

Lecture 22


Optimal Control

- **Discrete-Time Optimal Control**
 - ✓ Dynamic Programming
 - ✓ Affine Quadratic Problem
 - ✓ Infinite Horizon Linear Quadratic Problem
- **Continuous-Time Optimal Control**
 - ✓ Dynamic Programming (Hamilton-Jacobi-Bellman Equation)
 - ✓ Minimum Principle
 - ✓ Affine Quadratic Problem
 - ✓ Infinite Horizon Linear Quadratic Problem

Basic Principle to Analyze Dynamic Games

Equilibrium concept:


-Nash; Zero-sum; Stackelberg; Correlated



	Single Agent	Multi Agent
Static	Static optimization	Static Game
Dynamic	Dynamic Optimization	Dynamic Game

Dynamic optimization as a static optimization concept:

- Minimum principle (necessary condition)
- Dynamic programming principle (sufficient condition)
- Need to specify information structure


$$\begin{aligned} L^{1*} &\triangleq L^1(\mathbf{u}^{1*}; u^{2*}; \dots; u^{N*}) \leq L^1(\mathbf{u}^1; u^{2*}; \dots; u^{N*}), \\ L^{2*} &\triangleq L^2(u^{1*}; \mathbf{u}^{2*}; \dots; u^{N*}) \leq L^2(u^{1*}; \mathbf{u}^2; \dots; u^{N*}), \\ &\dots \\ L^{N*} &\triangleq L^N(u^{1*}; u^{2*}; \dots; \mathbf{u}^{N*}) \leq L^N(u^{1*}; u^{2*}; \dots; \mathbf{u}^{N*}) \end{aligned}$$

(Think in normal form game setting)

Overview

	Single Agent	Multi Agent
Static	Static optimization	Static Game
Dynamic	Dynamic Optimization	Dynamic Game

Action space

Time space	Model based	Finite	Infinite
	Discrete	Discrete time MDP $P(s_{t+1} s_t, a_t)$	Discrete-time dynamic system $x_{t+1} = f(x_t, u_t)$
	Continuous	Continuous time MDP $P(s_{t+h} s_t, a_t)$	Continuous-time dynamic system $\dot{x}_t = f(x_t, u_t)$

Overview

	Single Agent	Multi Agent
Static	Static optimization	Static Game
Dynamic	Dynamic Optimization	Dynamic Game

Action space

Time space	Model free	Finite	Infinite
	Discrete	Value-based Reinforcement Learning	Policy-based Reinforcement Learning
	Continuous		

Overview

	Single Agent	Multi Agent
Static	Static optimization	Static Game
Dynamic	Dynamic Optimization	Dynamic Game

Action space			
Time space	Model based	Finite	Infinite
	Discrete	Markov Game (Stochastic Game)	DT Infinite dynamic game (Stochastic Game)
	Continuous	Continuous time Markov Game	CT-time Infinite dynamic game (differential game)

