Week 3. Dynamic Programming Principle in Continuous Time

Primary reference: Hung Vinh Tran, *Hamilton–Jacobi Equations* §2.2, Theorem 2.6

1. Overview

In this course, we do not follow the textbook in a strictly linear order.
Instead, we first formulate continuous optimal control problems and value functions, and then introduce the **Dynamic Programming Principle (DPP)** as the fundamental bridge from control theory to Hamilton–Jacobi equations.

The main goal of this lecture is to understand why the value function satisfies a recursive relation in time. This recursive structure is the continuous-time analogue of the Bellman principle from discrete-time control, and it is the key step leading to the stationary Hamilton–Jacobi–Bellman equation.
See Tran, §2.2, Theorem 2.6.

2. Controlled Dynamics and Value Function

We consider the deterministic controlled system

$$\begin{cases} y'(s) = b(y(s), v(s)), & s > 0, \\ y(0) = x, \end{cases}$$

where

- $x \in \mathbb{R}^n$ is the initial state,
- V is a nonempty compact metric control space,
- $v(\cdot) : [0, \infty) \to V$ is a measurable admissible control,
- $b : \mathbb{R}^n \times V \to \mathbb{R}^n$ is the controlled velocity field.

Throughout this lecture, we assume:

1. $b \in C(\mathbb{R}^n \times V; \mathbb{R}^n)$;
2. there exists $C > 0$ such that

$$|b(x, v)| \le C \qquad \text{for all } x \in \mathbb{R}^n,\ v \in V;$$

3. there exists $C > 0$ such that

$$|b(x_1, v) - b(x_2, v)| \le C|x_1 - x_2| \qquad \text{for all } x_1, x_2 \in \mathbb{R}^n,\ v \in V;$$

4. the running cost $f : \mathbb{R}^n \times V \to \mathbb{R}$ is continuous, bounded, and Lipschitz continuous in x, uniformly in v;
5. the discount factor satisfies $\lambda > 0$.

For each admissible control $v(\cdot)$, let $y_{x,v(\cdot)}(\cdot)$ denote the corresponding trajectory starting from x. The infinite-horizon discounted cost is

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

and the value function is defined by

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

Thus $u(x)$ is the infimal total discounted cost when the system starts from x.
This is the setting used for the DPP in Tran, §2.1–2.2.

3. Why a Dynamic Programming Principle Should Hold

The DPP expresses the idea that an optimal control problem can be split at any intermediate time.

Suppose the system starts from x, and we fix some time $t > 0$.
A control on the interval $[0, t]$ determines a trajectory segment and hence a new state $y_x(t)$ at time t. From that point onward, the problem is again an optimal control problem of exactly the same form, but now starting from $y_x(t)$.

Therefore, the infimal total cost from x should be equal to:

1. the cost accumulated from time 0 to time t, and
2. the discounted continuation value starting from $y_x(t)$.

This is the continuous-time version of the Bellman principle of optimality.

4. Semigroup Property of Controlled Trajectories

The proof of the DPP is based on a simple but fundamental fact: the tail of a controlled trajectory is again a controlled trajectory.

Proposition 4.1 (Shift / Semigroup Property)

Let $v(\cdot)$ be an admissible control, and let $y_{x,v(\cdot)}(\cdot)$ be the corresponding trajectory.
Fix $t, s \geq 0$, and define the shifted control

$$\tilde{v}(\tau) := v(t + \tau), \qquad \tau \geq 0.$$

Then

$$y_{x,v(\cdot)}(t + s) = y_{y_{x,v(\cdot)}(t),\, \tilde{v}(\cdot)}(s).$$

Proof

Define

$$\eta(s) := y_{x,v(\cdot)}(t + s), \qquad s \geq 0.$$

Then

$$\eta'(s) = \frac{d}{ds} y_{x,v(\cdot)}(t + s) = b(y_{x,v(\cdot)}(t + s), v(t + s)).$$

Since $\tilde{v}(s) = v(t + s)$, we obtain

$$\eta'(s) = b(\eta(s), \tilde{v}(s)).$$

Also,

$$\eta(0) = y_{x,v(\cdot)}(t).$$

Hence η is exactly the trajectory starting from $y_{x,v(\cdot)}(t)$ with control $\tilde{v}(\cdot)$. Therefore

$$\eta(s) = y_{y_{x,v(\cdot)}(t),\, \tilde{v}(\cdot)}(s).$$

This proves the claim. □

Remark

This is called a semigroup property because the evolution from time 0 to time $t+s$ can be decomposed into:

- evolution from 0 to t, and then
- evolution from t to $t + s$.

