Week 1. Discrete Optimal Control and Dynamic Programming

Course

Hamilton–Jacobi Equations and Optimal Control

Primary References

- Hung Vinh Tran, *Hamilton–Jacobi Equations: Theory and Applications*, AMS GSM 213, 2021.
- Course lecture outline and syllabus.

1. Introduction

These notes serve as a conceptual prelude to the main themes of the course. The central objects of the semester are Hamilton–Jacobi equations, value functions arising from optimal control, and weak solution concepts—especially viscosity solutions. However, if one starts immediately with continuous-time control systems and nonlinear first-order PDEs, the guiding intuition can remain opaque. For that reason, we begin in discrete time, where every essential idea already appears in a simpler and more transparent form.

The aim of this first lecture is not to develop a separate theory for its own sake, but to isolate the recurring structural principle of optimal control:

> **The optimal cost from the present state is obtained by optimizing the immediate cost plus the optimal future cost.**

In discrete time this principle becomes the **Bellman equation**. In continuous time it becomes the **Dynamic Programming Principle (DPP)**, and in the infinitesimal limit it leads formally to the **Hamilton–Jacobi–Bellman (HJB) equation**.

A second purpose of this lecture is to make clear that value functions are typically **not smooth**. Even in the simplest examples, the value function may have corners or kinks. This is the first indication that classical C^1 solution theory is not sufficient for Hamilton–Jacobi equations, and that a more robust notion of solution will eventually be required.

2. Learning Objectives

By the end of this lecture, students should be able to:

1. formulate a discrete-time optimal control problem;
2. define and interpret the value function;
3. derive the Bellman equation;
4. understand the Bellman equation as a fixed-point equation;
5. solve basic examples explicitly;
6. recognize the natural emergence of non-smooth value functions;
7. see how discrete dynamic programming leads to continuous-time HJB equations.

3. The Basic Discrete-Time Control Model

Let X be the state space. In our examples, X will be one of the following:

- $\mathbb{Z}$,
- $\mathbb{Z}^2$,
- $\mathbb{R}$.

Let V be the control set. At each discrete time $k = 0, 1, 2, ...$, the controller chooses an action $v_k \in V$, and the state evolves according to the transition rule

$$y_{k+1} = B(y_k, v_k), \qquad y_0 = x.$$

Here:

- y_k is the state at time k,
- v_k is the control chosen at time k,
- $B : X \times V \to X$ is the one-step state transition map.

A control sequence is an infinite sequence

$$\mathbf{v} = (v_0, v_1, v_2, \dots), \qquad v_k \in V.$$

Once the initial state x and the control sequence $\mathbf{v}$ are fixed, the full trajectory $(y_k)_{k\geq 0}$ is determined recursively.

This is the discrete analogue of a controlled ODE. If one imagines a time step $\Delta t > 0$, then the forward Euler discretization of

$$y'(t) = b(y(t), v(t))$$

takes the form

$$y_{k+1} \approx y_k + \Delta t\, b(y_k, v_k),$$

which fits exactly into the discrete framework above.

4. Running Cost, Total Cost, and Value Function

Let

$$f : X \times V \to \mathbb{R}$$

be the running cost. The quantity $f(y_k, v_k)$ represents the cost incurred at time k when the state is y_k and the control is v_k.

For Sections 4–8, assume that f is bounded,

$$|f(x, v)| \leq M \qquad \text{for all } (x, v) \in X \times V,$$

and fix a discount factor $\beta \in (0, 1)$. Define the discounted infinite-horizon cost by

$$J(x, \mathbf{v}) = \sum_{k=0}^{\infty} \beta^k f(y_k, v_k).$$

The corresponding value function is

$$u(x) = \inf_{\mathbf{v}} J(x, \mathbf{v}).$$

Thus $u(x)$ is the infimal total cost when the system starts from the state x.

Then

$$|J(x, \mathbf{v})| \leq \sum_{k=0}^{\infty} \beta^k M = \frac{M}{1-\beta}.$$

Hence the total cost is well-defined and finite for every control sequence.

