ECE276B: Planning & Learning in Robotics
Lecture 13: Pontryagin's Minimum Principle

Lecturer:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Tianyu Wang: tiw161@eng.ucsd.edu

Yongxi Lu: yol070@eng.ucsd.edu

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Locally Extremal Trajectories

- **Deterministic continuous-time optimal control**:

$$\begin{aligned} \min_{\pi \in PC^0([t_0,T],\mathcal{U})} \quad & J^\pi(t_0, x_0) := \int_{t_0}^{T} g(x(t), \pi(t, x(t)))dt + g_T(x(T)) \\ \text{s.t.} \quad & \dot{x}(t) = f(x(t), u(t)), \;\; x(t_0) = x_0 \\ & x(t) \in \mathcal{X},\; \pi(t, x(t)) \in \mathcal{U} \end{aligned}$$

- **Hamiltonian**: $H(x, u, p) := g(x, u) + p^T f(x, u)$
- **Relationship to Mechanics**:
  - **Hamilton's principle of least action**: trajectories of mechanical systems are extremals of the action integral $\int_{t_0}^{T} L(t)dt$, where the Lagrangian $L(t) := K(t) - P(t)$ is the difference between kinetic and potential energy.
  - If we think of the stage cost as the Lagrangian of a mechanical system, the Hamiltonian is the total energy (kinetic plus potential) of the system
- We can compute **extremal open-loop trajectories** (i.e., local minima) by solving a boundary-value ODE problem with given $x(0)$ and **costate** $p(T) = \nabla_x g_T(x)$, where $p(t)$ is the gradient/sensitivity of the optimal cost-to-go with respect to the state x.

Pontryagin's Minimum Principle (PMP)

- **Hamiltonian**: $H(x,u,p) := g(x,u) + p^T f(x,u)$

Theorem: Pontryagin's Minimum Principle

- Let $u^*(t) : [t_0, T] \to \mathcal{U}$ be an optimal control trajectory
- Let $x^*(t) : [t_0, T] \to \mathcal{X}$ be the associated state trajectory from x_0
- Then, there exists a **costate trajectory** $p^*(t) : [t_0, T] \to \mathcal{X}$ satisfying:
  1. **Canonical equations with boundary conditions**:

$$\begin{aligned} \dot{x}^*(t) &= \nabla_p H(x^*(t), u^*(t), p^*(t)), \quad & x^*(t_0) &= x_0 \\ \dot{p}^*(t) &= -\nabla_x H(x^*(t), u^*(t), p^*(t)), \quad & p^*(T) &= \nabla_x g_T(x^*(T)) \end{aligned}$$

  2. **Minimum principle with constant (holonomic) constraint**:

$$\begin{aligned} u^*(t) &= \arg\min_{u \in \mathcal{U}} H(x^*(t), u, p^*(t)), \qquad \forall t \in [t_0, T] \\ H(x^*(t), u^*(t), p^*(t)) &= constant, \quad \forall t \in [t_0, T] \end{aligned}$$

- **Proof**: Liberzon, Calculus of Variations and Optimal Control Theory, Ch. 4.2

Proof of PMP (Step 0: Preliminaries)

First Order Necessary Condition for Optimality

Let f be a continuously differentiable function on $\mathbb{R}^m$ and $\mathcal{U} \subseteq \mathbb{R}^m$ be a convex set. If u^* is a minimizer of $\min_{u \in \mathcal{U}} f(u)$, then:

$$\nabla f(u^*)^T (v - u^*) \geq 0, \qquad \forall v \in \mathcal{U}$$

- **Proof**: Suppose $\exists w \in \mathcal{U}$ with $\nabla f(u^*)^T (w - u^*) < 0$. Consider $z(\lambda) := \lambda w + (1 - \lambda) u$ for $\lambda \in [0, 1]$. Since $\mathcal{U}$ is convex, $z(\lambda) \in \mathcal{U}$ and

$$\frac{d}{d\lambda} f(z(\lambda))\bigg|_{\lambda=0} = \nabla f(u^*)^T (w - u^*) < 0$$

implies that $f(z(\lambda)) < f(u^*)$ for small λ, which contradicts that u^* is optimal.