The DPP is a consequence of this dynamical decomposition together with the corresponding decomposition of the cost functional.

5. Splitting the Cost Functional at an Intermediate Time

Let $v(\cdot)$ be any admissible control. For any $t > 0$,

$$J(x, v(\cdot)) = \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + \int_t^\infty e^{-\lambda s} f(y_x(s), v(s))\, ds.$$

In the second integral, substitute $s = t + \tau$. Then

$$\int_t^\infty e^{-\lambda s} f(y_x(s), v(s))\, ds = e^{-\lambda t} \int_0^\infty e^{-\lambda \tau} f(y_x(t+\tau), v(t+\tau))\, d\tau.$$

By the semigroup property,

$$y_x(t+\tau) = y_{y_x(t), \tilde{v}(\cdot)}(\tau), \qquad \tilde{v}(\tau) = v(t+\tau).$$

Therefore,

$$J(x, v(\cdot)) = \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t} J(y_x(t), \tilde{v}(\cdot)).$$

This identity is the basic algebraic structure behind the DPP.
See Tran, §2.2.

6. Dynamic Programming Principle

Theorem 6.1 (Continuous-Time DPP)

For every $x \in \mathbb{R}^n$ and every $t > 0$,

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

Here $y_x(\cdot)$ denotes the trajectory corresponding to the control $v(\cdot)$ and initial state x.
This is Tran, Theorem 2.6.

7. Interpretation of the DPP

The theorem says that the optimal cost from x can be computed in two stages:

1. choose a control on the short interval $[0, t]$,
2. pay the discounted running cost accumulated on that interval,
3. arrive at the state $y_x(t)$,
4. then use the continuation value from that new state.

Thus the future cost is summarized completely by the value function $u(y_x(t))$, and the discount factor $e^{-\lambda t}$ accounts for the passage of time.

This recursive structure is the continuous-time analogue of the discrete Bellman equation studied earlier in the course.
See Tran, §2.2.

The picture below emphasizes that the DPP does not require us to know the entire optimal tail while optimizing the first interval. Once the intermediate state $y_x(t)$ is reached, all later decisions are compressed into the single number $u(y_x(t))$.