The value function is the mathematical encoding of optimal decision-making. It records, for every state x, the best achievable performance from that state onward. In this sense, once the value function is known, the control problem is essentially solved.

5. The Bellman Equation

We now derive the fundamental recursive identity satisfied by the value function.

Proposition 5.1 (Discrete Dynamic Programming Principle)

For every state $x \in X$,

$$\boxed{u(x) = \inf_{v \in V} \{ f(x,v) + \beta\, u(B(x,v)) \}\,.}$$

Proof

Take any control sequence

$$\mathbf{v} = (v_0, v_1, v_2, \dots),$$

and let the corresponding trajectory satisfy

$$y_0 = x, \qquad y_{k+1} = B(y_k, v_k).$$

Then

$$J(x, \mathbf{v}) = f(x, v_0) + \beta f(y_1, v_1) + \beta^2 f(y_2, v_2) + \cdots.$$

If we define the tail control sequence

$$\mathbf{v}' = (v_1, v_2, v_3, \dots),$$

then the total cost can be rewritten as

$$J(x, \mathbf{v}) = f(x, v_0) + \beta J(y_1, \mathbf{v}'), \qquad y_1 = B(x, v_0).$$

Since $u(y_1) \le J(y_1, \mathbf{v}')$, we obtain

$$J(x, \mathbf{v}) \ge f(x, v_0) + \beta u(y_1) = f(x, v_0) + \beta u(B(x, v_0)).$$

This holds for every control sequence $\mathbf{v}$, so taking the infimum over all such sequences gives

$$u(x) \ge \inf_{v \in V} \{ f(x,v) + \beta u(B(x,v)) \}\,.$$

For the reverse inequality, fix $v \in V$. After taking control v at the first step, let the controller act ε-optimally from the next state $B(x,v)$ onward. Then

$$u(x) \le f(x,v) + \beta u(B(x,v)) + \beta\varepsilon.$$

Since v was arbitrary, we may take the infimum over $v \in V$, obtaining

$$u(x) \le \inf_{v \in V} \{ f(x,v) + \beta u(B(x,v)) \} + \beta\varepsilon.$$

Letting $\varepsilon \to 0$ yields the result.

□

6. Interpretation of the Bellman Equation

The Bellman equation should be read as follows:

> The infimal cost from the current state equals the infimum, over all immediate control choices, of the current running cost plus the discounted continuation value.

This is the core principle of dynamic programming. The optimization problem is recursive: once the first decision is made, the remaining problem is again an optimal control problem of the same type.

Thus the Bellman equation is not merely an algebraic identity. It expresses the recursive structure of optimality; the proof only requires ε-optimal controls and does not assume that the infimum is attained.

7. The Bellman Operator and Fixed-Point Theory

Define the Bellman operator T by

$$(T\phi)(x) = \inf_{v \in V} \{f(x,v) + \beta\phi(B(x,v))\}.$$

Then the Bellman equation becomes

$$u = Tu.$$

So the value function is a fixed point of a nonlinear operator.

Proposition 7.1

Assume $\beta \in (0,1)$. Then T is a contraction on the space of bounded functions on X, equipped with the supremum norm:

$$\|T\phi - T\psi\|_\infty \leq \beta\|\phi - \psi\|_\infty.$$

Proof

Fix $x \in X$. For any $v \in V$,

$$f(x,v) + \beta\phi(B(x,v)) \leq f(x,v) + \beta\psi(B(x,v)) + \beta\|\phi - \psi\|_\infty.$$

Taking the infimum over $v \in V$, we obtain

$$(T\phi)(x) \leq (T\psi)(x) + \beta\|\phi - \psi\|_\infty.$$

By symmetry,

$$(T\psi)(x) \leq (T\phi)(x) + \beta\|\phi - \psi\|_\infty.$$

Hence

$$|(T\phi)(x) - (T\psi)(x)| \leq \beta\|\phi - \psi\|_\infty.$$

Taking the supremum over $x \in X$ gives the claim.

□

Corollary 7.2

There exists a unique bounded fixed point u of T. Therefore the Bellman equation has a unique bounded solution.