Proof of PMP (Step 0: Preliminaries)

Lemma: ∇-min Exchange

Let $F(t, x, u)$ be a cont.-diffable function of $t \in \mathbb{R}$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and let $\mathcal{U} \subseteq \mathbb{R}^m$ be a convex set. Furthermore, assume $\pi^*(t, x) = \arg\min_{u \in \mathcal{U}} F(t, x, u)$ exists and is cont.-diffable. Then, for all t and x:

$$\frac{\partial \left(\min_{u\in\mathcal{U}} F(t,x,u)\right)}{\partial t} = \left.\frac{\partial F(t,x,u)}{\partial t}\right|_{u=\pi^*(t,x)} \qquad \nabla_x \left(\min_{u\in\mathcal{U}} F(t,x,u)\right) = \left.\nabla_x F(t,x,u)\right|_{u=\pi^*(t,x)}$$

- **Proof**: Let $G(t, x) := \min_{u\in\mathcal{U}} F(t, x, u) = F(t, x, \pi^*(t, x))$. Then:

$$\frac{\partial G(t,x)}{\partial t} = \left.\frac{\partial F(t,x,u)}{\partial t}\right|_{u=\pi^*(t,x)} + \underbrace{\left.\frac{\partial F(t,x,u)}{\partial u}\right|_{u=\pi^*(t,x)} \frac{\partial \pi^*(t,x)}{\partial t}}_{\substack{=0 \text{ since } \nabla_u F(t,x,\pi^*)(\pi^*(t+\epsilon,x)-\pi^*(t,x))\geq 0 \\ \text{for all } \epsilon \text{ by 1st order optimality condition}}}$$

A similar derivation can be used for the partial derivative wrt x.

Proof of PMP (Step 1: HJB PDE gives $J^*(t,x)$)

- **Extra Assumption**: We prove the PMP under the assumption that $J^*(t,x)$ and $\pi^*(t,x)$ are cont-diffable in t and x and $\mathcal{U}$ is convex. These assumptions can be avoided in a more general proof.
- With cont-diffable cost-to-go, the HJB PDE is also a necessary condition for optimality:

$$J^*(T,x) = g_T(x), \quad \forall x \in \mathcal{X}$$
$$0 = \min_{u \in \mathcal{U}} \underbrace{\left(g(x,u) + \frac{\partial}{\partial t} J^*(t,x) + \nabla_x J^*(t,x)^T f(x,u) \right)}_{:=F(t,x,u)}, \quad \forall t \in [t_0, T], x \in \mathcal{X}$$

with $\pi^*(t,x)$ a corresponding optimal policy.

Proof of PMP (Step 2: ∇-min Exchange Lemma)

- Apply the ∇-min Exchange Lemma to the HJB PDE:

$$0 = \frac{\partial}{\partial t}\left(\min_{u \in \mathcal{U}} F(t,x,u)\right) = \frac{\partial^2 J^*(t,x)}{\partial t^2} + \left[\frac{\partial}{\partial t}\nabla_x J^*(t,x)\right]^T f(x, \pi^*(t,x))$$

$$\begin{aligned} 0 &= \nabla_x \left(\min_{u \in \mathcal{U}} F(t,x,u)\right) \\ &= \nabla_x g(x,u^*) + \nabla_x \frac{\partial J^*(t,x)}{\partial t} + [\nabla_x^2 J^*(t,x)] f(x,u^*) + [\nabla_x f(x,u^*)]^T \nabla_x J^*(t,x) \end{aligned}$$

where $u^* := \pi^*(t,x)$

- Evaluate these along the trajectory $x^*(t)$ resulting from $\pi^*(t, x^*(t))$:

$$\dot{x}^*(t) = f(x^*(t), u^*(t)) = \nabla_p H(x^*(t), u^*(t), p)^T, \qquad x^*(0) = x_0$$

Proof of PMP (Step 3: Evaluate along $x^*(t), u^*(t)$)

- Evaluate the results of Step 2 along $x^*(t)$:

$$0 = \left.\frac{\partial^2 J^*(t,x)}{\partial t^2}\right|_{x=x^*(t)} + \left[\frac{\partial}{\partial t}\nabla_x J^*(t,x)\Big|_{x=x^*(t)}\right]^T \dot{x}^*(t)$$

$$= \frac{d}{dt}\left(\underbrace{\left.\frac{\partial J^*(t,x)}{\partial t}\right|_{x=x^*(t)}}_{:=r(t)}\right) = \frac{d}{dt}r(t) \Rightarrow r(t) = const.\forall t$$

and

$$0 = \nabla_x g(x,u^*)|_{x=x^*(t)} + \frac{d}{dt}\left(\underbrace{\nabla_x J^*(t,x)|_{x=x^*(t)}}_{=:p^*(t)}\right) + [\nabla_x f(x,u^*)|_{x=x^*(t)}]^T[\nabla_x J^*(t,x)|_{x=x^*(t)}]$$

$$= \nabla_x g(x,u^*)|_{x=x^*(t)} + \dot{p}^*(t) + [\nabla_x f(x,u^*)|_{x=x^*(t)}]^T p^*(t)$$

