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), 1) \right]$$

$$\cdots$$

$$V^*(x,T) = \min_{u \in \mathcal{U}} \left[C(x,u) + V^*(f(x,u), T-1) \right]$$

Value iteration

Let us consider the **finite-horizon** problem with terminal cost. We recall the definition of the value function with $t \in [0, T]$ time-steps to go.

$$V^*(x, t) := \min_{\pi(\cdot)} \left[\sum_{s=T-t}^{T-1} C(x_s, u_s) + C_f(x_T) \right]$$

Recursive formula for the value function:

$$V^*(x, 0) = C_f(x),$$

and for all $0 \leq t \leq T - 1$

$$V^*(x, t) = \min_{u \in \mathcal{U}} [C(x, u) + V^*(f(x, u), t - 1)]$$

$$V^*(x, 1) = \min_{u \in \mathcal{U}} [C(x, u) + C_f(f(x, u))]$$

$$V^*(x, 2) = \min_{u \in \mathcal{U}} [C(x, u) + V^*(f(x, u), 1)]$$

$$\cdots$$

$$V^*(x, T) = \min_{u \in \mathcal{U}} [C(x, u) + V^*(f(x, u), T - 1)]$$

Example in a finite setting (MDP)

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$

Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

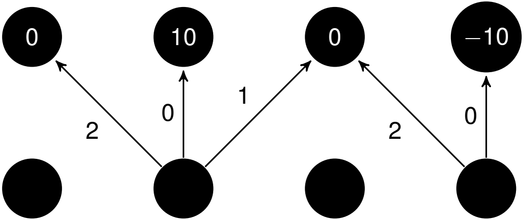
$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$

$C_f(\cdot) = V(\cdot, 0)$

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$
Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$



$C_f(\cdot) = V(\cdot, 0)$

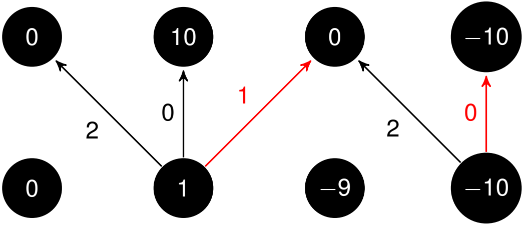
$V(\cdot, 1)$

Example in a finite setting (MDP)

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$
Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$



$C_f(\cdot) = V(\cdot, 0)$

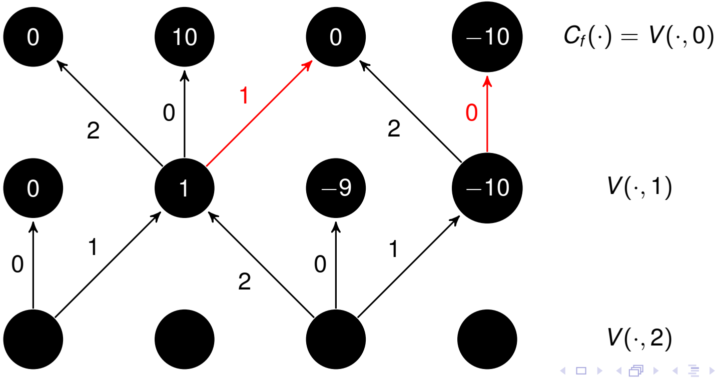
$V(\cdot, 1)$

Example in a finite setting (MDP)

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$
Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

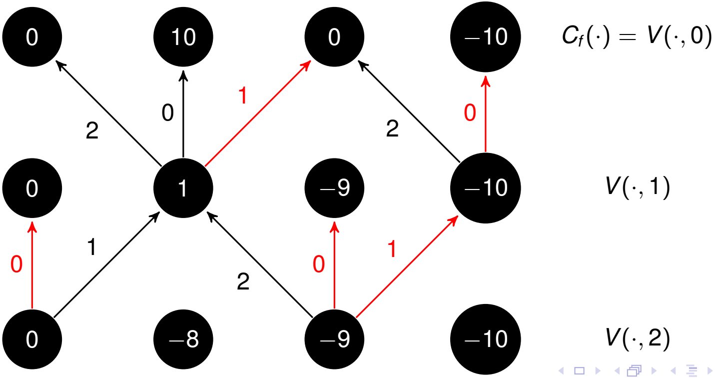
$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$



Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$
Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$

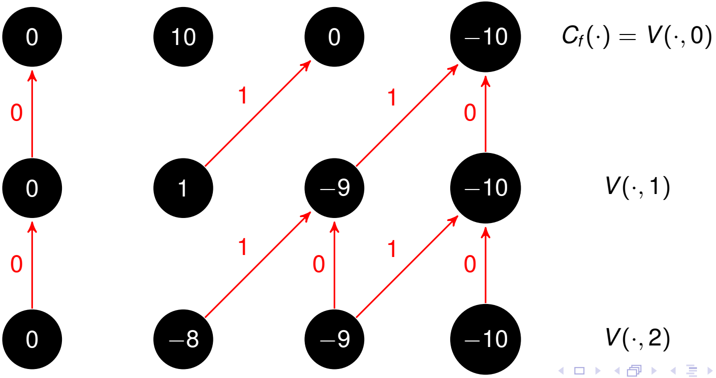


Example in a finite setting (MDP)

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}$ Set of possible actions: $\mathcal{U} = \{0, 1, -1\}$
Dynamics: $x_{t+1} = x_t + u_t$

Running and terminal cost:

$$C(u) := \begin{cases} 2 & u = -1 \\ 0 & u = 0 \\ 1 & u = 1 \end{cases} \qquad C_f(x) := \begin{cases} 0 & x = 1 \\ 10 & x = 2 \\ 0 & x = 3 \\ -10 & x = 4 \end{cases}$$



Let us consider the infinite-horizon problem with discounted factor $\gamma \in (0,1)$. Let X and $\mathcal{U}$ be the state space and the control space respectively (they can be continuous or discrete).

We recall the definition of the value function

$$V^*(x) := \min_{\pi(\cdot)} \left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right]$$

We look for a solution $V(\cdot)$ of the Bellman equation

$$V(x) = \min_{u \in \mathcal{U}} \{ C(x,u) + \gamma V(f(x,u)) \}$$

Definition

We define the **Bellman operator** $\mathcal{T} : L^\infty(X) \to L^\infty(X)$ as

$$\mathcal{T}V(x) := \min_{u \in \mathcal{U}} \{ C(x,u) + \gamma V(f(x,u)) \}, \qquad \text{for all } x \in X.$$

Value Iteration for infinite-horizon problems

Let us consider the infinite-horizon problem with discounted factor $\gamma \in (0, 1)$. Let X and $\mathcal{U}$ be the state space and the control space respectively (they can be continuous or discrete).

We recall the definition of the value function

$$V^*(x) := \min_{\pi(\cdot)} \left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right]$$

We look for a solution $V(\cdot)$ of the Bellman equation

$$V(x) = \min_{u \in \mathcal{U}} \{ C(x, u) + \gamma V(f(x, u)) \}$$

Definition

We define the **Bellman operator** $\mathcal{T} : L^\infty(X) \to L^\infty(X)$ as

$$\mathcal{T}V(x) := \min_{u \in \mathcal{U}} \{ C(x, u) + \gamma V(f(x, u)) \}, \qquad \text{for all } x \in X.$$

Value Iteration for infinite-horizon problems

Let us consider the infinite-horizon problem with discounted factor $\gamma \in (0, 1)$. Let X and $\mathcal{U}$ be the state space and the control space respectively (they can be continuous or discrete).

We recall the definition of the value function

$$V^*(x) := \min_{\pi(\cdot)} \left[\sum_{t=0}^{\infty} \gamma^t C(x_t, u_t) \right]$$

We look for a solution $V(\cdot)$ of the Bellman equation

$$V(x) = \min_{u \in \mathcal{U}} \{ C(x, u) + \gamma V(f(x, u)) \}$$

Definition

We define the **Bellman operator** $\mathcal{T} : L^\infty(X) \to L^\infty(X)$ as

$$\mathcal{T} V(x) := \min_{u \in \mathcal{U}} \{ C(x, u) + \gamma V(f(x, u)) \}, \qquad \text{for all } x \in X.$$

Value Iteration for infinite-horizon problems

Let $V, W : X \to \mathbb{R}$ be two function in $L^\infty(X)$.