8. Value Iteration

Starting from any bounded initial guess u_0, define

$$u_{n+1} = Tu_n.$$

Since T is a contraction,

$$\|u_n - u\|_\infty \le \beta^n \|u_0 - u\|_\infty,$$

where u is the unique fixed point. Therefore $u_n \to u$ uniformly.

This iterative scheme is called **value iteration**. It shows that repeated local optimization produces the global value function. In this sense, dynamic programming is not only a theoretical principle but also a computational method.

9. Example 1: A One-Dimensional Shortest-Path Problem

Let the state space be $X = \mathbb{Z}$, and let the target be the origin 0. At each step the controller may move one unit to the left or one unit to the right:

$$V = \{-1, +1\}, \qquad y_{k+1} = y_k + v_k.$$

The cost is one per step until the target is reached. This is an undiscounted stopped minimum-time problem, distinct from the discounted infinite-horizon problem of Sections 4–8.

Define the hitting time

$$\tau_x(\mathbf{v}) = \inf\{k \ge 0 : y_k = 0\}.$$

The cost of a control sequence is

$$J(x, \mathbf{v}) = \tau_x(\mathbf{v}),$$

and the value function is

$$u(x) = \inf_{\mathbf{v}} \tau_x(\mathbf{v}).$$

The Bellman equation becomes

$$u(0) = 0, \qquad u(x) = 1 + \min\{u(x-1), u(x+1)\} \quad \text{for } x \ne 0.$$

Every admissible move changes the state by one, so $\tau_x(\mathbf{v}) \ge |x|$. Moving toward the origin at each step attains this lower bound. Therefore

$$u(x) = |x|.$$

The formula also satisfies the Bellman equation. If $x > 0$, then

$$\min\{|x-1|, |x+1|\} = x - 1,$$

so

$$1 + \min\{|x-1|, |x+1|\} = 1 + (x-1) = x = |x|.$$

If $x < 0$, then

$$\min\{|x-1|, |x+1|\} = |x| - 1,$$

so again

$$1 + \min\{|x-1|, |x+1|\} = |x|.$$

Since $u(0) = 0$, the formula is proved.

Thus the value function is

$$\boxed{u(x) = |x|.}$$

Interpretation

This example is extremely important. The natural piecewise-linear extension of the grid value function to $\mathbb{R}$ is explicit and geometrically natural, yet it is **not differentiable at the origin**. Hence non-smoothness is not an artificial complication. It arises naturally from optimization.

10. Why Example 1 Matters

There are several lessons in the formula $u(x) = |x|$.

First, the value function reflects geometry: it is exactly the graph distance to the target.

Second, the Bellman equation is local, but the value function is global. The equation only compares neighboring states, while the solution encodes the minimal travel time from every state.

Third, the non-differentiability of the natural extension at the origin is not accidental. It reflects the change of optimal direction across the target. Changes in optimal behavior can manifest analytically as loss of smoothness.

This is the first concrete indication that later Hamilton–Jacobi equations will require a weak solution concept.

11. Example 2: A Discrete LQ-Type Problem

Let

$$V = [-1, 1], \qquad y_{k+1} = y_k + v_k, \qquad f(y, v) = \frac{1}{2}y^2, \qquad \beta \in (0, 1).$$

Then the Bellman equation is

$$u(x) = \min_{v \in [-1,1]} \left\{ \frac{1}{2}x^2 + \beta u(x + v) \right\}.$$

This is not the classical linear-quadratic regulator in the strict sense, since there is no quadratic control cost and the control set is bounded. Nevertheless, it has an LQ-type structure and is pedagogically useful.

For any admissible control, the bound $|v_k| \le 1$ gives

$$|y_k| \ge (|x| - k)_+.$$

The feedback below attains equality at every step. Since all discount weights are positive and $r \mapsto r^2/2$ is increasing for $r \ge 0$, it is optimal:

$$v^*(x) = -\operatorname{sgn}(x) \min\{|x|, 1\}.$$

Indeed:

- if $|x| \le 1$, one can choose $v = -x$ and move directly to the origin;
- if $|x| > 1$, one can reduce $|x|$ by at most one per step, so the best choice is to push maximally toward zero.

