CE 273
Markov Decision Processes

Lecture 26

Continuous-Time Control

Lecture Outline

Lecture Outline

HJB Equation

Consider a continuous-time dynamical system in which the state at time t is represented as $x(t) \in \mathbb{R}^n$ and control is represented as $u(t) \in \mathbb{R}^m$.

The system dynamics are deterministic and denoted as

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

These problems are usually formulated for a finite time horizon T and the initial state $x(0)$ is assumed to be known.

We will assume that $u(t)$ is continuous and belongs to U for all $t \in [0, T]$.

Given a control trajectory $\{u(t)|t \in [0, T]\}$, we assume that $\{x(t)|t \in [0, T]\}$ is called the state trajectory and can be uniquely identified.

HJB Equation

Objective

The objective of continuous-time models is very similar to finite-horizon dynamic programs.

The goal is to find $u(t)$ and also the evolution of states $x(t)$ over the time period of interest $[0, T]$ that and minimizes

$$h(x(T)) + \int_0^T g(x(t), u(t))$$

where h is the terminal cost and g is the cost in dt when the state is $x(t)$ and a control $u(t)$ is applied.

We will further assume that f, g, and h are continuously differentiable. Much of our discussion in this lecture will be informal and more rigorous development of these ideas can be found in control theory texts.

HJB Equation

Example

Consider a continuous-time investment problem in which $x(t)$ is the wealth at time t and $u(t)$ is the proportion of wealth that is reinvested and $(1 - u(t))x(t)$ is kept as savings.

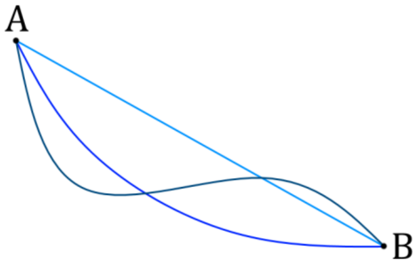
Assume that the initial wealth is $x(0)$ and the dynamics are

$$\dot{x}(t) = \gamma u(t)x(t)$$

where γ is some interest rate-like constant. Suppose we want to maximize the total savings over a fixed time horizon, then the problem can be written as

$$\max \int_0^T (1 - u(t))x(t)dt$$
$$\text{s.t.}\, 0 \leq u(t) \leq 1 \, \forall\, t \in [0, T]$$

Brachistrochrone Problem:
Given two points A and B in a vertical plane, what is the curve traced out by a point acted on only by gravity, which starts at A and reaches B in the shortest time.

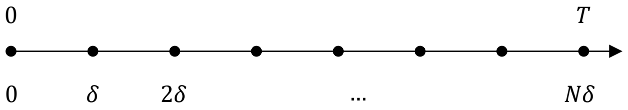


HJB Equation

Discrete Version

Just as with the DP equations, we can develop equivalent optimality conditions for the continuous time problem. To this end, let us first discretize the problem and then take some limits.

Suppose the time period of interest $[0, T]$ is divided into N intervals



Let the state and control at each of these control points be denoted as x_k and u_k, i.e.,

$$x_k = x(k\delta)$$
$$u_k = u(k\delta)$$

HJB Equation

Discrete Version

The system dynamics is represented as

$$x_{k+1} = x_k + f(x_k, u_k)\delta$$

The objective function can be written as

$$h(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k)\delta$$

The optimal value functions of the discrete problem $\hat{J}^*(k, x_k)$ satisfy the following Bellman equations

$$\hat{J}^*(N\delta, x) = h(x)$$

$$\hat{J}^*(k\delta, x) = \min_{u \in U} \left\{ g(x, u)\delta + \hat{J}^*\Big((k+1)\delta, x + f(x, u)\delta\Big) \right\}$$