$$\begin{aligned}
\mathcal{T}V(x) - \mathcal{T}W(x) &= \min_{u\in\mathcal{U}}\{C(x,u) + \gamma V(f(x,u))\} - \min_{w\in\mathcal{U}}\{C(x,w) + \gamma W(f(x,w))\} \\
&\leq C(x,w^*) + \gamma V(f(x,w^*)) - C(x,w^*) + \gamma W(f(x,w^*)) \\
&= \gamma \max_{x\in X}\{V(x) - W(x)\} \\
&\leq \gamma \|V(\cdot) - W(\cdot)\|_\infty .
\end{aligned}$$

Interchanging the roles of V and W we obtain that $\mathcal{T}$ satisfies the contraction property

$$\|\mathcal{T}V(\cdot) - \mathcal{T}W(\cdot)\|_\infty \leq \gamma \|V(\cdot) - W(\cdot)\|_\infty ,$$

where $\gamma \in (0,1)$ is the discount factor.

As a consequence of Banach's fix-point Theorem we have

$$V_k(\cdot) := \underbrace{\mathcal{T}\circ \cdots \circ \mathcal{T}}_{k \text{ times}} V(\cdot) \longrightarrow V^*(\cdot), \qquad \text{as } k\to\infty \text{ in } L^\infty(X),$$

where V^* is **the unique** fix point of the Bellman operator, i.e.

$$V^*(x) = \mathcal{T}V^*(x) = \min_{u\in\mathcal{U}}\left\{C(x,u) + \gamma V^*(f(x,u))\right\}, \qquad \text{for all } x\in X.$$

Let $V, W : X \to \mathbb{R}$ be two function in $L^\infty(X)$.

$$
\begin{aligned}
\mathcal{T}V(x) - \mathcal{T}W(x) &= \min_{u\in\mathcal{U}}\{C(x,u) + \gamma V(f(x,u))\} - \min_{w\in\mathcal{U}}\{C(x,w) + \gamma W(f(x,w))\} \\
&\leq C(x,w^*) + \gamma V(f(x,w^*)) - C(x,w^*) + \gamma W(f(x,w^*)) \\
&= \gamma \max_{x\in X}\{V(x) - W(x)\} \\
&\leq \gamma \|V(\cdot) - W(\cdot)\|_\infty .
\end{aligned}
$$

Interchanging the roles of V and W we obtain that $\mathcal{T}$ satisfies the contraction property

$$
\|\mathcal{T}V(\cdot) - \mathcal{T}W(\cdot)\|_\infty \leq \gamma \|V(\cdot) - W(\cdot)\|_\infty ,
$$

where $\gamma \in (0,1)$ is the discount factor.
As a consequence of Banach's fix-point Theorem we have

$$
V_k(\cdot) := \mathcal{T} \circ \underbrace{\cdots}_{k \text{ times}} \circ \mathcal{T}V(\cdot) \longrightarrow V^*(\cdot), \qquad \text{as } k \to \infty \text{ in } L^\infty(X),
$$

where V^* is **the unique** fix point of the Bellman operator, i.e.

$$
V^*(x) = \mathcal{T}V^*(x) = \min_{u\in\mathcal{U}} \{C(x,u) + \gamma V^*(f(x,u))\}, \qquad \text{for all } x \in X.
$$

Value iteration to approximate V^*

We initialize $V_0(x)$ arbitrarily (for instance $V_0(x) \equiv 0$).

For each x, we update the value function as follows:

$$V_{k+1}(x) = \min_{u \in \mathcal{U}} \{C(x,u) + \gamma V_k(f(x,u))\} .$$

- The discount factor ensures the convergence of the method with rate γ^k.
- **Remark:**

$$V_k(x) = \min_{u_1 \dots u_k} \left\{ \sum_{t=0}^{k-1} \gamma^t C(x_t, u_t) + \gamma^k V_0(x_k) \right\} .$$

  The function V_k is the value function of a finite-horizon problem with terminal cost $\gamma^k V_0(x)$.
- **Question:** Can we consider the non-discounted infinite-horizon problem? Under which conditions?

Value iteration to approximate V^*

We initialize $V_0(x)$ arbitrarily (for instance $V_0(x) \equiv 0$).