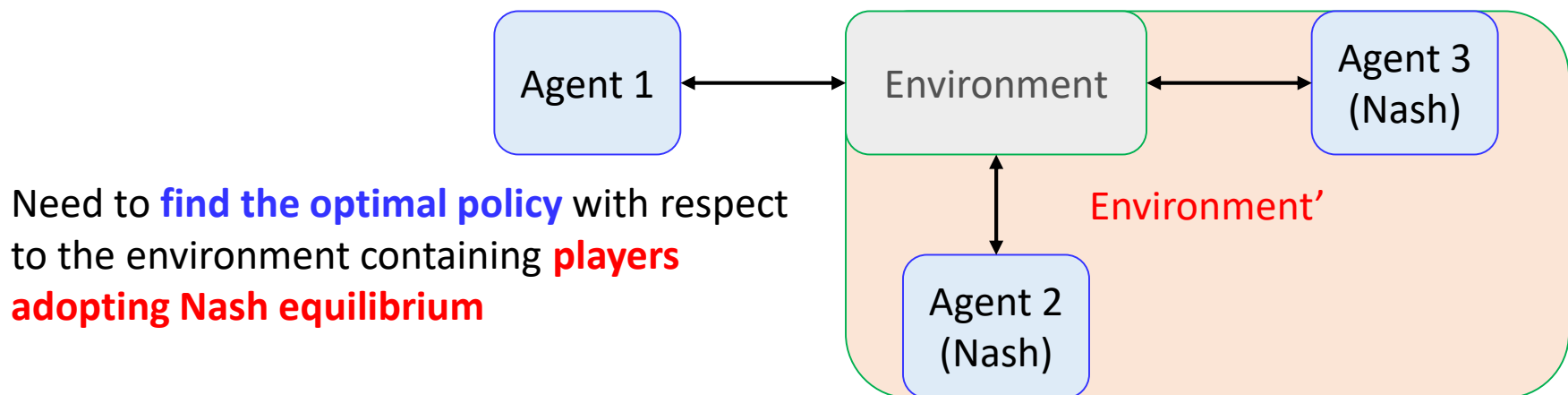
Overview

	Single Agent	Multi Agent
Static	Static optimization	Static Game
Dynamic	Dynamic Optimization	Dynamic Game

		Action space	
		Finite	Infinite
Time space	Model free		
	Discrete	Multi-Agent Value-based RL	Multi-Agent Policy-based RL
	Continuous		

Why We Learn Optimal Control Theory?

- **Optimal control problems** constitute a special class of infinite dynamic games with one player and one criterion
 - The mathematical tools available for such problems are useful in dynamic game theory
- If the players adopt the **non-cooperative Nash equilibrium solution concept**, in which case each player is faced with a single criterion optimization problem (i.e., optimal control problem) with the strategies of the remaining players taken to be **fixed** at their equilibrium values.
 - Hence, in order to verify whether a given set of strategies is in Nash equilibrium, we inevitably have to utilize the tools of **optimal control theory**.



Discrete-Time Optimal Control Problems

Want to find **a control sequence** $u = \{u_k, k \in \mathbf{K}\}$ which minimizes

$$L(u) = \sum_{k=0}^{K-1} g_k(x_k, u_k) + g_K(x_K) \quad (1)$$

where the state variable x_k satisfies discrete-time system constraints:

$$x_{k+1} = f_k(x_k, u_k), \quad u_k \in U_k, \quad k \in \mathbf{K} \quad (2)$$

on time steps $k \in \mathbf{K} = \{0, 1, \dots, K-1\}$

- $u = \{u_k, k \in \mathbf{K}\}, u_k = \gamma_k(x_k)$
 - ✓ $\gamma_k(\cdot)$ is a permissible control strategy at stage $k \in \mathbf{K}$

Dynamic Programming for Discrete-Time Optimal Control Problems

- In order to determine the minimizing control strategy, we define the expression for the minimum cost from **any starting point x** at any **initial time k** , which is so called value function

$$V(k, x) = \min_{\gamma_k, \dots, \gamma_K} \left[\sum_{i=k}^K g_i(x_{i+1}, u_i, x_i) \right] \quad (3)$$

with $u_i = \gamma_i(x_i) \in U_i$ and $x_k = x$

- A direct application of the **principle of optimality** now readily leads to the recursive relation

$$V(k, x) = \min_{u_k} \left[g_k(\overset{x_{k+1}}{f_k(x, u_k)}, u_k, x) + V(k+1, \overset{x_{k+1}}{f_k(x, u_k)}) \right] \quad (4)$$

- If the optimal control problem admits a solution $u^* = \{\gamma_k^*, k \in \mathbf{K}\}$, then the solution $V(1, x)$ of Eq.(4) should be equal to $L(u^*)$
- Each u_k^* should be determined as an argument of the RHS of Eq.(4) (**single-shot optimization**)

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Infinite horizon linear-quadratic problem (Discrete Time Optimal Control)

- discrete-time system $x_{t+1} = Ax_t + Bu_t$, $x_0 = x^{init}$
- problem: choose u_0, u_1, \dots to minimize

$$J = \sum_{\tau=0}^{\infty} (x_{\tau}^T Q x_{\tau} + u_{\tau}^T R u_{\tau})$$

with given constant state and input weight matrices

$$Q = Q^T \geq 0, \quad R = R^T > 0$$

- ✓ this is an infinite dimensional problem

Infinite horizon linear-quadratic problem (Discrete Time Optimal Control)

- **Problem:** it's possible that $J = \infty$ for all input sequences u_0, \dots

$$x_{t+1} = 2x_t + 0u_t, \quad x^{init} = 1$$

- Let's assume (A, B) is controllable
- then for any x^{init} there's an input sequence

$$u_0, \dots, u_{n-1}, 0, 0, \dots$$

that steers x to zero at $t = n$, and keeps it there

- ✓ For this u , $J < \infty$
- ✓ And therefore, $\min_u J < \infty$ for any x^{init}

