Dynamic Programming

Author: Wenxiao Yang

Institute: Haas School of Business, University of California Berkeley

Date: 2025

Mind offline, notes online.

Contents

Chapter 1 Dynamic Programming in Continuous Time

This chapter is based on the notes of Jesús Fernández-Villaverde and Galo Nuño, Wikipedia, the notes of Benjamin Moll, and the notes of J. Miguel Villas-Boas.

1.1 Optimal Control and Hamiltonian

1.1.1 Hamiltonian

Consider a problem

$$\max_{u(t)} J = \int_{t_0}^{t_1} I(x(t), u(t), t) dt$$
$$\text{s.t. } \dot{x}(t) = f(x(t), u(t), t)$$
$$x(0) = x_0$$

The solution method involves defining an ancillary function known as the control Hamiltonian

Definition 1.1 (Hamiltonian)

$\mathcal{H}(x(t), u(t), \lambda(t), t) \equiv I(x(t), u(t), t) + \lambda^T(t) f(x(t), u(t), t)$

Proposition 1.1 (Conditions for Maximum)

The **first-order necessary conditions** for a maximum are given by

1. Maximum principle:

$$\mathcal{H}_u(x(t), u(t), \lambda(t), t) = I_u(x(t), u(t), t) + \lambda^T(t) f_u(x(t), u(t), t) = 0$$

2. Costate equations:

$$\dot{\lambda}(t) = -\mathcal{H}_x(x(t), u(t), \lambda(t), t) = -[I_x(x(t), u(t), t) + \lambda^T(t) f_x(x(t), u(t), t)]$$

with boundary condition for co-state variable(s) $\lambda(t)$, called "transversality condition"

$$\lambda(t_1) = 0 \text{ or } \lim_{t_1 \to \infty} \lambda(t_1) = 0 \text{ for infinite time horizons}$$

3. State transition function:

$$\dot{x}(t) = \mathcal{H}_\lambda(x(t), u(t), \lambda(t), t) = f(x(t), u(t), t)$$
$$x(0) = x_0$$

A **sufficient condition** for a maximum is the concavity of the Hamiltonian evaluated at the solution, i.e.

$$\mathcal{H}_{uu}\left(x^*(t), u^*(t), \lambda(t), t\right) \leq 0$$

where $u^*(t)$ is the optimal control and $x^*(t)$ is the resulting optimal trajectory for the state variable. Therefore, the necessary conditions are sufficient if the functions $I(x(t), u(t), t)$ and $f(x(t), u(t), t)$ are concave in both $x(t)$ and $u(t)$.

Proof 1.1 (Derivation from the Lagrangian)

The Lagrangian is

$$\mathcal{L} = \int_{t_0}^{t_1} \left(I(x(t), u(t), t) + \lambda^T(t)\left[f(x(t), u(t), t) - \dot{x}(t)\right]\right) dt$$

In order to eliminate $\dot{x}(t)$, we can use integration by parts on the last term:

$$\begin{aligned}\int_{t_0}^{t_1} \lambda^T(t)\dot{x}(t)dt &= \int_{t_0}^{t_1} \lambda^T(t)dx(t) \\ &= \lambda^T(t)x(t)\Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \dot{\lambda}^T(t)x(t)dt \\ &= \lambda^T(t_1)x(t_1) - \lambda^T(t_0)x(t_0) - \int_{t_0}^{t_1} \dot{\lambda}^T(t)x(t)dt\end{aligned}$$

which can be substituted back into the Lagrangian expression to give

$$\mathcal{L} = \int_{t_0}^{t_1} \Big(I(x(t), u(t), t) + \lambda^T(t)f(x(t), u(t), t) + \dot{\lambda}^T(t)x(t)\Big)\, dt - \lambda^T(t_1)x(t_1) + \lambda^T(t_0)x(t_0)$$