For each x, we update the value function as follows:

$$V_{k+1}(x) = \min_{u \in \mathcal{U}} \{C(x,u) + \gamma V_k(f(x,u))\}.$$

- The discount factor ensures the convergence of the method with rate γ^k.
- **Remark:**
$$V_k(x) = \min_{u_1 \dots u_k} \left\{ \sum_{t=0}^{k-1} \gamma^t C(x_t, u_t) + \gamma^k V_0(x_k) \right\}.$$
The function V_k is the value function of a finite-horizon problem with terminal cost $\gamma^k V_0(x)$.
- **Question:** Can we consider the non-discounted infinite-horizon problem? Under which conditions?

Example in finite setting

Value iteration to approximate V^*

We initialize $V_0(x)$ arbitrarily (for instance $V_0(x) \equiv 0$).

For each x, we update the value function as follows:

$$V_{k+1}(x) = \min_{u \in \mathcal{U}} \left\{ C(x,u) + \gamma V_k(f(x,u)) \right\}.$$

- The discount factor ensures the convergence of the method with rate γ^k.
- **Remark:**

$$V_k(x) = \min_{u_1 \ldots u_k} \left\{ \sum_{t=0}^{k-1} \gamma^t C(x_t, u_t) + \gamma^k V_0(x_k) \right\}.$$

  The function V_k is the value function of a finite-horizon problem with terminal cost $\gamma^k V_0(x)$.
- **Question:** Can we consider the non-discounted infinite-horizon problem? Under which conditions?

Example in finite setting

Value iteration to approximate V^*

We initialize $V_0(x)$ arbitrarily (for instance $V_0(x) \equiv 0$).

For each x, we update the value function as follows:

$$V_{k+1}(x) = \min_{u \in \mathcal{U}} \{C(x,u) + \gamma V_k(f(x,u))\} .$$

- The discount factor ensures the convergence of the method with rate γ^k.
- **Remark:**
$$V_k(x) = \min_{u_1 \ldots u_k} \left\{ \sum_{t=0}^{k-1} \gamma^t C(x_t, u_t) + \gamma^k V_0(x_k) \right\} .$$
The function V_k is the value function of a finite-horizon problem with terminal cost $\gamma^k V_0(x)$.
- **Question:** Can we consider the non-discounted infinite-horizon problem? Under which conditions?

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.9	1.9	0	0
1.9	1.9	0	0
0	0	4.0	−0.5
0	0	−0.5	−9.5

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.7	2.0	0	0
2.0	2.0	0	0
0	0	3.5	−4.5
0	0	−4.5	−14.0

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

3.4	2.0	0	0
2.0	2.0	0	−3.1
0	0	−0.095	−8.2
0	−3.1	−8.2	−17.0

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

3.8	2.0	0	−1.8
2.0	2.0	−1.8	−6.4
0	−1.8	−3.4	−11.0
−1.8	−6.4	−11.0	−20.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

3.8	2.0	−0.61	−4.7
2.0	0.39	−4.7	−9.3
−0.61	−4.7	−6.3	−14.0
−4.7	−9.3	−14.0	−23.0

Set of states: $X = \{1,2,3,4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

3.8	1.5	−3.3	−7.4
1.5	−2.3	−7.4	−12.0
−3.3	−7.4	−9.0	−17.0
−7.4	−12.0	−17.0	−26.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

3.3	−0.94	−5.7	−9.8
−0.94	−4.7	−9.8	−14.0
−5.7	−9.8	−11.0	−19.0
−9.8	−14.0	−19.0	−28.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.2	−3.1	−7.8	−12.0
−3.1	−6.8	−12.0	−17.0
−7.8	−12.0	−14.0	−22.0
−12.0	−17.0	−22.0	−31.0

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−0.78	−5.0	−9.7	−14.0
−5.0	−8.7	−14.0	−18.0
−9.7	−14.0	−15.0	−24.0
−14.0	−18.0	−24.0	−33.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−2.5	−6.8	−11.0	−16.0
−6.8	−10.0	−16.0	−20.0
−11.0	−16.0	−17.0	−25.0
−16.0	−20.0	−25.0	−34.0