HJB Equation

Discrete Version

The second term inside the minimization in

$$\hat{J}^*(k\delta, x) = \min_{u \in U} \left\{ g(x,u)\delta + \hat{J}^*\Big((k+1)\delta, x + f(x,u)\delta\Big) \right\}$$

using Taylor series, $\hat{J}^*\Big(k\delta + \delta, x + f(x,u)\delta\Big)$ can be approximated as

$$\hat{J}^*(k\delta, x) + \partial_{k\delta}\hat{J}^*(k\delta, x)\delta + (\nabla_x \hat{J}^*(k\delta, x))^\mathsf{T} f(x,u)\delta + o(\delta)$$

Hence, we can write the second set of Bellman equations as

$$0 = \min_{u \in U} \left\{ g(x,u)\delta + \partial_{k\delta}\hat{J}^*(k\delta, x)\delta + (\nabla_x \hat{J}^*(k\delta, x))^\mathsf{T} f(x,u)\delta + o(\delta) \right\}$$

Divide both sides by δ and take limit as $\delta \to 0$, $k\delta \to t$.

HJB Equation

Continuous Version

Suppose $\lim_{\delta \to 0, k\delta \to t} \hat{J}^*(k\delta, x) = J^*(t, x) \, \forall \, t \in [0, T]$. Then, the optimality conditions in continuous time take the form

$$J^*(T, x) = h(x)$$

$$0 = \min_{u \in U} \left\{ g(x, u) + \partial_t J^*(t, x) + (\nabla_x J^*(t, x))^\mathsf{T} f(x, u) \right\}$$

This equation is called the Hamilton-Jacobi-Bellman (HJB) equation.

As with discrete time problems, the optimal policy $\mu^*(t, x)$ is attained at u's which minimize the RHS in the above expression.

We assumed that $J^*(t, x)$ is continuously differentiable but sufficiency can also be established, i.e., if a function $J(t, x)$ solves this equation, then it is continuously differentiable and is the optimal value function.

HJB Equation

Example: Linear Quadratic Control

Consider a n-dimensional linear system

$$\dot{x}(t) = Ax(t) + Bu(t)$$

where A and B are matrices and the objective is to minimize

$$x(T)^\mathsf{T} Q_T x(T) + \int_0^T \Big(x(t)^\mathsf{T} Q x(t) + u(t)^\mathsf{T} R u(t) \Big) dt$$

where Q, Q_T are symmetric positive semidefinite and R is symmetric positive definite.

For instance you can formulate the bus bunching problem in Assignment 2 as a LQ control problem in which state represents the deviation from the ideal spacing and the controls are the operating speeds.

Likewise, think about controlling a drone using some force and the objective penalizes deviations from the flight path and usage of fuel.

HJB Equation

Example: Linear Quadratic Control

Assuming that the optimal value functions are of the form $J^*(t,x) = x^\mathsf{T} K(t) x$, where $K(t)$ is symmetric, write the HJB equations to this problem.

$$0 = \min_{u \in \mathbb{R}^m} \left\{ x^\mathsf{T} Q x + u^\mathsf{T} R u + \partial_t J(t,x) + (\nabla_x J(t,x))^\mathsf{T} (Ax + Bu) \right\}$$

Substituting $x^\mathsf{T} K(t) x$ for $J(t,x)$ in the above equation,

$$0 = \min_{u \in \mathbb{R}^m} \left\{ x^\mathsf{T} Q x + u^\mathsf{T} R u + x^\mathsf{T} \dot{K}(t) x + 2x^\mathsf{T} K(t) A x + 2x^\mathsf{T} K(t) B u \right\}$$

Setting the gradient of the objective wrt u to 0

$$2B^\mathsf{T} K(t) x + 2Ru = 0$$

Thus, the optimal control is linear and is given by $u = -R^{-1} B^\mathsf{T} K(t) x$.

For $J(t,x) = x^\mathsf{T} K(t) x$ to solve the HJB equation, what conditions must $K(t)$ satisfy?