In the form of Lagrangian functional, the first-order necessary condition is, for $t \in (t_0, t_1)$

$$\underbrace{I_u(x(t), u(t), t) + \lambda^T(t)f_u(x(t), u(t), t)}_{=\mathcal{H}_u(x(t), u(t), \lambda(t), t)} = 0$$

$$\underbrace{I_x(x(t), u(t), t) + \lambda^T(t)f_x(x(t), u(t), t)}_{=\mathcal{H}_x(x(t), u(t), \lambda(t), t)} + \dot{\lambda}(t) = 0$$

If both the initial value $x(t_0)$ and terminal value $x(t_1)$ are fixed, no conditions on $\lambda(t_0)$ and $\lambda(t_1)$ are needed. If the terminal value is free, as is often the case, the additional condition $\lambda(t_1) = 0$ is necessary for optimality. The latter is called a transversality condition for a fixed horizon problem.

1.1.2 Optimal Control Problem

Definition 1.2 (Generic Deterministic Optimal Control Problem)

The generic deterministic optimal control problem is

$$v(x_0) = \max_{\{\alpha(t)\}_{t\geq 0}} \int_0^{\infty} e^{-\rho t} r(x(t), \alpha(t)) dt$$

$$\text{s.t. } \dot{x}(t) = f(x(t), \alpha(t))$$

$$x(0) = x_0$$

Note $\frac{dx(t)}{dt}$ *can be written as* $\dot{x}(t)$.

Here, $x \in \mathbb{X} \subset \mathbb{R}^N$ is the **state** vector, $\alpha \in \mathbb{A} \subset \mathbb{R}^M$ is the **control** vector, $\rho > 0$ is the **discount factor**, $f(\cdot) : \mathbb{A} \times \mathbb{X} \to \mathbb{R}^N$ the **drift**, and $r(\cdot) : \mathbb{A} \times \mathbb{X} \to \mathbb{R}$ is the instantaneous **reward** (utility).

We can obtain conditions for an optimum using the Hamiltonian. The current-value Hamiltonian is

$$\mathcal{H}(x, \alpha, \lambda) = r(x, \alpha) + \lambda f(x, \alpha)$$

where $\lambda \in \mathbb{R}^N$ is the "co-state" vector.

Proposition 1.2

The first-order necessary conditions for an optimum are

$$\mathcal{H}_\alpha(x(t), \alpha(t), \lambda(t)) = 0$$

$$\dot{\lambda}(t) = \rho\lambda(t) - \mathcal{H}_x(x(t), \alpha(t), \lambda(t))$$

$$\dot{x}(t) = f(x(t), \alpha(t))$$

for all $t \geq 0$. The boundary condition (``transversality condition'') is

$$\lim_{T\to\infty} e^{-\rho T}\lambda(T)x(T) = 0$$

Proof 1.2

Let $\tilde{H}(x(t), \alpha(t), L(t)) = e^{-\rho t} r(x(t), \alpha(t)) + L(t) f(x(t), \alpha(t))$. By the Proposition 1.1, the first-order necessary for the optimal $\alpha(t)$ is

1. $\tilde{H}_\alpha(x(t), \alpha(t), L(t)) = 0$:

$$\begin{aligned} \tilde{H}_\alpha(x(t), \alpha(t), L(t)) = e^{-\rho t} r_\alpha(x(t), \alpha(t)) + L(t) f_\alpha(x(t), \alpha(t)) = 0 \\ \Leftrightarrow \quad r_\alpha(x(t), \alpha(t)) + \underbrace{e^{\rho t} L(t)}_{:=\lambda(t)} f_\alpha(x(t), \alpha(t)) = 0 \\ \Leftrightarrow \quad \mathcal{H}_\alpha(x(t), \alpha(t), \lambda(t)) = 0 \end{aligned}$$