If $|x| \le 1$, then the next state can be made equal to zero, so

$$u(x) = \frac{1}{2}x^2.$$

If $|x| > 1$, let

$$m = \lfloor |x| \rfloor, \qquad r = |x| - m \in [0, 1).$$

Under the optimal strategy, the magnitudes visited are

$$|x|,\ |x|-1,\ |x|-2,\ \ldots,\ |x|-(m-1),\ r,\ 0,\ 0, \ldots$$

and hence

$$u(x) = \frac{1}{2} \sum_{j=0}^{m-1} \beta^j (|x| - j)^2 + \frac{1}{2} \beta^m r^2.$$

This example shows how the Bellman equation naturally generates a **feedback law** and how the value function may be given by different formulas in different regions of the state space.

12. Example 3: Path Planning on $\mathbb{Z}^2$

Let

$$X = \mathbb{Z}^2, \qquad V = \{(\pm 1, 0), (0, \pm 1)\}.$$

At each step the controller may move one unit horizontally or vertically. The target is the origin $(0, 0)$, and the cost is one per step until the target is reached.

The Bellman equation becomes

$$u(0, 0) = 0,$$

and for $(x_1, x_2) \neq (0, 0)$,

$$u(x_1, x_2) = 1 + \min\Big\{u(x_1 + 1, x_2), u(x_1 - 1, x_2), u(x_1, x_2 + 1), u(x_1, x_2 - 1)\Big\}.$$

Every path to the origin has at least $|x_1| + |x_2|$ steps, because one move changes only one coordinate by one unit. Moving a nonzero coordinate toward zero at each step attains this lower bound. Hence

$$u(x_1, x_2) = |x_1| + |x_2|.$$

To verify this, note that as long as $(x_1, x_2) \neq (0, 0)$, at least one coordinate is nonzero. Moving one step toward zero in a nonzero coordinate decreases $|x_1| + |x_2|$ by exactly one. Therefore the smallest possible future value is

$$|x_1| + |x_2| - 1,$$

and so

$$1 + \min(\text{future value}) = |x_1| + |x_2|.$$

Hence

$$\boxed{u(x_1, x_2) = |x_1| + |x_2|.}$$

This is the Manhattan distance to the origin.

Interpretation

Once again, the natural piecewise-linear extension of the grid value function is not smooth. Nondifferentiability appears along the coordinate axes and at the origin.

The geometry of this example is summarized below. Each admissible move changes exactly one coordinate by one unit, so every shortest path from (x_1, x_2) to the origin has length $|x_1| + |x_2|$. The diamond-shaped level sets are therefore not merely a plotting convention: they are the metric balls of the control system.