HJB Equation

Example: Linear Quadratic Control

Substituting $u = -R^{-1}B^{\mathsf{T}}K(t)x$ in the HJB equation, we get

$$0 = x^{\mathsf{T}}\Big(\dot{K}(t) + K(t)A + A^{\mathsf{T}}K(t) - K(t)BR^{-1}B^{\mathsf{T}}K(t) + Q\Big)x$$

Therefore, for $J(t,x) = x^{\mathsf{T}}K(t)x$ to solve the HJB equation for all t and x, we need

$$\dot{K}(t) = -K(t)A - A^{\mathsf{T}}K(t) + K(t)BR^{-1}B^{\mathsf{T}}K(t) - Q$$

with the boundary condition $K(T) = Q_T$.

The above equation is an ordinary matrix differential equation and is quadratic. Hence it is also called the Riccati equation.

Lecture Outline

Pontryagin Minimum Principle

Pontryagin Minimum Principle

Given an optimal state trajectory $x^*(t)$, the optimal control trajectory can be derived for all $t \in [0, T]$ using

$$
\begin{aligned}
u^*(t) &= \arg\min_{u \in U} \left\{ g(x^*(t), u) + \partial_t J^*(t, x^*(t)) + \nabla_x (J^*(t, x^*(t)))^\mathsf{T} f(x^*(t), u) \right\} \\
&= \arg\min_{u \in U} \left\{ g(x^*(t), u) + \nabla_x (J^*(t, x^*(t)))^\mathsf{T} f(x^*(t), u) \right\}
\end{aligned}
$$

Therefore, to find the optimal control trajectory, we just need $\nabla_x J^*(t, x^*(t))$ values only along the state trajectory but not along the entire state space for every t.

It turns out that $\nabla_x J^*(t, x^*(t))$ satisfies a certain differential equation called the *adjoint equation* and can be calculated more easily and hence we can avoid solving the HJB equations.

Pontryagin Minimum Principle

Adjoint Equation

The adjoint equation is a system of n first-order differential equations of the form

$$\dot{p}(t) = -\nabla_x f\big(x^*(t), u^*(t)\big)\, p(t) - \nabla_x g\big(x^*(t), u^*(t)\big)$$

where

$$p(t) = \nabla_x J^*(t, x^*(t))$$

These can be derived informally by differentiating the HJB equations along the optimal state and control trajectories.

What is the terminal boundary condition for the above system of equations?

$$p(T) = \nabla_x J^*(T, x^*(T)) = \nabla_x h(x^*(T))$$

Plugging $p(t)$ back into HJB equations, we can write that $u^*(t)$ satisfies the following set of equations

$$u^*(t) = \arg\min_{u \in U} \left\{ g(x^*(t), u) + p(t)^\mathsf{T} f(x^*(t), u) \right\} \forall\, t \in [0, T]$$

Define a new function $H(x, u, p)$, called the Hamiltonian, where $(x, u, p) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ as

$$H(x, u, p) = g(x, u) + p^\mathsf{T} f(x, u)$$

Hence, we can write the optimal control trajectory in terms of the Hamiltonian as

$$u^*(t) = \arg\min_{u \in U} H(x^*(t), u, p(t)) \,\forall\, t \in [0, T]$$

Pontryagin Minimum Principle

Adjoint Equation

Recall that $p(t)$ satisfies the adjoint equation

$$\dot{p}(t) = -\nabla_x f\big(x^*(t), u^*(t)\big) p(t) - \nabla_x g\big(x^*(t), u^*(t)\big)$$

and $H(x, u, p) = g(x, u) + p^\mathsf{T} f(x, u)$. What is $\nabla_x H(x^*(t), u^*(t), p(t))$?

Hence, the adjoint equation can also be expressed in terms of the Hamiltonian as

$$\dot{p}(t) = -\nabla_x H(x^*(t), u^*(t), p(t))$$

Proposition (Minimum Principle)

Let $u^*(t)$ and $x^*(t)$ be an optimal control and state trajectory and $p(t)$ be a solution to the adjoint equation

$$\dot{p}(t) = -\nabla_x H(x^*(t), u^*(t), p(t))$$

with boundary condition $p(T) = \nabla_x h(x^*(T))$. Then, for all $t \in [0, T]$,

$$u^*(t) = \arg\min_{u \in U} H(x^*(t), u, p(t))$$

Note that this is a necessary condition. It can further be shown that the Hamiltonian along the optimal control and state trajectory $H(x^*(t), u^*(t), p(t))$ is a constant for all $t \in [0, T]$.