2. $\dot{x}(t) = f(x(t), \alpha(t))$

3. $\dot{L}(t) = -[e^{-\rho t} r_x(x(t), \alpha(t)) + L(t) f_x(x(t), \alpha(t))]$: By substituting $L(t) = e^{-\rho t}\lambda(t)$, we have

$$
\begin{aligned}
& -\rho e^{-\rho t}\lambda(t) + e^{-\rho t}\dot{\lambda}(t) = -[e^{-\rho t} r_x(x(t), \alpha(t)) + e^{-\rho t}\lambda(t) f_x(x(t), \alpha(t))] \\
\Leftrightarrow \quad & -\rho\lambda(t) + \dot{\lambda}(t) = -\underbrace{[r_x(x(t), \alpha(t)) + \lambda(t) f_x(x(t), \alpha(t))]}_{:=\mathcal{H}_x(x(t), \alpha(t), \lambda(t))} \\
\Leftrightarrow \quad & \dot{\lambda}(t) = \rho\lambda(t) - \mathcal{H}_x(x(t), \alpha(t), \lambda(t))
\end{aligned}
$$

1.2 Hamilton-Jacobi-Bellman Equation

Consider the generic deterministic optimal control problem is

$$
\begin{aligned}
v(x_0) = \max_{\{\alpha(t)\}_{t\geq 0}} & \int_0^\infty e^{-\rho t} r(x(t), \alpha(t)) dt \\
\text{s.t. } & \dot{x}(t) = f(x(t), \alpha(t)) \\
& x(0) = x_0
\end{aligned}
$$

More generally, define the problem begins at t with $x(t) = x$ as

$$
\begin{aligned}
v(t, x) = \max_{\{\alpha(s)\}_{s\geq t}} & \int_t^\infty e^{-\rho(s-t)} r(x(s), \alpha(s)) ds \\
\text{s.t. } & \dot{x}(t) = f(x(t), \alpha(t)) \\
& x(t) = x
\end{aligned}
$$

which may be written as $v_t(x)$.

Note *$v_t(x)$'s value only depends on x, i.e., $v_{t_1}(x) = v_{t_2}(x) = v(x), \forall t_1 \neq t_2$.*

Proposition 1.3 (Hamilton-Jacobi-Bellman Equation)

The value function of the generic optimal control problem satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$
\rho v(x) = \underbrace{\frac{\partial v}{\partial t}}_{=0} + \max_{\alpha \in A}\left\{ r(x, \alpha) + \frac{\partial v(x)}{\partial x} \cdot f(x, \alpha) \right\}
$$

with a transversality condition $\lim_{T\to\infty} e^{-\rho T} v_T(x) = 0$.
($\frac{\partial v(x)}{\partial x} \in \mathbb{R}^N$ is the gradient of v if $N > 1$.)

Proof 1.3 (Derivation from Discrete-time Bellman)

By the discrete-time Bellman equation,

$$
v_t(x(t)) = \max_{\alpha(t)} \; \Delta r(x(t), \alpha(t)) + e^{-\rho\Delta} v_{t+\Delta}(x(t+\Delta))
$$

where $x(t+\Delta) = x(t) + \Delta[f(x(t), \alpha(t))]$.

For small Δ (will take $\Delta \to 0$), $e^{-\rho\Delta} = 1 - \rho\Delta$.

$$v_t(x(t)) = \max_{\alpha(t)} \ \Delta r(x(t), \alpha(t)) + (1-\rho\Delta) v_{t+\Delta}(x(t+\Delta))$$

Subtract $(1-\rho)\Delta v_t(x(t))$ from both sides:

$$\begin{aligned}\rho\Delta v_t(x(t)) = \max_{\alpha(t)} \ \Delta r(x(t), \alpha(t)) + (1-\rho\Delta)[v_{t+\Delta}(x(t+\Delta)) - v_t(x(t+\Delta))] \\ +(1-\rho\Delta)[v_t(x(t+\Delta)) - v_t(x(t))]\end{aligned}$$

Divide by Δ and manipulate last term

$$\begin{aligned}\rho v_t(x(t)) = \max_{\alpha(t)} \ r(x(t), \alpha(t)) + (1-\rho\Delta)\frac{v_{t+\Delta}(x(t+\Delta)) - v_t(x(t+\Delta))}{\Delta} \\ +(1-\rho\Delta)\frac{v_t(x(t+\Delta)) - v_t(x(t))}{x(t+\Delta) - x(t)}\frac{x(t+\Delta) - x(t)}{\Delta}\end{aligned}$$