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

Proof of PMP (Step 4: Done)

- The boundary condition $J^*(T,x) = g_T(x)$ implies that $\nabla_x J^*(T,x) = \nabla_x g_T(x)$ for all $x \in \mathcal{X}$ and thus $p^*(T) = \nabla_x g_T(x^*(T))$
- From the HJB PDE we have:
$$-\frac{\partial J^*(t,x)}{\partial t} = \min_{u \in \mathcal{U}} H(x, u, \nabla_x J^*(t, \cdot))$$
which along the optimal trajectory $x^*(t), u^*(t)$ becomes:
$$-r(t) = H(x^*(t), u^*(t), p^*(t)) = const$$
- Finally, note that
$$\begin{aligned}
u^*(t) &= \arg\min_{u \in \mathcal{U}} F(t, x^*(t), u) \\
&= \arg\min_{u \in \mathcal{U}} \left\{ g(x^*(t), u) + [\nabla_x J^*(t,x)|_{x=x^*(t)}]^T f(x^*(t), u) \right\} \\
&= \arg\min_{u \in \mathcal{U}} \left\{ g(x^*(t), u) + p^*(t)^T f(x^*(t), u) \right\} \\
&= \arg\min_{u \in \mathcal{U}} H(x^*(t), u, p^*(t))
\end{aligned}$$

HJB PDE vs PMP

- The HJB PDE provides a lot of information – the optimal cost-to-go and an optimal policy for all time and all states!
- Often, we only care about the optimal trajectory for a specific initial condition x_0. Exploiting that we need less information, we can arrive at simpler conditions for optimality – Pontryagin's Minimum Principle
- The PMP does **not apply to infinite horizon problems**, so one has to use the HJB equations in that case
- The HJB PDE is a **sufficient condition** for optimality (it is possible that the optimal solution does not satisfy it but a solution that satisfies it is guaranteed to be optimal)
- The PMP is a **necessary condition** for optimality (it is possible that non-optimal trajectories satisfy it) so further analysis is necessary to determine if the candidate PMP policy is optimal
- The PMP requires solving an ODE with split boundary conditions (not easy but easier than the nonlinear HJB PDE!)

Example: Resource Allocation for a Martian Base

- A fleet of reconfigurable, general purpose robots is sent to Mars at $t = 0$
- The robots can 1) replicate or 2) make human habitats
- The number of robots at time t is $x(t)$, while the number of habitats is $z(t)$ and they evolve according to:

$$\begin{aligned}
\dot{x}(t) &= u(t)x(t), \quad x(0) = x > 0 \\
\dot{z}(t) &= (1 - u(t))x(t), \quad z(0) = 0 \\
0 \leq u(t) &\leq 1
\end{aligned}$$

  where $u(t)$ denotes the percentage of the $x(t)$ robots used for replication

- Goal: Maximize the size of the Martian base by a terminal time T, i.e.:

$$\max\ z(T) = \int_0^T (1 - u(t))x(t)dt$$

  with $f(x, u) = ux$, $g(x, u) = (1 - u)x$ and $g_T(x) = 0$

Example: Resource Allocation for a Martian Base

- Hamiltonian: $H(x, u, p) = (1 - u)x + pux$
- Apply the PMP:

$$\dot{x}^*(t) = \nabla_p H(x^*, u^*, p^*) = x^*(t)u^*(t), \quad x^*(0) = x$$
$$\dot{p}^*(t) = -\nabla_x H(x^*, u^*, p^*) = -1 + u^*(t) - p^*(t)u^*(t), \quad p^*(T) = 0$$
$$u^*(t) = \arg\max_{0 \leq u \leq 1} H(x^*(t), u, p^*(t)) = \arg\max_{0 \leq u \leq 1} (x^*(t) + x^*(t)(p^*(t) - 1)u)$$

- Since $x^*(t) > 0$ for $t \in [0, T]$:

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

Example: Resource Allocation for a Martian Base

- Work backwards from $t = T$ to determine $p^*(t)$:
  - Since $p^*(T) = 0$ for t close to T, we have $u^*(t) = 0$ and the costate dynamics become $\dot{p}^*(t) = -1$
  - At time $t = T - 1$, $p^*(t) = 1$ and the control input switches to $u^*(t) = 1$
  - For $t < T - 1$:

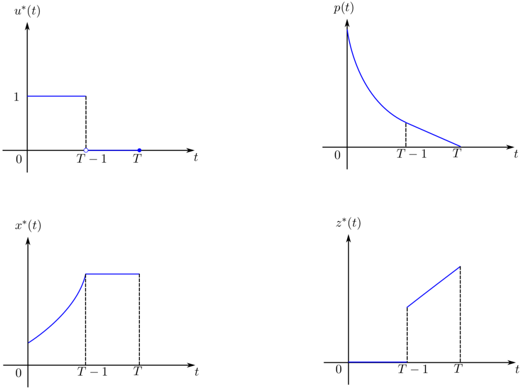
$$\begin{aligned} \dot{p}^*(t) &= -p^*(t), \quad p(T-1) = 1 \\ &\Rightarrow p^*(t) = e^{(T-1)-t} > 1 \text{ for } t < T - 1 \end{aligned}$$

- Optimal control:

$$u^*(t) = \begin{cases} 1 & \text{if } 0 \le t \le T - 1 \\ 0 & \text{if } T - 1 \le t \le T \end{cases}$$

Example: Resource Allocation for a Martian Base

- Optimal trajectories for the Martian resource allocation problem:



- **Conclusions**:
  - Use all robots to replicate themselves from $t = 0$ to $t = T - 1$ and then use all robots to build habitats
  - If $T < 1$, then the robots should only build habitats
  - If the Hamiltonian is linear in u, its min can only be attained on the boundary of $\mathcal{U}$, known as **bang-bang control**

PMP with Fixed Terminal State

- Suppose that in addition to $x(0) = x_s$, a final state $x(T) = x_\tau$ is given.
- The terminal cost $g_T(x(T))$ is not useful since $J^*(T, x) = \infty$ if $x(T) \neq x_\tau$. The terminal boundary condition for the costate $p(T) = \nabla_x g_T(x(T))$ does not hold but as compensation we have a different boundary condition $x(T) = x_\tau$.
- We still have $2n$ ODEs with $2n$ boundary conditions:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)), \qquad x(0) = x_s, x(T) = x_\tau \\ \dot{p}(t) &= -\nabla_x H(x(t), u(t), p(t))\end{aligned}$$

- If only some terminal state are fixed $x_j(T) = x_{\tau,j}$ for $j \in I$, then:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)), \qquad x(0) = x_s, \;\; x_j(T) = x_{\tau,j}, \;\; \forall j \in I \\ \dot{p}(t) &= -\nabla_x H(x(t), u(t), p(t)), \qquad p_j(T) = \frac{\partial}{\partial x_j} g_T(x(T)), \;\; \forall j \notin I\end{aligned}$$

PMP with Fixed Terminal Set

- **Terminal set**: a k dim surface in $\mathbb{R}^n$ requires:

$$x(T) \in \mathcal{X}_\tau = \{x \in \mathbb{R}^n \mid h_j(x) = 0, j = 1, \ldots, n-k\}$$

- The costate boundary condition requires that $p(T)$ is orthogonal to the tangent space $T_{x(T)}\mathcal{X}_\tau = \{d \in \mathbb{R}^n \mid \nabla_x h_j(x(T))^T d = 0, j = 1, \ldots, n-k\}$:

$$\begin{aligned} &\dot{x}(t) = f(x(t), u(t)), \qquad x(0) = x_s, && h_j(x(T)) = 0,\ j = 1, \ldots, n-k \\ &\dot{p}(t) = -\nabla_x H(x(t), u(t), p(t)), && p(T) \in \textbf{span}\{\nabla_x h_j(x(T)), \forall j\} \\ & && \text{OR} \quad d^T p(T) = 0, \forall d \in T_{x(T)}\mathcal{X}_\tau \end{aligned}$$

PMP with Free Initial State

- Suppose that x_0 is free and subject to optimization with additional cost $g_0(x_0)$ term
- The total cost becomes $g_0(x_0) + J(0, x_0)$ and the necessary condition for an optimal initial state x_0 is:

$$\nabla_x g_0(x)|_{x=x_0} + \underbrace{\nabla_x J(0,x)|_{x=x_0}}_{=p(0)} = 0 \quad \Rightarrow \quad p(0) = -\nabla_x g_0(x_0)$$