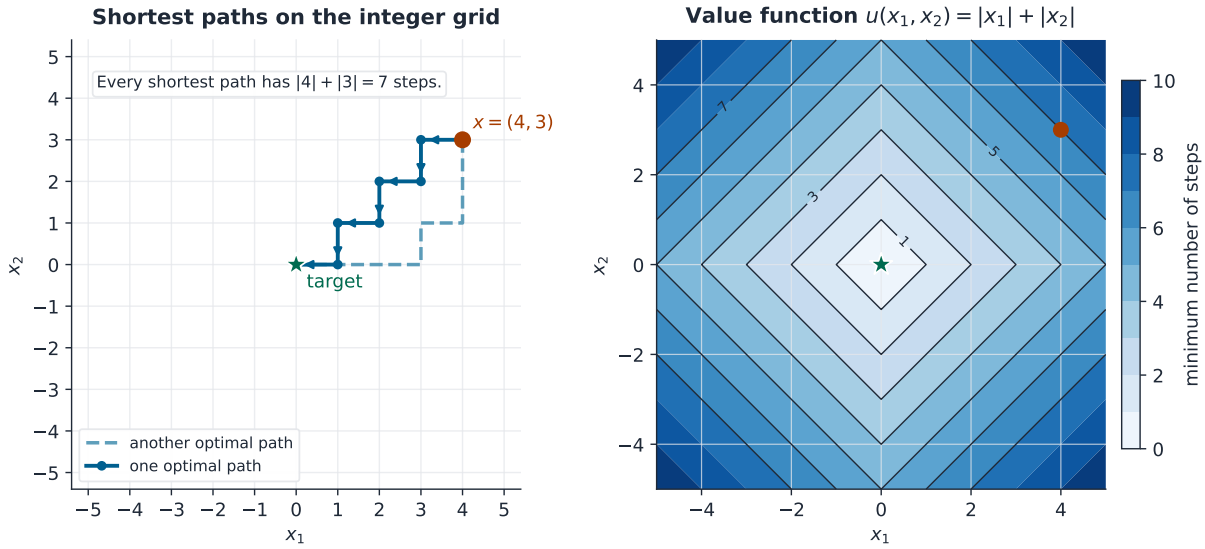


Figure 1.1. Left: two of the many optimal paths from $(4, 3)$ to the target. Right: level sets of the value function $u(x_1, x_2) = |x_1| + |x_2|$. The Bellman equation is local, but its solution records the global shortest-path geometry.

13. Feedback Interpretation

The Bellman equation does more than identify the value function. It also suggests the optimal control at each state.

If u is known and the infimum is attained at each state, an optimal feedback map $F(x)$ may be chosen to satisfy

$$F(x) \in \operatorname*{argmin}_{v \in V} \{f(x, v) + \beta u(B(x, v))\}.$$

Thus the value function contains not only optimal cost information but also optimal policy information. This is one of the main reasons why the value function is such a central object.

Later in the course, the same idea will reappear in continuous time through the Hamiltonian

$$H(x, p) = \sup_{v \in V} \{-b(x, v) \cdot p - f(x, v)\},$$

where the optimizing control plays the role of an infinitesimal feedback law dictated by the gradient $p = Du(x)$.

14. From Discrete Bellman to Continuous Dynamic Programming

We now compare the discrete setting with the continuous-time optimal control framework that will be introduced in the next lectures.

In continuous time, one considers:

- a compact control set V,
- measurable controls $v(\cdot) : [0, \infty) \to V$,
- a discount rate $\lambda > 0$,
- continuous, bounded functions b and f that are uniformly Lipschitz in x, uniformly in v,

- a controlled ODE

$$y'(s) = b(y(s), v(s)), \qquad y(0) = x,$$

- a discounted cost functional

$$J(x, v(\cdot)) = \int_0^\infty e^{-\lambda s} f(y(s), v(s))\, ds,$$

- and the value function

$$u(x) = \inf_{v(\cdot)} J(x, v(\cdot)).$$

The corresponding Dynamic Programming Principle is

$$u(x) = \inf_{v(\cdot)} \left\{ \int_0^t e^{-\lambda s} f(y(s), v(s))\, ds + e^{-\lambda t} u(y(t)) \right\}.$$

This is the continuous-time analogue of the Bellman equation. In discrete time one step of cost is extracted; in continuous time a short time interval is extracted.

Thus the continuous DPP should be understood as the same recursive optimality principle in a time-continuous form.

15. Formal Derivation of the HJB Equation

We now show, at a heuristic level, how the Bellman equation leads to the stationary Hamilton–Jacobi–Bellman equation.

Let $\Delta t > 0$ be a small time step, and consider the discrete approximation

$$B_{\Delta t}(x, v) = x + \Delta t\, b(x, v).$$

To match the continuous-time running cost, we scale the one-step cost as

$$\Delta t\, f(x, v),$$

and choose the discount factor

$$\beta = e^{-\lambda \Delta t}.$$

Then the Bellman equation becomes

$$u_{\Delta t}(x) = \inf_{v \in V} \left\{ \Delta t\, f(x, v) + e^{-\lambda \Delta t} u_{\Delta t}(x + \Delta t\, b(x, v)) \right\}.$$