Taking $\Delta \to 0$,

$$\rho v_t(x(t)) = \underbrace{\frac{\partial v}{\partial t}}_{=0} + \max_{\alpha(t)}\left\{ r(x(t), \alpha(t)) + \frac{\partial v(x(t))}{\partial x}\dot{x}(t)\right\}$$

Note *The HJB can be written in the terms of maximizing the Hamiltonian,*

$$\rho v(x) = \max_{\alpha \in A}\left\{ \underbrace{r(x, \alpha) + \underbrace{\lambda \cdot f(x, \alpha)}_{=\sum_{i=1}^N \lambda_i f_i(x,\alpha)}}_{:=\mathcal{H}(x,\alpha,\lambda)} \right\}$$

where $\lambda_i := \frac{\partial v}{\partial x_i}(x(t))$ *and* $\frac{d\lambda_i}{dt} = \frac{\partial^2 v}{\partial x_i \partial t}(x(t)) + \frac{\partial^2 v}{\partial x_i^2}\dot{x}(t)$.

1.3 Brownian Motion and Diffusion Process

Definition 1.3 (Standard Brownian Motion)

A standard Brownian motion (Wiener Process) is a stochastic process W which satisfies

$$W(t+\Delta t) - W(t) = \epsilon_t \sqrt{\Delta t}, \ \epsilon_t \sim \mathcal{N}(0,1), \ W(0) = 0$$

We can see $W(t) \sim \mathcal{N}(0, t)$.

A more generalized Brownian motion can be defined as

Definition 1.4 (Brownian Motion)

A Brownian motion with drift μ and variance σ^2 is given as

$$x(t) = x(0) + \mu t + \sigma W(t)$$

where $W(t)$ is a Wiener Process.

Since $W(t) \sim \mathcal{N}(0, t)$, we have $x(t) - x(0) \sim \mathcal{N}(\mu t, \sigma^2 t)$.

Let $dW(t) := W(t + \Delta t) - W(t) = \epsilon_t \sqrt{\Delta t}$, where $\epsilon_t \sim \mathcal{N}(0, 1)$. Then, we can write

$$dx(t) = \mu dt + \sigma dW(t)$$

which is called a **stochastic differential equation**.

1.3.1 Diffusion Process

Definition 1.5 (Diffusion Process)

The stochastic differential equation can be generalized further to the case that considering the dependence of x on μ and σ:

$$dx = \mu(x)dt + \sigma(x)dW$$

which is called a **diffusion process**. The $\mu(\cdot)$ is called the drift and the $\sigma(\cdot)$ is called the diffusion.

(All results can be extended to the case where they depend on t, $\mu(x, t), \sigma(x, t)$ but abstract from this for now.)

Note *By choosing functions μ and σ, you can get pretty much any stochastic process you want (except jumps).*

1.4 Stochastic HJB Equations

The generic problem of stochastic optimal control is

$$v(x_0) = \max_{\{\alpha(t)\}_{t\geq 0}} \mathbb{E} \int_0^\infty e^{-\rho t} r(x(t), \alpha(t))dt$$

$$\text{s.t. } dx(t) = f(x(t), \alpha(t))dt + \sigma(x(t))dW(t)$$

$$\alpha(t) \in A, \forall t \geq 0, \; x(0) = x_0$$

Proposition 1.4 (HJB of Stochastic Optimal Control)

The HJB equation is

$$\rho v(x) = \max_{\alpha \in A} r(x, \alpha) + v'(x) f(x, \alpha) + \frac{1}{2} v''(x) \sigma^2(x)$$

Multivariate Case:

$$\rho v(x) = \max_{\alpha \in A} r(x,\alpha) + \sum_{i=1}^{N} \frac{\partial v(x)}{\partial x_i} f_i(x,\alpha) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 v(x)}{\partial x_i \partial x_j}\sigma_{ij}^2(x)$$

In vector notation:

$$\rho v(x) = \max_{\alpha \in A} r(x,\alpha) + \nabla_x v(x) \cdot f(x,\alpha) + \frac{1}{2}\text{tr}(\nabla^2 v(x)\sigma^2(x))$$

where $\nabla_x v(x)$ is the gradient of $v(x)$ and $\nabla^2 v(x)$ is the Hessian of $v(x)$.