- We lose the initial state boundary condition but gain an adjoint state boundary condition:

$$\begin{aligned}
&\dot{x}(t) = f(x(t), u(t)) \\
&\dot{p}(t) = -\nabla_x H(x(t), u(t), p(t)), \;\; p(0) = -\nabla_x g_0(x_0), \;\; p(T) = -\nabla_x g_T(x(T))
\end{aligned}$$

- Similarly, we can deal with some parts of the initial state being free and some not

PMP with Free Terminal Time

- Suppose that the initial and/or terminal state are given but the terminal time T is free and subject to optimization

- We can compute the total cost of optimal trajectories for various terminal times T and look for the best choice, i.e.:

$$\frac{\partial}{\partial t} J^*(t,x)\bigg|_{t=T, x=x(T)} = 0$$

- Recall that on the optimal trajectory:

$$H(x^*(t), u^*(t), p^*(t)) = -\frac{\partial}{\partial t} J^*(t,x)\bigg|_{x=x^*(t)} = const. \quad \forall t$$

- Hence, in the free terminal time case, we gain an extra degree of freedom with free T but lose one degree of freedom by the constraint:

$$H(x^*(t), u^*(t), p^*(t)) = 0, \qquad \forall t \in [t_0, T]$$

PMP with Time-varying System and Cost

- Suppose that the system and stage cost vary with time:

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

- A usual trick is to convert the problem to a time-invariant one by making t part of the state. Let $y(t) = t$ with dynamics:

$$\dot{y}(t) = 1, \quad y(0) = t_0$$

- Augmented state $z(t) := (x(t), y(t))$ and system:

$$\dot{z}(t) = \bar{f}(z(t), u(t)) := \begin{bmatrix} f(x(t), u(t), y(t)) \\ 1 \end{bmatrix}$$

$$\bar{g}(z, u) := g(x, u, y) \quad \bar{g}_T(z) := g_T(x)$$

- The Hamiltonian need not to be constant along the optimal trajectory:

$$\begin{aligned}
H(x, u, p, t) &= g(x, u, t) + p^T f(x, u, t) \\
\dot{x}^*(t) &= f(x^*(t), u^*(t), t), & x^*(0) &= x_0 \\
\dot{p}^*(t) &= -\nabla_x H(x^*(t), u^*(t), p^*(t), t), & p^*(T) &= \nabla_x g_T(x^*(T)) \\
u^*(t) &= \arg\min_{u \in \mathcal{U}} H(x^*(t), u, p^*(t), t)
\end{aligned}$$

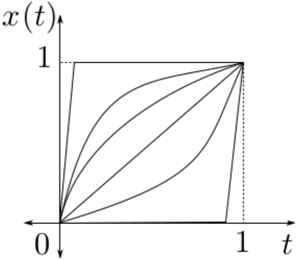
$$H(x^*(t), u^*(t), p^*(t), t) \neq const$$

Singular Problems

- **Singular Problems**: in some cases, the minimum condition $u(t) = \arg\min_{u \in \mathcal{U}} H(x^*(t), u, p^*(t), t)$ might be insufficient to determine $u^*(t)$ for all t because the values of $x^*(t)$ and $p^*(t)$ are such that $H(x^*(t), u, p^*(t), t)$ is independent of u over a nontrivial interval of time
- The optimal trajectories consist of portions where $u^*(t)$ can be determined from the minimum condition (**regular arcs**) and where $u^*(t)$ cannot be determined from the minimum condition since the Hamiltonian is independent of u (**singular arcs**)

Example: Fixed Terminal State

- System: $\dot{x}(t) = u(t),\ x(0) = 0,\ x(1) = 1,\ u(t) \in \mathbb{R}$
- Cost: $\min \frac{1}{2} \int_0^1 (x(t)^2 + u(t)^2) dt$
- Want $x(t)$ and $u(t)$ to be small but need to meet $x(1) = 1$



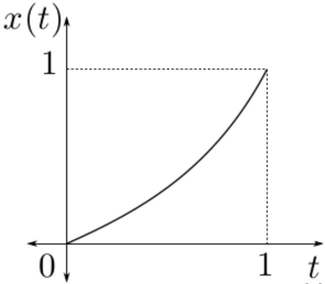
- Approach: use PMP to find a locally optimal open-loop policy