Infinite horizon linear-quadratic problem (Discrete Time Optimal Control)

- To apply dynamic programming approach define value function $V : \mathbf{R}^n \rightarrow \mathbf{R}$

$$V(z) = \min_{u_0, \dots} \sum_{\tau=0}^{\infty} (x_{\tau}^T Q x_{\tau} + u_{\tau}^T R u_{\tau})$$

subject to $x_0 = z, \quad x_{\tau+1} = Ax_{\tau} + Bu_{\tau}$

- $V(z)$ is the minimum LQR cost-to-go, starting from state z
- doesn't depend on time-to-go, which is always ∞ ; infinite horizon problem is *shift invariant*

Infinite horizon linear-quadratic problem (Discrete Time Optimal Control)

- **Fact:** V is quadratic, i.e., $V(z) = z^T P z$, where $P = P^T \geq 0$
- Applying Bellman minimum principle (or HJB equation for discrete system)

$$V(z) = \min_w (z^T Q z + w^T R w + V(Az + Bw))$$

or

$$z^T P z = \min_w (z^T Q z + w^T R w + (Az + Bw)^T P (Az + Bw))$$

- minimizing $w^* = -(R + B^T P B)^{-1} B^T P A z$
- so HJB equation is

$$\begin{aligned} z^T P z &= z^T Q z + w^{*T} R w^* + (Az + Bw^*)^T P (Az + Bw^*) \\ &= z^T (Q + A^T P A - A^T P B (R + B^T P B)^{-1} B^T P A) z \end{aligned}$$

- ✓ This must hold for all z , so we can conclude that P satisfies the **ARE**

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

- The optimal input is constant state feedback $u_t = K x_t$

$$K = -(R + B^T P B)^{-1} B^T P A$$

Infinite horizon linear-quadratic problem (Discrete Time Optimal Control)

- **Fact:** the ARE has only one positive semidefinite solution P
 - ✓ ARE plus $P = P^T \geq 0$ uniquely characterizes value function
- Consequence: The Riccati recursion

$$P_{k+1} = Q + A^T P_k A - A^T P_k B (R + B^T P_k B)^{-1} B^T P_k A, \quad P_1 = Q$$

converges to the unique PSD solution of the ARE (when (A,B) controllable)

- Infinite-horizon LQR optimal control is same as steady-state finite horizon optimal control

Continuous-Time Optimal Control Problems

Consider the following optimal control problem defined by

$$L(u) = \int_0^T g(t, x(t), u(t)) dt + q(T, x(T)) \quad (5)$$

where the state variable $x(t)$ satisfies the differential equation:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0, \quad t \geq 0 \quad (6)$$

- $u(t) = \gamma(t, x(t)) \in U$,
✓ $\gamma \in \Gamma$ is the class of all admissible feedback strategies

Dynamic Programming for Continuous-Time Optimal Control Problems

- The minimum cost-to-go from any initial state x and any initial time t is described by the so-called **value function** defined by

$$V(t, x) = \min_{u(s), t \leq s \leq T} \left[\int_t^T g(s, x(s), u(s)) ds + q(T, x(T)) \right] \quad (7)$$

Satisfying the boundary condition

$$V(T, x) = q(T, x(T)) \quad (8)$$

Hamilton-Jacobi-Bellman (HJB) equation

- The dynamic programming approach, when applied to optimal control problems defined in continuous time, leads to a partial differential equation (PDE) which is known as the **Hamilton-Jacobi-Bellman (HJB) equation**.
- A direct application of **the principle of optimality** on Eq. (7), under the assumption of continuous differentiability of V , leads to the HJB equation

$$-\frac{\partial V(t, x)}{\partial t} = \min_u \left[\frac{\partial V(t, x)}{\partial x} f(t, x, u) + g(t, x, u) \right] \quad (9)$$

with $V(T, x) = q(T, x(T))$ boundary condition

- In general, it is not easy to compute $V(t, x)$ and the continuous differentiability assumption is rather restrictive.
- If $V(t, x)$ exists, the HJB equation provides a means of obtaining the optimal control strategy **(sufficient condition)**

Proof of Hamilton-Jacobi-Bellman (HJB) equation

Proof:

According to Bellman's optimality principle, the following identity holds for small δ

$$V(t, x(t)) = \min_{u \in U} [g(t, x, u)\delta + V(t + \delta, x(t + \delta))]$$

where $V(t + \delta, x(t + \delta)) = V(t + \delta, x(t) + f(t, x, u) \cdot \delta)$

$$= V(t, x(t)) + \frac{\partial V(t, x)}{\partial x} f(t, x, u) \cdot \delta + \frac{\partial V(t, x)}{\partial t} \delta + o(\delta)$$

Substituting into (3) gives

$$V(t, x(t)) = \min_{u \in U} \left[g(t, x, u) \cdot \delta + V(t, x(t)) + \frac{\partial V(t, x)}{\partial x} f(t, x, u) \cdot \delta + \frac{\partial V(t, x)}{\partial t} \delta + o(\delta) \right]$$

Canceling $V(t, x(t))$ both side and divide by δ gives:

$$-\frac{\partial V(t, x)}{\partial t} = \min_u \left[\frac{\partial V(t, x)}{\partial x} f(t, x, u) + g(t, x, u) \right]$$

Obtaining the optimal control strategy from HJB equation

Theorem

If a continuously differentiable function $V(t, x)$ can be found that satisfies the HJB equation (9), then it generate the optimal strategy through the static (pointwise) minimization problem defined by the RHS of (9)

$$-\frac{\partial V(t, x)}{\partial t} = \min_{u \in U} \left[\frac{\partial V(t, x)}{\partial x} f(t, x, u) + g(t, x, u) \right] \quad (9)$$



$$u^*(t, x) = \operatorname{argmin}_{u \in U} \left[\frac{\partial V(t, x)}{\partial x} f(t, x, u) + g(t, x, u) \right] \quad (10)$$

Obtaining the optimal control strategy from HJB equation

Proof:

- If we are given two strategies, $\gamma^* \in \Gamma$ (the optimal one) and $\gamma \in \Gamma$ (an arbitrary one), with the corresponding terminating trajectories x^* and x , and terminal times T^* and T , respectively, then Eq. (9) reads

$$g(t, x, u) + \frac{\partial V(t, x)}{\partial x} f(t, x, u) + \frac{\partial V(t, x)}{\partial t} \geq 0 \quad (11)$$

$$g(t, x^*, u^*) + \frac{\partial V(t, x^*)}{\partial x} f(t, x^*, u^*) + \frac{\partial V(t, x^*)}{\partial t} \equiv 0 \quad (12)$$

where γ^* and γ have been replaced by the corresponding controls u^* and u , respectively.

- Integrating Eq.(11) from 0 to T and Eq.(12) from 0 to T^* leads to

$$\begin{aligned} \int_0^T g(t, x, u) + V(T, x(T)) - V(0, x_0) &\geq 0 \\ \int_0^{T^*} g(t, x^*, u^*) + V(T^*, x^*(T^*)) - V(0, x_0) &= 0 \end{aligned}$$

- Elimination of $V(0, x_0)$ yields

$$\int_0^T g(t, x, u) + q(T, x(T)) \geq \int_0^{T^*} g(t, x^*, u^*) + q(T^*, x^*(T^*)) \geq 0$$

➤ u^* is the optimal control (sequence of actions), and therefore γ^* is the optimal strategy

Infinite horizon linear-quadratic problem (Continuous Time Optimal Control)

- continuous-time system $\dot{x}(t) = Ax(t) + Bu(t)$
- problem: choose $u(t)$ to minimize

$$J(u) = \int_0^T (x(t)^T Q x(t) + u(t)^T R u(t)) dt + x(T)^T Q_f x(T)$$

- ✓ $g = x^T Q x + u^T R u$
- ✓ $f = Ax + Bu$
- ✓ Let's assume $V(t, x) = x^T P_t x$

- HJB equation is employed

$$-\frac{\partial V(t, x)}{\partial t} = \min_{u \in U} \left[\frac{\partial V(t, x)}{\partial x} f(t, x, u) + g(t, x, u) \right]$$

$$\Rightarrow -x \dot{P}_t x = \min_u \{ 2P_t x (Ax + Bu) + x^T Q x + u^T R u \}$$

- minimizing over u yields $u^* = -R^{-1} B^T P_t$, and substitute to HJB;

$$-\dot{P}_t = A^T P_t + P_t A - P_t B R^{-1} B^T P_t + Q$$

which is called as **Riccati differential equation** in optimal control.