1.4.1 Ito's Lemma

Let x be a scalar diffusion.

$$dx = \mu(x)dt + \sigma(x)dW$$

We are interested in the evolution of $y(t) = f(x(t))$ where f is any twice differentiable function.

Lemma 1.1 (Ito's Lemma)

$y(t) = f(x(t))$ follows

$$df(x) = \left(f'(x)\mu(x) + \frac{1}{2}f''(x)\sigma^2(x)\right)dt + f'(x)\sigma(x)dW$$

Multivariate Case:

$$df(x) = \left(\sum_{i=1}^{N}\frac{\partial f(x)}{\partial x_i}\mu_i(x) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\sigma_{ij}^2(x)\right)dt + \sum_{i=1}^{N}\frac{\partial f(x)}{\partial x_i}\sigma_i(x)dW_i$$

In vector notation:

$$df(x) = \left(\nabla_x f(x)\cdot\mu(x) + \frac{1}{2}\text{tr}(\nabla^2 f(x)\sigma^2(x))\right)dt + \nabla_x f(x)\cdot\sigma(x)dW$$

where $\nabla_x f(x)$ is the gradient of $f(x)$ and $\nabla^2 f(x)$ is the Hessian of $f(x)$.

Note *It is extremely powerful because it says that any (twice differentiable) function of a diffusion is also a diffusion.*

1.4.2 Kolmogorov Forward Equations

Let x be a scalar diffusion.

$$dx = \mu(x)dt + \sigma(x)dW, \quad x(0) = x_0$$

We are interested in the evolution of the *distribution* of x, $g(x,t)$, and in particular in the stationary distribution $g(x)$. Natural thing to care about especially in heterogenous agent models.

Proposition 1.5 (Kolmogorov Forward Equation)

Given an initial distribution $g(x,0) = g_0(x)$, $g(x,t)$ satisfies the PDE

$$\frac{\partial g(x,t)}{\partial t} = -\frac{\partial[\mu(x)g(x,t)]}{\partial x} + \frac{1}{2}\frac{\partial^2[\sigma^2(x)g(x,t)]}{\partial x^2}$$

Corollary 1.1

If a stationary distribution $g(x)$ exists, it satisfies the ODE

$$0 = \frac{\partial g(x,t)}{\partial t} = -\frac{\partial[\mu(x)g(x,t)]}{\partial x} + \frac{1}{2}\frac{\partial^2[\sigma^2(x)g(x,t)]}{\partial x^2}$$

Multivariate Case:

$$\frac{\partial g(x,t)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial[\mu_i(x)g(x,t)]}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2[\sigma_{ij}^2(x)g(x,t)]}{\partial x^2}$$

Chapter 2 [?]: Dynamic Bargaining with Private Information

2.1 "Classic" Coase Conjecture

Consider a seller facing a buyer has private value of the good $v \in [1, 2]$ that is distributed according to an atomless distribution with full support, $F(v)$. The seller has cost $c \geq 0$ to serve the buyer. Every period of an infinite horizon game, the uninformed seller makes an offer p_t. If p_t is accepted, the game ends with payoffs $v - p_t$ and $p_t - c$ for the buyer and the seller, respectively. If p_t is rejected, the seller makes another offer at time $t + \Delta$. Both buyer and seller discount future with the rate r. Thus, Δ can be thought of as the commitment power of the seller.

Since higher-value buyers lose more from delay, the Skimming Property holds in any equilibrium.

Lemma 2.1 (Skimming Property)

If type v is willing to accept an offer p_t, then all higher types strictly prefer to accept this offer.

Denote by k_t the highest type remaining before the offer at time t is made. Then, the seller beliefs at time t are $v \in [1, k_t]$ with truncated distribution $F(v)/F(k_t)$, where $k_0 = 2$.