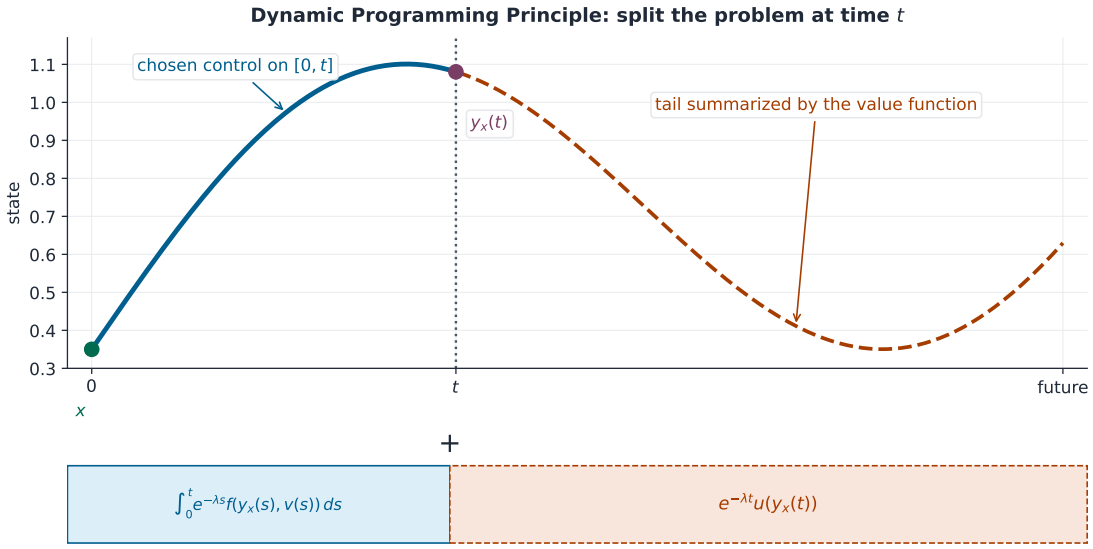


Figure 3.1. Splitting at time t decomposes the cost into the discounted running cost on $[0, t]$ and the discounted continuation value $e^{-\lambda t} u(y_x(t))$. The dashed tail represents the future optimization summarized by the value function.

8. Proof of the DPP

We prove the theorem by showing both inequalities.

Step 1. Proof of

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

Fix any admissible control $v(\cdot)$. By the cost splitting formula,

$$J(x, v(\cdot)) = \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t} J(y_x(t), \tilde{v}(\cdot)).$$

Since

$$u(y_x(t)) \leq J(y_x(t), \tilde{v}(\cdot)),$$

we obtain

$$J(x, v(\cdot)) \geq \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t} u(y_x(t)).$$

Taking the infimum over all admissible controls $v(\cdot)$, we conclude

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

Step 2. Proof of

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

Fix an admissible control $v(\cdot)$ on $[0, t]$, and let $y_x(t)$ be the corresponding state at time t.
Let $\varepsilon > 0$. By the definition of the infimum, there exists an admissible control $w(\cdot)$ such that

$$J(y_x(t), w(\cdot)) \leq u(y_x(t)) + \varepsilon.$$

Now define the concatenated control

$$v^*(s) = \begin{cases} v(s), & 0 \le s < t, \\ w(s-t), & s \ge t. \end{cases}$$

Then

$$J(x, v^*(\cdot)) = \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t} J(y_x(t), w(\cdot)).$$

Hence

$$J(x, v^*(\cdot)) \le \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t}(u(y_x(t)) + \varepsilon).$$

Since $u(x) \le J(x, v^*(\cdot))$, we obtain

$$u(x) \le \int_0^t e^{-\lambda s} f(y_x(s), v(s))\, ds + e^{-\lambda t} u(y_x(t)) + e^{-\lambda t}\varepsilon.$$

Now take the infimum over all admissible controls $v(\cdot)$ on $[0, t]$, and then let $\varepsilon \to 0$. Every measurable prefix on $[0, t]$ has a measurable extension to $[0, \infty)$, and the expression below depends only on that prefix. Thus this infimum agrees with the one over full admissible controls, and we conclude

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

Combining the two inequalities proves the theorem. □

9. Why the Discount Factor Matters

The discount factor $e^{-\lambda s}$, with $\lambda > 0$, is important for two reasons.

9.1 Convergence of the infinite-horizon cost

Since f is assumed bounded, say $|f| \le M$, we have

$$|J(x, v(\cdot))| \le \int_0^\infty e^{-\lambda s} M\, ds = \frac{M}{\lambda}.$$

Thus the cost functional is finite for every admissible control.

9.2 Appearance of the zero-order term in the HJB equation

When we formally expand the DPP for small t, the term $e^{-\lambda t}$ produces the λu term in the stationary Hamilton–Jacobi equation.
So the discount factor is not merely technical: it directly determines the form of the PDE.

This point is emphasized in Tran, Chapter 2.

10. From DPP to the Stationary HJB Equation

The DPP leads formally to the stationary Hamilton–Jacobi–Bellman equation.
Assume for the moment that u is C^1, and consider a constant control $v(s) \equiv v$ on a short interval $[0, t]$.