The Minimum Principle

- Dynamic optimization using Dynamic Programming approach → HJB equation
 - Handy to use
 - Easy to extend to stochastic feedback control
- Dynamic optimization using the Lagrangian Method → Minimum Principle
 - Intuitively show how dynamic optimization is solved using static optimization technique
 - Will be used to state the equilibrium condition (necessary conditions) for various dynamic game

From Static Optimization to Dynamic Optimization

- We derive the first-order necessary conditions for the basic dynamic optimization problem to find a control function $u(\cdot)$ that minimizes the cost functional

$$J(u) = \int_0^T g(t, x(t), u(t)) dt + h(x(T)) \quad \text{where } h(x(T)) = q(T, x(T))$$

where the state variable $x(t)$ satisfies the differential equation:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0$$

- ✓ $x(t) \in \mathbf{R}^n, u(t) \in \mathbf{R}^m$
- ✓ It is assumed that the dynamic process starts from $t = 0$ and ends at the fixed terminal time $T > 0$.
- ✓ $f(t, x, u)$ and $g(t, x, u)$ are continuous functions on \mathbf{R}^{n+m+1}
- ✓ The set of admissible control functions $u(t) \in U$ consists of the set of functions that are continuous on $[0, T]$.

The Euler-Lagrange Equation

- Inspired by the theory of static optimization we introduce for each t in the interval $[0, T]$ the quantity $\lambda(t)[f(t, x(t), u(t)) - \dot{x}(t)]$ that satisfies

$$\int_0^T \lambda(t)[f(t, x(t), u(t)) - \dot{x}(t)]dt = 0$$

- ✓ the Lagrange multiplier $\lambda(t)$ (or costate variable) is an arbitrarily chosen row vector
- Define the new cost function \bar{J} which coincides with the original cost function J if the dynamic constraint, $\dot{x}(t) = f(t, x(t), u(t))$, is satisfied as

$$\begin{aligned}\bar{J} &:= \int_0^T \{g(t, x, u) + \lambda(t)f(t, x, u) - \lambda(t)\dot{x}(t)\}dt + h(x(T)) \\ &= \int_0^T \{H(t, x, u, \lambda) - \lambda(t)\dot{x}(t)\}dt + h(x(T))\end{aligned}$$

- ✓ Where the Hamiltonian function $H(t, x, u, \lambda) = g(t, x, u) + \lambda(t)f(t, x, u)$
- Conducting integration by parts, \bar{J} can be rewritten as

$$\bar{J} := \underbrace{\int_0^T \{H(t, x, u, \lambda) + \dot{\lambda}(t)x(t)\}dt}_{\bar{J}_1} + \underbrace{h(x(T)) - \lambda(T)x(T)}_{\bar{J}_2} + \underbrace{\lambda(0)x_0}_{\bar{J}_3}$$

The Euler-Lagrange Equation

- Assume that $u^*(t) \in U$ is an optimal control path generating the minimum value of \bar{J} and $x^*(t)$ is the corresponding optimal state trajectory
- If we perturb this optimal $u^*(t)$ path with a continuous perturbing curve $p(t)$, we can generate ‘neighboring’ control paths

$$u(t) = u^*(t) + \epsilon p(t)$$

✓ ϵ is small scalar

✓ $u(t)$ induces a corresponding ‘neighboring’ state trajectory $x(t, \epsilon, p)$ with $t \in [0, T]$

- The cost function associated with the perturbation becomes

$$\bar{J}(\epsilon) := \int_0^T \{H(t, x(t, \epsilon, p), u^* + \epsilon p, \lambda) + \dot{\lambda}(t)x(t, \epsilon, p),\} dt + h(x(T, \epsilon, p), \lambda(T)) - \lambda(T)x(T, \epsilon, p) + \lambda(0)x_0$$

- By assumption $\bar{J}(\epsilon)$ has a minimum at $\epsilon = 0 \rightarrow \frac{d\bar{J}(\epsilon)}{d\epsilon} = 0$ at $\epsilon = 0$