Example: Fixed Terminal State

- Pontryagin's Minimum Principle
  - Hamiltonian: $H(x,u,p) = \frac{1}{2}(x^2 + u^2) + pu$
  - Minimum principle: $u(t) = \arg\min_{u \in \mathbb{R}} \left\{ \frac{1}{2}(x(t)^2 + u^2) + p(t)u \right\} = -p(t)$
  - Canonical equations with boundary conditions:

$$\dot{x}(t) = \nabla_p H(x(t), u(t), p(t)) = u(t) = -p(t), \quad x(0) = 0, \; x(1) = 1$$
$$\dot{p}(t) = -\nabla_x H(x(t), u(t), p(t)) = -x(t)$$

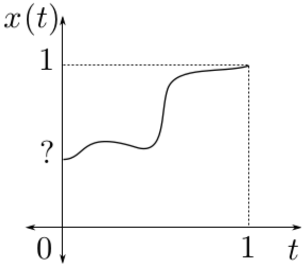
- Candidate trajectory: $\ddot{x}(t) = x(t) \quad \Rightarrow \quad x(t) = Ae^t + Be^{-t} = \frac{e^t - e^{-t}}{e - e^{-1}}$
  - $x(0) = 0 \quad \Rightarrow \quad A + B = 0$
  - $x(1) = 1 \quad \Rightarrow \quad Ae + Be^{-1} = 1$



- Open-loop control: $u(t) = \dot{x}(t) = \frac{e^t + e^{-t}}{e - e^{-1}}$

Example: Free Initial State

- System: $\dot{x}(t) = u(t)$, $x(0) = \textit{free}$, $x(1) = 1$, $u(t) \in \mathbb{R}$
- Cost: $\min \frac{1}{2} \int_0^1 (x(t)^2 + u(t)^2) dt$
- Picking $x(0) = 1$ will allow $u(t) = 0$ but we will accumulate cost due to $x(t)$. On the other hand, picking $x(0) = 0$ will accumulate cost due to $u(t)$ having to drive the state to $x(1) = 1$.



- Approach: use PMP to find a locally optimal open-loop policy

Example: Free Initial State

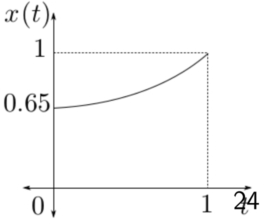
- Pontryagin's Minimum Principle
  - Hamiltonian: $H(x, u, p) = \frac{1}{2}(x^2 + u^2) + pu$
  - Minimum principle: $u(t) = \arg\min_{u \in \mathbb{R}} \left\{ \frac{1}{2}(x(t)^2 + u^2) + p(t)u \right\} = -p(t)$
  - Canonical equations with boundary conditions:

$$\dot{x}(t) = \nabla_p H(x(t), u(t), p(t)) = u(t) = -p(t), \quad x(1) = 1$$
$$\dot{p}(t) = -\nabla_x H(x(t), u(t), p(t)) = -x(t), \quad p(0) = 0$$

- Candidate trajectory:

$$\ddot{x}(t) = x(t) \quad \Rightarrow \quad x(t) = Ae^t + Be^{-t} = \frac{e^t + e^{-t}}{e + e^{-1}}$$
$$p(t) = -\dot{x} = -Ae^t + Be^{-t} = \frac{-e^t + e^{-t}}{e + e^{-1}}$$

  - $x(1) = 1 \quad \Rightarrow \quad Ae + Be^{-1} = 1$
  - $p(0) = 0 \quad \Rightarrow \quad -A + B = 0$
  - $x(0) \approx 0.65$
  - Open-loop control: $u(t) = \dot{x}(t) = \frac{e^t - e^{-t}}{e + e^{-1}}$



Example: Free Terminal Time

- System: $\dot{x}(t) = u(t),\ x(0) = 0,\ x(T) = 1,\ u(t) \in \mathbb{R}$
- Cost: $\min \int_0^T 1 + \frac{1}{2}(x(t)^2 + u(t)^2)dt$
- Free terminal time: $T = \mathit{free}$
- Note: if we do not include 1 in the stage-cost (i.e., use the same cost as before), we would get $T^* = \infty$ (see next slide for details)
- Approach: use PMP to find a locally optimal open-loop policy

Example: Free Terminal Time

- Pontryagin's Minimum Principle
  - Hamiltonian: $H(x(t), u(t), p(t)) = \frac{1}{2}(x(t)^2 + u(t)^2) + p(t)u(t)$
  - Minimum principle: $u(t) = \arg\min_{u \in \mathbb{R}} \left\{ \frac{1}{2}(x(t)^2 + u^2) + p(t)u \right\} = -p(t)$
  - Canonical equations with boundary conditions:

$$
\begin{aligned}
\dot{x}(t) &= \nabla_p H(x(t), u(t), p(t)) = u(t) = -p(t), \quad x(0) = 0, \; x(T) = 1 \\
\dot{p}(t) &= -\nabla_x H(x(t), u(t), p(t)) = -x(t)
\end{aligned}
$$