To find u^* one still needs x^*. One option is to find u^* as a function of x^* and p from the Hamiltonian and substitute it into the adjoint equation and solve some ODEs.

Pontryagin Minimum Principle

Example

Recall the investment problem that was discussed earlier in which $\dot{x} = \gamma u x$.

$$\max \int_0^T (1 - u(t))x(t)dt$$
$$\text{s.t. } 0 \leq u(t) \leq 1 \, \forall \, t \in [0, T]$$

- Write the Hamiltonian $H(x, u, p)$ and the adjoint equation.
- Maximize the Hamiltonian over $u \in [0, 1]$
- Write $u^*(t)$ in terms of $x^*(t)$ and $p(t)$
- Plug it into the adjoint equation and solve an ODE

Pontryagin Minimum Principle

Example

- Write the Hamiltonian $H(x, u, p)$ and the adjoint equation.

$$H(x, u, p) = (1 - u)x + p\gamma ux$$

$$\dot{p}(t) = \nabla_x H(x^*(t), u^*(t), p(t)) = -\gamma u^*(t)p(t) - 1 + u^*(t)$$
$$p(T) = 0$$

- Maximize the Hamiltonian over $u \in [0, 1]$

$$\begin{aligned} u^*(t) &= \arg\max_{u \in U} H(x^*(t), u, p(t)) \\ &= \arg\max \left\{ (1 - u(t))x^*(t) + p(t)\gamma u(t)x^*(t) \right\} \\ &= \arg\max \left\{ (p(t)\gamma - 1)u(t)x^*(t) \right\} \end{aligned}$$

Hence, $u^*(t)$ is 0 if $p(t) < 1/\gamma$ and 1 if $p(t) >= 1/\gamma$.

Pontryagin Minimum Principle

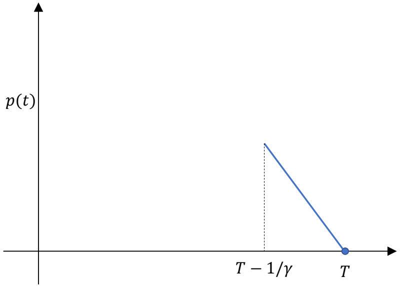
Example

- Write $u^*(t)$ in terms of $x^*(t)$ and $p(t)$

$$u^*(t) = \begin{cases} 0 & \text{if } p(t) < 1/\gamma \\ 1 & \text{if } p(t) \geq 1/\gamma \end{cases}$$

- Plug it into the adjoint equation and solve an ODE

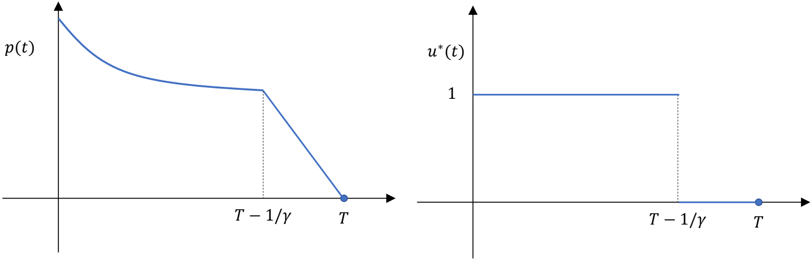
What is $u^*(T)$? $p(T) = 0$ and hence $u^*(T) = 0$. The same is true for t in $(T-\delta, T]$. Therefore, $\dot{p}(t) = -1$



For $t < T - 1/\gamma$, $u^*(t) = 1$. Hence, the adjoint equation is

$$\dot{p}(t) = -\gamma u^*(t) p(t)$$

Hence, $p(t)$ has an exponential shape in this region.



Knowledge of $u^*(t)$ implies that we can calculate $x^*(t)$ using the dynamics $\dot{x}(t) = f(x(t), u(t))$.

Your Moment of Zen