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−4.1	−8.3	−13.0	−17.0
−8.3	−12.0	−17.0	−22.0
−13.0	−17.0	−19.0	−27.0
−17.0	−22.0	−27.0	−36.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−5.5	−9.8	−14.0	−19.0
−9.8	−13.0	−19.0	−23.0
−14.0	−19.0	−20.0	−28.0
−19.0	−23.0	−28.0	−37.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−6.8	−11.0	−16.0	−20.0
−11.0	−15.0	−20.0	−24.0
−16.0	−20.0	−21.0	−30.0
−20.0	−24.0	−30.0	−39.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−7.9	−12.0	−17.0	−21.0
−12.0	−16.0	−21.0	−26.0
−17.0	−21.0	−23.0	−31.0
−21.0	−26.0	−31.0	−40.0

Example in finite setting

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−8.9	−13.0	−18.0	−22.0
−13.0	−17.0	−22.0	−27.0
−18.0	−22.0	−24.0	−32.0
−22.0	−27.0	−32.0	−41.0

Set of states: $X = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.9$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

−9.9	−14.0	−19.0	−23.0
−14.0	−18.0	−23.0	−28.0
−19.0	−23.0	−25.0	−33.0
−23.0	−28.0	−33.0	−42.0

This is the approximation of the value function with a tolerance error of 1.

Example in finite setting

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	-5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Example in finite setting

Set of states: $\mathcal{S} = \{1,2,3,4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.6	1.6	0	0
1.6	1.6	0	0
0	0	4.0	1.0
0	0	1.0	−8.0

Example in finite setting

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	-5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

1.96	1.96	0	0
1.96	1.96	0	0
0	0	4.0	-0.8
0	0	-0.8	-9.8

Example in finite setting

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	-5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.18	2.0	0	0
2.0	2.0	0	0
0	0	3.52	-1.88
0	0	-1.88	-10.9

Example in finite setting

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.31	2.0	0	0
2.0	2.0	0	−0.128
0	0	2.87	−2.53
0	−0.128	−2.53	−11.5

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	-5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.38	2.0	0	0
2.0	2.0	0	-0.517
0	0	2.48	-2.92
0	-0.517	-2.92	-11.9

Example in finite setting

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.43	2.0	0	0
2.0	2.0	0	−0.75
0	0	2.25	−3.15
0	−0.75	−3.15	−12.2

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0, 0), \pm(1, 0), \pm(0, 1)\}$
Running cost: $C(x, u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.46	2.0	0	0
2.0	2.0	0	−0.89
0	0	2.11	−3.29
0	−0.89	−3.29	−12.3

Set of states: $\mathcal{S} = \{1, 2, 3, 4\}^2$
Set of possible action: $\mathcal{U} = \{(0,0), \pm(1,0), \pm(0,1)\}$
Running cost: $C(x,u) = c(x) + |u|$, where $c(x)$ is defined by the following table:

1.0	1.0	0	0
1.0	1.0	0	0
0	0	3.0	3.0
0	0	3.0	−5.0

Discount factor: $\gamma = 0.5$.

Value iteration

We initialize the value function $V_0(x) \equiv 0$, and then iterate using the Bellman operator.

2.47	2.0	0	0
2.0	2.0	0	−0.974
0	0	2.03	−3.37
0	−0.974	−3.37	−12.4

This is the approximation of the value function with a tolerance error of 0.1.

Example: Linear Quadratic Regulator

We consider the following **finite-time horizon** problem with quadratic final cost

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{T-1} (x_t^* Q x_t + u_t^* R u_t) + x_T^* P_0 x_T$$

$$\text{s.t. } x_{t+1} = A x_t + B u_t$$

$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

$$V(x,0) = x^* P_0 x$$

$$V(x,1) = \min_u \left[\underbrace{x^* Q x + u^* R u}_{C(x,u)} + \underbrace{(Ax+Bu)^* P_0 (Ax+Bu)}_{V(f(x,u))} \right]$$

$$\overline{u} = -(B^* P_t B + R)^{-1} B^* P_0 A x$$

$$V(x,1) = x^* \underbrace{\left(Q + A^* P_0 A - A^* P_0 B (B^* P_0 B + R)^{-1} B^* P_0 A \right)}_{P_1} x$$