Then

$$y_x(t) = x + t\, b(x, v) + o(t),$$

and

$$\int_0^t e^{-\lambda s} f(y_x(s), v)\, ds = t\, f(x, v) + o(t).$$

Also,

$$e^{-\lambda t} u(y_x(t)) = (1 - \lambda t + o(t))(u(x) + t\, b(x, v) \cdot Du(x) + o(t)),$$

so

$$e^{-\lambda t} u(y_x(t)) = u(x) + t(b(x, v) \cdot Du(x) - \lambda u(x)) + o(t).$$

Since every constant control is admissible, substituting these expansions into the DPP first gives

$$f(x, v) + b(x, v) \cdot Du(x) - \lambda u(x) \geq 0 \qquad \text{for every } v \in V.$$

The reverse inequality is obtained formally by choosing, for each small t, a control whose DPP expression is within $o(t)$ of the infimum and using the uniform short-time expansions. Consequently,

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

Equivalently,

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

where

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

This is the stationary HJB equation associated with the control problem.
A fully rigorous treatment requires viscosity solutions, which will be introduced later.
See Tran, §2.3.

11. Example: DPP in a Discounted Exit-Time Eikonal Model

To illustrate the DPP in a mathematically consistent way, we consider a **discounted exit-time** problem rather than the pure infinite-horizon cost with $f \equiv 1$.

Example 3.1

Let $n = 1$, $V = [-1, 1]$, and

$$b(x, v) = v.$$

Let the target set be

$$K = \{0\},$$

and define the exit time

$$\tau_x(v) := \inf\{s \geq 0 : y_{x,v(\cdot)}(s) = 0\}.$$

Consider the discounted cost

$$J_K(x, v(\cdot)) := \int_0^{\tau_x(v)} e^{-s}\, ds.$$

The corresponding value function is

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

Since the speed is bounded by 1, the fastest way to reach the origin is to move toward it at maximal speed. Hence the minimum exit time is

$$\tau_x^* = |x|.$$

Therefore

$$u(x) = \int_0^{|x|} e^{-s}\, ds = 1 - e^{-|x|}.$$

This is the standard discounted minimum-time type profile.

Verification of the DPP

Fix $x > 0$, and let $0 < t < x$.
The optimal control on $[0, t]$ is $v \equiv -1$, so

$$y_x(t) = x - t.$$

The DPP becomes

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

For the optimal short-time motion,

$$\int_0^t e^{-s}\, ds = 1 - e^{-t},$$

and

$$u(y_x(t)) = u(x - t) = 1 - e^{-(x-t)}.$$

Hence

$$1 - e^{-t} + e^{-t} u(x - t) = 1 - e^{-t} + e^{-t}(1 - e^{-(x-t)}) = 1 - e^{-x} = u(x).$$

So the DPP is verified exactly for this example.

The same argument works for $x < 0$, using the control $v \equiv +1$.
At $x = 0$, the value is $u(0) = 0$, and the function is not differentiable there. This is one of the earliest indications that classical differentiability is too restrictive, and that a weak notion of solution will be needed later.

Remark

The formula

$$u(x) = 1 - e^{-|x|}$$

is therefore naturally associated with a **discounted minimum-time / exit-time problem**, not with the unconstrained infinite-horizon cost

$$\int_0^\infty e^{-s} \cdot 1\, ds.$$

12. Conceptual Summary

The Dynamic Programming Principle is the fundamental recursive identity for the value function. Its meaning is simple but profound:

- the control problem can be split at any time,
- the future cost is summarized by the value function,
- the value is obtained by taking the infimum of the short-time cost plus the discounted continuation value.

The DPP is more primitive than the Hamilton–Jacobi equation.
It is formulated directly in terms of trajectories and costs, and it remains meaningful even when the value function is not differentiable. The PDE emerges only after performing a small-time expansion of the DPP.

Thus the logical structure of the theory is:

$$\text{controlled dynamics} \longrightarrow \text{value function} \longrightarrow \text{DPP} \longrightarrow \text{HJB equation}.$$

13. Exercises

Exercise 3.1

Prove the semigroup property of controlled trajectories in full detail.

Exercise 3.2

Starting from the definition of $J(x, v(\cdot))$, derive the cost splitting formula at time t.

Exercise 3.3

Rewrite the proof of the DPP, clearly separating the two inequalities.

Exercise 3.4

In the discounted exit-time example, verify directly that for $x > 0$ and $0 < t < x$,

$$u(x) = 1 - e^{-t} + e^{-t}u(x - t).$$

Exercise 3.5

Assuming $u \in C^1$, derive formally the stationary HJB equation from the DPP.

References

- Hung Vinh Tran, *Hamilton–Jacobi Equations*, §2.1–2.3.