A distinction arises in this game: (1). $c < 1$ (the gap case); (2). $c = 1$ (the no-gap case).

As shown by [**?**] and [**?**], the game must end in a finite number of rounds (uniformly bounded for all Δ). The intuition is: the difference between distinguishing k_t and an immediate trading 1 is $k_t - 1$. If the $k_t - 1$ is relative small to $1 - c$. The seller prefers an immediate trade rather than keep bargaining. So, there is a critical k_t that is close enough to 1 and the seller offers $p_t = 1$ to get an immediate trade.

By the backward induction, equilibrium must be unique (up to seller's randomization at time 0). These equilibria have the property that the seller value depends only on the state of the game, k_t, a property called "stationarity" that is similar to the equilibria being Markov.

Stationary equilibria continue to exist when $c = 1$, but other nonstationary equilibrium can be constructed by [**?**].

Formally, the Coase conjecture ([**?**]) states that

Proposition 2.1 (Coase Conjecture)

As $\Delta \to 0$ (i.e., the seller loses all ability to commit to prices), prices fall to the lowest buyer valuation, $p_0 \to 1$, and there is no inefficient delay.

2.2 Exogenously Interdependent Values

Suppose the cost of selling to type v is an increasing function of the buyer's type $c(v)$ with $c'(v)$. We assume that immediate trade is strictly efficient for all $v \in (1, 2]$, $c(v) < v$.

The equilibrium dynamics depend on the degree of adverse selection.

1. If there is little adverse selection, i.e., $\mathbb{E}[c(v)] \leq 1$, the continuous-time limit of the unique stationary equilibrium has immediate trade ([**?**]).
2. If there is more adverse selection, $\mathbb{E}[c(v)] > 1$, the seller would prefer not to trade at all rather than offering $p_0 = 1$ and trading with all types, so trade cannot be efficient in equilibrium.

   No trade cannot be an equilibrium either:

   (a). For no trade to be an equilibrium, prices would always have to be above 2. If the price were ever lower, a mass of buyers $(2 - \epsilon, 2]$ would accept this price, contradicting to no trade.

   (b). If the price were always above 2, there would be a profitable deviation for the seller. The seller could offer $p_0 = c(2) < 2$ and then offer a sequence of prices that would lead to no losses.

   What then do equilibria look like?

Suppose the continuous-time limit equilibrium strategy of the buyer is characterized by a continuous downward-sloping demand function, $P(v)$. Then, the seller's best response problem is to choose the speed to maximize her expected value given k_t:

$$rV(k_t) = \max_{\dot{k}_t \in [0,\infty]} \left(P(k_t) - c(k_t) - V(k_t)\right)(-\dot{k}_t)\frac{f(k_t)}{F(k_t)} + V'(k_t)\dot{k}_t$$

When trade happens, the seller collects $P(k_t) - c(k_t)$ and the game ends. Trade happens with a probability flow

Chapter 3 [?]: Bargaining Between Collaborators of A Stochastic Project

Two risk-neutral firms with discount rate r, firm 1 and firm 2, can undertake an irreversible joint project with a one-time payoff. Let x_t denote the expected return of the project at time t. Suppose x_t follows a Brownian motion, $dx_t = \sigma dW_t$, with volatility σ and initial position x_0, where W_t is a standard Wiener process.

The order of movement at time t is determined by a recognition process $f_t \in \{1, 2\}$. Firm i is the *Proposer* at t if $f_t = i$ and *Responder* at time t if $f_t \neq i$. The roles are switched upon the arrival of a Poisson process with rate λ.[1]

The offer proposed by the proposer is p_t, and denote $p_t = -\infty$ if the proposer does not make an offer. If accepted, responder receives p_t and the proposer gets $x_t - p_t$. The responder can choose an outside option, which gives a one-time payoff of ω for both firms.

[1] WLOG, let $f_0 = 1$. Let $(\Sigma, \mathcal{F}, P)$ be the probability space that supports the Wiener process W_t and the Poisson counting process, and $\mathcal{F} = (F_t)_{t\in[0,\infty)}$ be the filtration process satisfying the usual assumptions.