To check formal consistency, apply the one-step scheme to a smooth candidate u. This calculation is heuristic; convergence $u_{\Delta t} \to u$ alone would not justify replacing $u_{\Delta t}$ by u with an $o(\Delta t)$ error. We use

$$e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t),$$

and

$$u(x + \Delta t\, b(x, v)) = u(x) + \Delta t\, b(x, v) \cdot Du(x) + o(\Delta t).$$

Substituting these expansions into the Bellman equation gives

$$u(x) = \inf_{v \in V} \left\{ \Delta t\, f(x, v) + (1 - \lambda \Delta t)(u(x) + \Delta t\, b(x, v) \cdot Du(x)) \right\} + o(\Delta t).$$

Subtract $u(x)$ from both sides, divide by Δt, and let $\Delta t \to 0$. One obtains

$$0 = \inf_{v \in V} \{ f(x,v) - \lambda u(x) + b(x,v) \cdot Du(x) \}.$$

Multiplying by -1, this becomes

$$\lambda u(x) + \sup_{v \in V} \{ -b(x,v) \cdot Du(x) - f(x,v) \} = 0.$$

Define the Hamiltonian by

$$H(x,p) = \sup_{v \in V} \{ -b(x,v) \cdot p - f(x,v) \}.$$

Then the equation becomes

$$\boxed{\lambda u(x) + H(x, Du(x)) = 0.}$$

This is the stationary Hamilton–Jacobi–Bellman equation.

The same scaling also appears directly at the level of the dynamics. For a fixed control, the update

$$y_{k+1} = y_k + \Delta t \, b(y_k, v_k)$$

is an Euler approximation of the controlled ODE. As the time step decreases, the polygonal discrete trajectory approaches the continuous trajectory. The following example uses the rotation field $b(y_1, y_2) = (-y_2, y_1)$; it is not an optimal-control computation, but it isolates the discrete-to-continuous limit underlying the formal Bellman expansion.

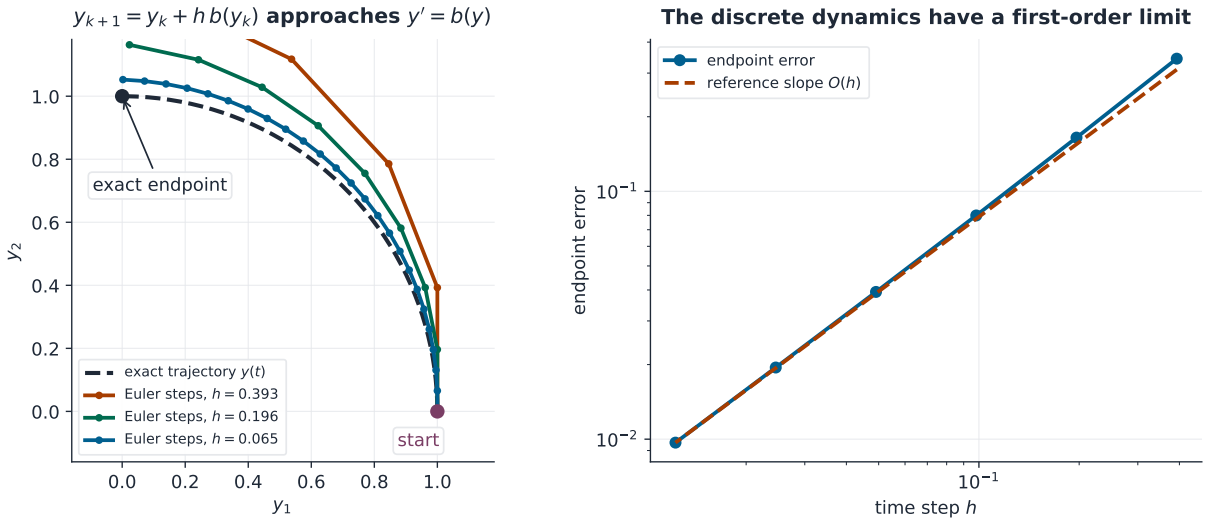


Figure 1.2. Left: explicit Euler trajectories for three time steps approach the exact circular arc. Right: the endpoint error decreases at first order in the time step. This is the dynamical counterpart of passing from the discrete Bellman equation to the continuous HJB equation.