The Euler-Lagrange Equation

- Evaluating the derivative of $\bar{J}(\epsilon)$ yields

$$\frac{d\bar{J}(\epsilon)}{d\epsilon} = \int_0^T \left\{ \frac{\partial H}{\partial x} \frac{dx(t, \epsilon, p)}{d\epsilon} + \frac{\partial H}{\partial u} p(t) + \lambda(t) \frac{dx(t, \epsilon, p)}{d\epsilon} \right\} dt + \frac{\partial h(x(T))}{\partial x} \frac{dx(T, \epsilon, p)}{d\epsilon} - \lambda(t) \frac{dx(T, \epsilon, p)}{d\epsilon}$$

$$\frac{d\bar{J}(\epsilon)}{d\epsilon} = \int_0^T \left\{ \left(\frac{\partial H}{\partial x} + \lambda(t) \right) \frac{dx(t, \epsilon, p)}{d\epsilon} + \frac{\partial H}{\partial u} p(t) \right\} dt + \left(\frac{\partial h(x(T))}{\partial x} - \lambda(T) \right) \frac{dx(T, \epsilon, p)}{d\epsilon}$$

✓ $\frac{d\bar{J}(\epsilon)}{d\epsilon} = 0$ at $\epsilon = 0$ if we choose $\lambda(t)$

$$\frac{\partial H(t, x^*, u^*, \lambda)}{\partial x} + \lambda(t) = 0 \text{ with } \frac{\partial h(x(T))}{\partial x} - \lambda(T)$$

✓ $\frac{d\bar{J}(\epsilon)}{d\epsilon} = 0$ at $\epsilon = 0$ if and only if

$$\int_0^T \frac{\partial H(t, x^*, u^*, \lambda)}{\partial u} p(t) dt = 0$$

- This should hold for any continuous function $p(t)$ on $[0, T]$. Choosing $p(t) = \frac{\partial H^T(t, x^*, u^*, \lambda)}{\partial u}$ shows that another necessary condition

$$\frac{\partial H(t, x^*, u^*, \lambda)}{\partial u} = 0$$

The Euler-Lagrange Equation

- We derive the first-order necessary conditions for the basic dynamic optimization problem to find a control function $u(\cdot)$ that minimizes the cost functional

$$J(u) = \int_0^T g(t, x(t), u(t)) dt + h(x(T)) \quad \text{where } h(x(T)) = q(T, x(T))$$

where the state variable $x(t)$ satisfies the differential equation:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0$$

Theorem

If $u^*(t) \in U$ is a control that yields a local minimum for the cost function above and $x^*(t)$ and $\lambda^*(t)$ are the corresponding state and costate, then it is necessary that

$$\begin{aligned} \dot{x}(t) &= f(t, x^*, u^*) \left(= \frac{\partial H(t, x^*, u^*, \lambda)}{\partial \lambda} \right), & x^*(0) &= x_0 \\ \dot{\lambda}^*(t) &= \frac{\partial H(t, x^*, u^*, \lambda^*)}{\partial x}; & \lambda^*(T) &= \frac{\partial h(x^*(T))}{\partial x} \\ \frac{\partial H(t, x^*, u^*, \lambda^*)}{\partial u} &= 0 \end{aligned}$$

- The dynamic optimization problem has been reduced to a static optimization problem which should hold at every single instant of time

From the Euler Lagrange to Minimum Principle (Generalization)

- **Euler-Lagrange method** assumes that
 - ✓ $u(t)$ could (in principle) to be chosen arbitrarily in R^m
 - ✓ $u(t)$ is continuous for all $t \in [0, T]$
- However, these assumptions are too restrictive, thus a more general problem setting is required
- **Pontryagin Minimum principle** assumes that
 - ✓ There is a subset $U \subset R^m$ such that for all $t \in [0, T], u(t) \in U$
 - ✓ This is the next generalization of the Euler-Lagrange method

The Minimum Principle

Consider the following optimal control problem defined by

$$L(u) = \int_0^T g(t, x(t), u(t)) dt + q(T, x(T))$$

where the state variable $x(t)$ satisfies the differential equation:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0, \quad t \geq 0$$

Theorem (The minimum principle)