We consider the following **finite-time horizon** problem with quadratic final cost

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{T-1} (x_t^* Q x_t + u_t^* R u_t) + x_T^* P_0 x_T$$
$$\text{s.t. } x_{t+1} = A x_t + B u_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

$$V(x, 0) = x^* P_0 x$$

$$V(x, 1) = \min_u \left[\underbrace{x^* Q x + u^* R u}_{C(x,u)} + \underbrace{(Ax + Bu)^* P_0 (Ax + Bu)}_{V(f(x,u))} \right]$$

$$\overline{u} = -(B^* P_t B + R)^{-1} B^* P_0 A x$$

$$V(x, 1) = x^* \underbrace{\left(Q + A^* P_0 A - A^* P_0 B (B^* P_0 B + R)^{-1} B^* P_0 A \right)}_{P_1} x$$

Example: Linear Quadratic Regulator

We consider the following **finite-time horizon** problem with quadratic final cost

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{T-1} (x_t^* Q x_t + u_t^* R u_t) + x_T^* P_0 x_T$$

$$\text{s.t. } x_{t+1} = A x_t + B u_t$$

$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

$$V(x,0) = x^* P_0 x$$

$$V(x,1) = \min_u \left[\underbrace{x^* Q x + u^* R u}_{C(x,u)} + \underbrace{(Ax + Bu)^* P_0 (Ax + Bu)}_{V(f(x,u))} \right]$$

$$\overline{u} = -(B^* P_t B + R)^{-1} B^* P_0 A x$$

$$V(x,1) = x^* \underbrace{\left(Q + A^* P_0 A - A^* P_0 B (B^* P_0 B + R)^{-1} B^* P_0 A \right)}_{P_1} x$$

Example: Linear Quadratic Regulator

We consider the following **finite-time horizon** problem with quadratic final cost

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{T-1}(x_t^* Q x_t + u_t^* R u_t) + x_T^* P_0 x_T$$

$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$

$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

$$V(x,0) = x^* P_0 x$$

$$V(x,1) = \min_u \left[\underbrace{x^* Q x + u^* R u}_{C(x,u)} + \underbrace{(Ax+Bu)^* P_0 (Ax+Bu)}_{V(f(x,u))} \right]$$

$$\overline{u} = -(B^* P_t B + R)^{-1} B^* P_0 A x$$

$$V(x,1) = x^* \underbrace{\left(Q + A^* P_0 A - A^* P_0 B (B^* P_0 B + R)^{-1} B^* P_0 A \right)}_{P_1} x$$

Example: Linear Quadratic Regulator

We consider the following **finite-time horizon** problem with quadratic final cost

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{T-1} (x_t^* Q x_t + u_t^* R u_t) + x_T^* P_0 x_T$$

$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$

$$x_0 = x$$

$$u_t = \pi(\tau_t)$$

Value Iteration

$$V(x, 0) = x^* P_0 x$$

$$V(x, t) = x^* P_t x$$

$$P_{t+1} = Q + A^* P_t A - A^* P_t B(B^* P_t B + R)^{-1} B^* P_t A$$

$$\pi_t^*(x_t) = \underbrace{-(B^* P_t B + R)^{-1} B^* P_{T-t} A}_{K_t} x_t$$

Example: Linear Quadratic Regulator

Infinite-horizon LQR: Let (A, B) be stabilizable, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$

$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$

$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

1. **Initialization:** $V_0(x) = 0$.
2. **Iterative procedure:**

$$V_{k+1}(x) = \min_u [x^* Qx + u^* Ru + V_k(x)] = x^* P_k x.$$

Observe that

$$V_k(x) = V(x, T), \qquad \text{with } T = k,$$

then

$$V(x) = \lim_{T\to\infty} V(x, T) \qquad \text{(if it exists)}$$

Infinite-horizon LQR: Let (A, B) be stabilizable, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

1. **Initialization:** $V_0(x) = 0$.
2. **Iterative procedure:**

$$V_{k+1}(x) = \min_u \left[x^* Q x + u^* R u + V_k(x)\right] = x^* P_k x.$$

Observe that

$$V_k(x) = V(x, T), \qquad \text{with } T = k,$$

then

$$V(x) = \lim_{T\to\infty} V(x, T) \qquad \text{(if it exists)}$$

Example: Linear Quadratic Regulator

Infinite-horizon LQR: Let (A, B) be stabilizable, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