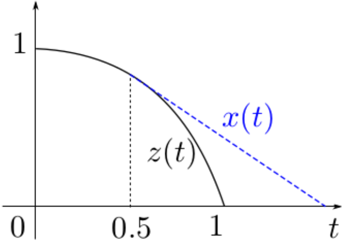
- Candidate trajectory: $\ddot{x}(t) = x(t) \quad \Rightarrow \quad x(t) = Ae^t + Be^{-t} = \frac{e^t - e^{-t}}{e^T - e^{-T}}$
  - $x(0) = 0 \quad \Rightarrow \quad A + B = 0$
  - $x(T) = 1 \quad \Rightarrow \quad Ae^T + Be^{-T} = 1$

- Free terminal time:

$$
\begin{aligned}
0 = H(x(t), u(t), p(t)) &= 1 + \frac{1}{2}(x(t)^2 - p(t)^2) \\
&= 1 + \frac{1}{2}\left(\frac{(e^t - e^{-t})^2 - (e^t + e^{-t})^2}{(e^T - e^{-T})^2} \right) = 1 - \frac{2}{(e^T - e^{-T})^2} \\
\Rightarrow \quad & T \approx 0.66
\end{aligned}
$$

Example: Time-varying Singular Problem

- System: $\dot{x}(t) = u(t),\ x(0) = \mathit{free},\ x(1) = \mathit{free},\ u(t) \in [-1, 1]$
- Time-varying cost: $\min \frac{1}{2}\int_0^1 (x(t) - z(t))^2 dt$ for $z(t) = 1 - t^2$
- Example feasible state trajectory that tracks the desired $z(t)$ until the slope of $z(t)$ becomes less than -1 and the input $u(t)$ saturates:



- Approach: use PMP to find a locally optimal open-loop policy

Example: Time-varying Singular Problem

- Pontryagin's Minimum Principle
  - Hamiltonian: $H(x, u, p, t) = \frac{1}{2}(x - z(t))^2 + pu$
  - Minimum principle:

$$u(t) = \arg\min_{|u|\leq 1} H(x(t), u, p(t), t) = \begin{cases} -1 & \text{if } p(t) > 0 \\ \text{undetermined} & \text{if } p(t) = 0 \\ 1 & \text{if } p(t) < 0 \end{cases}$$

  - Canonical equations with boundary conditions:

$$\begin{aligned} \dot{x}(t) &= \nabla_p H(x(t), u(t), p(t)) = u(t), \\ \dot{p}(t) &= -\nabla_x H(x(t), u(t), p(t)) = -(x(t) - z(t)), \quad p(0) = 0,\ p(1) = 0 \end{aligned}$$

- **Singular arc**: when $p(t) = 0$ for a non-trivial time interval, the control cannot be determined from PMP

- In this problem, the singular arc can be determined from the costate ODE:

$$0 \equiv \dot{p} = -x(t) + z(t) \quad \Rightarrow \quad u(t) = \dot{x}(t) = \dot{z}(t) = -2t \qquad \text{for } p(t) = 0$$

Example: Time-varying Singular Problem

- Since $p(0) = 0$, the state trajectory follows a singular arc until $t_s \leq \frac{1}{2}$ (since $u(t) = -2t \in [-1, 1]$) when it switches to a regular arc with $u(t) = -1$ (since $z(t)$ is decreasing and we are trying to track it).
- For $0 \leq t \leq t_s \leq \frac{1}{2}$: $\qquad x(t) = z(t) \qquad p(t) = 0$
- For $t_s < t \leq 1$:

$$\dot{x}(t) = -1 \quad \Rightarrow \quad x(t) = z(t_s) - \int_{t_s}^{t} ds = 1 - t_s^2 - t + t_s$$