In the continuous time dynamic system defined by equation (1) and (2), optimal control $u^*(t)$ and corresponding trajectory $x^*(t)$ satisfy following equations:

$$H(t, \lambda, x, u) := g(t, x, u) + \lambda(t)f(t, x, u)$$

$$\dot{x}^*(t) = f(t, x^*, u^*) \left(= \frac{\partial H(t, x^*, u^*, \lambda)}{\partial \lambda} \right), x^*(0) = x_0;$$

$$\dot{\lambda}^*(t) = - \frac{\partial H(t, \lambda, x^*, u^*)}{\partial x}; \lambda^*(T) = \frac{\partial h(x^*(T))}{\partial x}$$

$$u^*(t) = \operatorname{argmin}_{u \in U} H(t, \lambda^*, x^*, u)$$

Proof of The Minimum Principle

Proof)

We start with Hamilton-Jacobi-Bellman equation,

$$\nabla_t V(t, x) + \min_{u \in U} [\nabla_x V(t, x) f(t, x, u) + g(t, x, u)] = 0$$

Let $u^* = \operatorname{argmin}_{u \in U} [\nabla_x V(t, x) \cdot f(t, x, u) + g(t, x, u)]$ then

$$0 = \nabla_x V(t, x) f(t, x, u^*) + g(t, x, u^*) + \nabla_t V(t, x)$$

Derivatives w.r.t. x and t yields

$$0 = \nabla_{xx}^2 V(t, x) f(t, x, u^*) + \nabla_x V(t, x) \nabla_x f(t, x, u^*) + \nabla_x g(t, x, u^*) + \nabla_{xt}^2 V(t, x) \cdots (*)$$

$$0 = \nabla_{xt}^2 V(t, x) \cdot f(t, x, u^*) + \nabla_{tt}^2 V(t, x) \cdots (**)$$

(*) and (**) holds for all (t, x) .

Let us specialize them along an optimal state and trajectory $x^*(t), u^*(t)$, then we have

Proof of The Minimum Principle

Proof)

$$0 = \nabla_{xx}^2 V(t, x) f(t, x, u^*) + \nabla_x V(t, x) \nabla_x f(t, x, u^*) + \nabla_x g(t, x, u^*) + \nabla_{xt}^2 V(t, x) \cdots (*)$$

$$0 = \nabla_{xt}^2 V(t, x) \cdot f(t, x, u^*) + \nabla_{tt}^2 V(t, x) \cdots (**)$$

the term $\nabla_{xx}^2 V(t, x^*) f(t, x^*, u^*) + \nabla_{xt}^2 V(t, x^*)$ in (*) is equal the following total derivative w.r.t. t

$$\frac{d}{dt} (\nabla_x V(t, x^*)), \quad (\text{let } \nabla_x V(t, x^*) := p(t))$$

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

Similarly, the term $\nabla_{xt}^2 V(t, x^*) f(t, x^*, u^*) + \nabla_{tt}^2 V(t, x^*)$ in (**) is equal the total derivative

$$\frac{d}{dt} (\nabla_t V(t, x^*)), \quad (\text{let } \nabla_t V(t, x^*) := p_0(t))$$

$$\text{and } (**) \text{ becomes } \dot{p}_0(t) = 0 \rightarrow p_0(t) = \text{constant}$$

We have boundary condition

$$p(T) = \nabla_x V(T(x^*), x^*) = \nabla_x q(T(x^*), x^*)$$

Proof of The Minimum Principle

Proof)

So, the original HJB $u^* = \operatorname{argmin}_u [\nabla_x V(t, x) f(t, x, u) + g(t, x, u)]$ become

$$u^* = \operatorname{argmin}_{u \in U} [p(t) \cdot f(t, x, u) + g(t, x, u)]$$

Define $H(x, u, p) := p(t)f(t, x, u) + g(t, x, u)$, then system and ad joint equations can be written in terms of Hamiltonian

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

■

Comments (Minimum principle v.s. HJB)

- When satisfied along a trajectory, Pontryagin's minimum principle is a ***necessary condition***, *not sufficient contrition*, for an optimum.
- The Hamilton–Jacobi–Bellman equation provides a ***necessary and sufficient*** condition for an optimum, but this condition must be satisfied over the whole of the state space.
- While the Hamilton-Jacobi-Bellman equation admits a straightforward extension to stochastic optimal control problems, the minimum principle does not
 - DP not only solves the original optimization problem, but tells us even ***at arbitrary times and arbitrary states*** what the best action is
 - Having the optimal policy available as a function of times and state (“**feedback**”) is often viewed as even better than having it available only as a function of time (“**open-loop form**”)
- The Minimum Principle requires solving an ODE with split boundary conditions. It is not trivial to solve, but easier than solving a PDE in the HJB.