1. **Initialization:** $V_0(x) = 0$.
2. **Iterative procedure:**

$$V_{k+1}(x) = \min_u [x^* Q x + u^* R u + V_k(x)] = x^* P_k x.$$

Observe that

$$V_k(x) = V(x, T), \qquad \text{with } T = k,$$

then

$$V(x) = \lim_{T \to \infty} V(x, T) \qquad \text{(if it exists)}$$

Infinite-horizon LQR: Let (A, B) be stabilizable, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

1. **Initialization:** $V_0(x) = 0$.
2. **Iterative procedure:**

$$V_{k+1}(x) = \min_u \left[x^* Q x + u^* R u + V_k(x)\right] = x^* P_k x.$$

Observe that

$$V_k(x) = V(x, T), \qquad \text{with } T = k,$$

then

$$V(x) = \lim_{T \to \infty} V(x, T) \qquad \text{(if it exists)}$$

Infinite-horizon LQR: Let (A, B) be stabilizable, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Value Iteration

1. **Initialization:** $V_0(x) = 0$.
2. **Iterative procedure:**

$$V_{k+1}(x) = \min_u \left[x^* Q x + u^* R u + V_k(x)\right] = x^* P_k x.$$

Observe that

$$V_k(x) = V(x, T), \qquad \text{with } T = k,$$

then

$$V(x) = \lim_{T \to \infty} V(x, T) \qquad \text{(if it exists)}$$

Example: Linear Quadratic Regulator

Infinite-horizon LQR: Let (A, B) be stabilizable and Q, R positive definite matrices, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = A x_t + B u_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Long-time behavior for $V(x, T)$

In [E.-Kouhkouh-Pighin-Zuazua, 2020], it is proved (in the cont. setting) that

$$V(x, T) - V_s T \to W(x) + \lambda, \qquad \text{as } T \to \infty,$$

where

$$V_s = \min\{x^* Q x + u^* R u \ : \ (x, u) \text{ s.t. } Ax + Bu = 0\} = 0,$$

and $W(x) = x^* P x$, with P the unique pos. def. sol. to DARE:

$$P = Q + A^* P A - A^* P B (B^* P B + R)^{-1} B^* P A.$$

Question: Is it possible to extend this to more general cases?

Example: Linear Quadratic Regulator

Infinite-horizon LQR: Let (A, B) be stabilizable and Q, R positive definite matrices, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Long-time behavior for $V(x, T)$

In [E.-Kouhkouh-Pighin-Zuazua, 2020], it is proved (in the cont. setting) that

$$V(x, T) - V_s T \to W(x) + \lambda, \qquad \text{as } T \to \infty,$$

If $V_s \neq 0$, we can consider a modified cost functional

$$\tilde{C}(x, u) = C(x, u) - V_s,$$

and then

$$\tilde{V}(x, T) \to W(x) + \lambda, \qquad \text{as } T \to \infty.$$

Question: Is it possible to extend this to more general cases?

Example: Linear Quadratic Regulator

Infinite-horizon LQR: Let (A, B) be stabilizable and Q, R positive definite matrices, NO discount factor

$$\underset{\pi(\cdot)}{\text{minimize}} \sum_{t=0}^{\infty} (x_t^* Q x_t + u_t^* R u_t)$$
$$\text{s.t. } x_{t+1} = A x_t + B u_t$$
$$x_0 = x, \quad u_t = \pi(\tau_t)$$

Long-time behavior for $V(x, T)$

In [E.-Kouhkouh-Pighin-Zuazua, 2020], it is proved (in the cont. setting) that

$$V(x, T) - V_s T \to W(x) + \lambda, \qquad \text{as } T \to \infty,$$

If $V_s \neq 0$, we can consider a modified cost functional

$$\tilde{C}(x, u) = C(x, u) - V_s,$$

and then

$$\tilde{V}(x, T) \to W(x) + \lambda, \qquad \text{as } T \to \infty.$$

Question: Is it possible to extend this to more general cases?