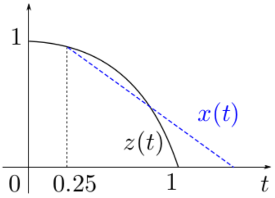
$$\dot{p}(t) = -(x(t) - z(t)) = t_s^2 - t_s - t^2 + t, \qquad p(t_s) = p(1) = 0$$

$$\Rightarrow p(s) = p(t_s) + \int_{t_s}^{s} (t_s^2 - t_s - t^2 + t)dt, \quad s \in [t_s, 1]$$

$$\Rightarrow 0 = p(1) = t_s^2 - t_s - \frac{1}{3} + \frac{1}{2} - t_s^3 + t_s^2 + \frac{t_s^3}{3} - \frac{t_s^2}{2}$$

$$\Rightarrow 0 = (t_s - 1)^2(1 - 4t_s)$$

$$\Rightarrow \boxed{t_s = \frac{1}{4}}$$



Discrete-time PMP

- Consider a discrete-time problem with dynamics $x_{t+1} = f(x_t, u_t)$
- Introduce Lagrange multipliers $p_{0:T}$ to relax the constraints:

$$
\begin{aligned}
L(x_{0:T}, u_{0:T-1}, p_{0:T}) &= g_T(x_T) + x_0^T p_0 + \sum_{t=0}^{T-1} g(x_t, u_t) + (f(x_t, u_t) - x_{t+1})^T p_{t+1} \\
&= g_T(x_T) + x_0^T p_0 - x_T^T p_T + \sum_{t=0}^{T-1} H(x_t, u_t, p_{t+1}) - x_t^T p_t
\end{aligned}
$$

- Setting $\nabla_x L = \nabla_p L = 0$ and explicitly minimizing wrt $u_{0:T-1}$ yields:

Theorem: Discrete-time PMP

If $x^*_{0:T}, u^*_{0:T-1}$ is an optimal state-control trajectory starting at x_0, then there exists a **costate trajectory** $p^*_{0:T}$ such that:

$$
\begin{aligned}
x^*_{t+1} &= \nabla_p H(x^*_t, u^*_t, p^*_{t+1}) = f(x^*_t, u^*_t), & x^*_0 &= x_0 \\
p^*_t &= \nabla_x H(x^*_t, u^*_t, p^*_{t+1}) = \nabla_x g(x^*_t, u^*_t) + \nabla_x f(x^*_t, u^*_t)^T p^*_{t+1}, & p^*_T &= \nabla_x g_T(x^*_T) \\
u^*_t &= \arg\min_u H(x^*_t, u, p^*_{t+1})
\end{aligned}
$$

Gradient of the Cost-to-go via the PMP

- The discrete-time PMP provides an efficient way to evaluate the gradient of the cost-to-go with respect to u and thus optimize control trajectories locally and numerically

Theorem: Cost-to-go Gradient

Given an initial state x_0 and trajectory $u_{0:T-1}$, let $x_{1:T}, p_{0:T}$ be such that:

$$\begin{aligned} x_{t+1} &= f(x_t, u_t), \qquad x_0 \text{ given} \\ p_t &= \nabla_x g(x_t, u_t) + [\nabla_x f(x_t, u_t)]^T p_{t+1}, \quad p_T = \nabla_x g_T(x_T) \end{aligned}$$

Then:

$$\nabla_{u_t} J(x_{0:T}, u_{0:T-1}) = \nabla_u H(x_t, u_t, p_{t+1}) = \nabla_u g(x_t, u_t) + \nabla_u f(x_t, u_t)^T p_{t+1}$$

- Note that x_t can be found in a forward pass (since it does not depend on p) and then p_t can be found in a backward pass

Proof by Induction

- The accumulated cost can be written recursively:

$$J_t(x_{t:T}, u_{t:T-1}) = g(x_t, u_t) + J_{t+1}(x_{t+1:T}, u_{t+1:T-1})$$

- Note that u_t affects the future costs only through $x_{t+1} = f(x_t, u_t)$:

$$\nabla_{u_t} J_t(x_{t:T}, u_{t:T-1}) = \nabla_u g(x_t, u_t) + [\nabla_u f(x_t, u_t)]^T \nabla_{x_{t+1}} J_{t+1}(x_{t+1:T}, u_{t+1:T-1})$$

- **Claim**: $p_t = \nabla_{x_t} J_t(x_{t:T}, u_{t:T-1})$:
  - Base case: $p_T = \nabla_{x_T} g_T(x_T)$
  - Induction: for $t \in [t_0, T)$:

$$\underbrace{\nabla_{x_t} J_t(x_{t:T}, u_{t:T-1})}_{=p_t} = \nabla_x g(x_t, u_t) + [\nabla_x f(x_t, u_t)]^T \underbrace{\nabla_{x_{t+1}} J_{t+1}(x_{t+1:T}, u_{t+1:T-1})}_{=p_{t+1}}$$

which is identical with the costate ODE.