16. Why Smoothness Fails

The formal derivation above assumes differentiability of the limiting value function. But our discrete examples already show that differentiability is generally too much to expect.

The natural piecewise-linear extensions of the discrete value functions

$$u(x) = |x| \qquad \text{and} \qquad u(x_1, x_2) = |x_1| + |x_2|$$

agree with the grid values, yet they are not smooth.

This is not merely a peculiarity of the discrete setting. In continuous first-order PDEs, similar phenomena arise in eikonal equations and more general Hamilton–Jacobi equations. Classical C^1 solutions may fail to exist, even in very simple situations.

Therefore the following conceptual chain should be kept in mind:

$$\text{optimal control} \longrightarrow \text{value function} \longrightarrow \text{dynamic programming} \longrightarrow \text{HJB equation} \longrightarrow \text{need for viscosity solutions.}$$

This is the main roadmap of the course.

17. Key Takeaways

The main points of this lecture are the following.

1. A discrete-time optimal control problem is defined by a state space, a control set, a transition rule, and a running cost.
2. The value function records the infimal total cost achievable from each initial state.
3. The value function satisfies the Bellman equation.
4. The Bellman equation is a nonlinear fixed-point equation.
5. In the discounted setting, the Bellman operator is a contraction, so the value function is uniquely determined.
6. Explicit examples show that value functions are often non-smooth.
7. The discrete Bellman equation is the natural precursor of the continuous-time Dynamic Programming Principle and the Hamilton–Jacobi–Bellman equation.

18. Exercises

Exercise 1.1

For the one-dimensional shortest-path problem on $\mathbb{Z}$, prove carefully that $u(x) = |x|$ satisfies

$$u(0) = 0, \qquad u(x) = 1 + \min\{u(x-1), u(x+1)\} \quad (x \neq 0).$$

Exercise 1.2

For the two-dimensional path-planning problem on $\mathbb{Z}^2$, verify rigorously that

$$u(x_1, x_2) = |x_1| + |x_2|$$

satisfies the Bellman equation.

Exercise 1.3

In the discrete LQ-type example, derive the formula

$$u(x) = \frac{1}{2} \sum_{j=0}^{m-1} \beta^j (|x| - j)^2 + \frac{1}{2} \beta^m r^2,$$

where $m = \lfloor |x| \rfloor$ and $r = |x| - m$.

Exercise 1.4

Let

$$(T\phi)(x) = \inf_{v \in V} \{ f(x, v) + \beta \phi(B(x, v)) \}.$$

Show that T is monotone: if $\phi \leq \psi$, then $T\phi \leq T\psi$.

Exercise 1.5

Starting from

$$u_{\Delta t}(x) = \inf_{v \in V} \left\{ \Delta t\, f(x,v) + e^{-\lambda \Delta t} u_{\Delta t}(x + \Delta t\, b(x,v)) \right\},$$

carry out the formal derivation of

$$\lambda u + H(x, Du) = 0.$$

Exercise 1.6

Compare the function $u(x) = |x|$ with the continuous eikonal intuition $|Du| = 1$. Explain informally why one should expect similar cusp structures in continuous shortest-path or minimum-time problems.

19. Preview of Week 2

In the next lecture we replace the discrete transition rule by a controlled ODE,

$$y'(s) = b(y(s), v(s)), \qquad y(0) = x,$$

where the control $v(\cdot)$ is a measurable function taking values in a compact control set V.

We then define the discounted continuous-time cost

$$J(x, v(\cdot)) = \int_0^\infty e^{-\lambda s} f(y(s), v(s))\, ds$$

and the corresponding value function

$$u(x) = \inf_{v(\cdot)} J(x, v(\cdot)).$$

This will place us in the standard framework of continuous-time optimal control and prepare the rigorous formulation of the Dynamic Programming Principle and the Hamilton–Jacobi–Bellman equation.