ECE276B: Planning & Learning in Robotics
Lecture 15: Continuous-Time Optimal Control

Nikolay Atanasov

natanasov@ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Outline

Continuous-Time Motion Model

- **time**: $t \in [0, T]$
- **state**: $\mathbf{x}(t) \in \mathcal{X} \subseteq \mathbb{R}^n$, $\forall t \in [0, T]$
- **control**: $\mathbf{u}(t) \in \mathcal{U} \subseteq \mathbb{R}^m$, $\forall t \in [0, T]$
- **motion model**: a stochastic differential equation (SDE):

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)) + C(\mathbf{x}(t), \mathbf{u}(t))\boldsymbol{\omega}(t)$$

  defined by functions $f : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^n$ and $C : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{n \times d}$
- **white noise**: $\boldsymbol{\omega}(t) \in \mathbb{R}^d$, $\forall t \in [0, T]$

Gaussian Process

- A **Gaussian Process** with mean function $\boldsymbol{\mu}(t)$ and covariance function $k(t, t')$ is an $\mathbb{R}^d$-valued continuous-time stochastic process $\{\mathbf{g}(t)\}_t$ such that every finite set $\mathbf{g}(t_1), \ldots, \mathbf{g}(t_n)$ of random variables has a joint Gaussian distribution:

$$\begin{bmatrix} \mathbf{g}(t_1) \\ \vdots \\ \mathbf{g}(t_n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(t_1) \\ \vdots \\ \boldsymbol{\mu}(t_n) \end{bmatrix}, \begin{bmatrix} k(t_1, t_1) & \ldots & k(t_1, t_n) \\ \vdots & \ddots & \vdots \\ k(t_n, t_1) & \cdots & k(t_n, t_n) \end{bmatrix} \right)$$

- Short-hand notation: $\mathbf{g}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), k(t, t'))$
- Intuition: a GP is a Gaussian distribution for a function $\mathbf{g}(t)$

Brownian Motion

- Robert Brown made microscopic observations in 1827 that small particles in plant pollen, when immersed in liquid, exhibit highly irregular motion
- **Brownian Motion** is an $\mathbb{R}^d$-valued continuous-time stochastic process $\{\boldsymbol{\beta}(t)\}_{t \geq 0}$ with the following properties:
  - $\boldsymbol{\beta}(t)$ has stationary independent increments, i.e., for $0 \leq t_0 < t_1 < \ldots < t_n$, $\boldsymbol{\beta}(t_0), \boldsymbol{\beta}(t_1) - \boldsymbol{\beta}(t_0), \ldots, \boldsymbol{\beta}(t_n) - \boldsymbol{\beta}(t_{n-1})$ are independent
  - $\boldsymbol{\beta}(t) - \boldsymbol{\beta}(s) \sim \mathcal{N}(\mathbf{0}, (t-s)Q)$ for $0 \leq s \leq t$ and diffusion matrix Q
  - $\boldsymbol{\beta}(t)$ is almost surely continuous (but nowhere differentiable)
- **Standard Brownian Motion**: $\boldsymbol{\beta}(0) = \mathbf{0}$ and $Q = I$
- Brownian motion is a Gaussian process $\boldsymbol{\beta}(t) \sim \mathcal{GP}(\mathbf{0}, \min\{t, t'\}\, Q)$

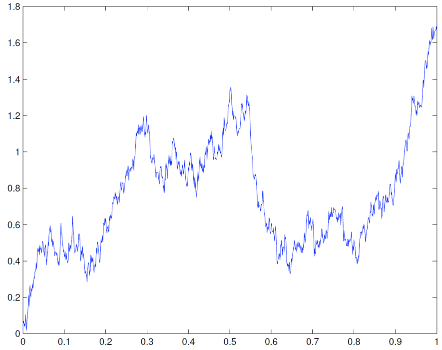
White Noise

- **White Noise** is an $\mathbb{R}^d$-valued continuous-time stochastic process $\{\boldsymbol{\omega}(t)\}_{t\geq 0}$ with the following properties:
  - $\boldsymbol{\omega}(t_1)$ and $\boldsymbol{\omega}(t_2)$ are independent if $t_1 \neq t_2$
  - $\boldsymbol{\omega}(t)$ is a Gaussian process $\mathcal{GP}(\mathbf{0}, \delta(t-t')Q)$ with spectral density Q, where δ is the Dirac delta function.
- The sample paths of $\boldsymbol{\omega}(t)$ are discontinuous almost everywhere
- White noise is unbounded: it takes arbitrarily large positive and negative values at any finite interval
- White noise can be considered the derivative of Brownian motion:

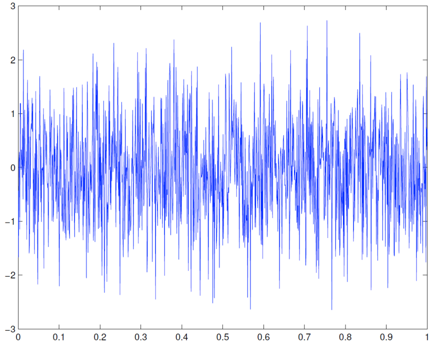
$$d\boldsymbol{\beta}(t) = \boldsymbol{\omega}(t)dt, \qquad \text{where } \boldsymbol{\beta}(t) \sim \mathcal{GP}(\mathbf{0}, \min\{t, t'\}\, Q)$$

- White noise is used to model motion noise in continuous-time systems of ordinary differential equations

Brownian Motion and White Noise



(a) Brownian Motion



(b) White Noise

Continuous-Time Stochastic Optimal Control

- **Problem statement**:

$$\min_{\pi} V^{\pi}(\tau, \mathbf{x}_0) := \mathbb{E}\left\{ \underbrace{\mathfrak{q}(\mathbf{x}(T))}_{\text{terminal cost}} + \int_{\tau}^{T} \underbrace{\ell(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))}_{\text{stage cost}} dt \,\middle|\, \mathbf{x}(\tau) = \mathbf{x}_0 \right\}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \pi(t, \mathbf{x}(t))) + C(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))\boldsymbol{\omega}(t).$$

$$\mathbf{x}(t) \in \mathcal{X}, \; \pi(t, \mathbf{x}(t)) \in PC^0([0, T], \mathcal{U})$$

- **Admissible policies**: set $PC^0([0, T], \mathcal{U})$ of piecewise continuous functions from $[0, T]$ to $\mathcal{U}$

- **Problem variations**:
  - $\mathbf{x}(\tau)$ can be given or free for optimization
  - $\mathbf{x}(T)$ can be in a given target set $\mathcal{T}$ or free for optimization
  - T can be given (**finite-horizon**) or free for optimization (**first-exit**)
  - State and control constraints can be imposed via $\mathcal{X}$ and $\mathcal{U}$

Assumptions

- Motion model $f(\mathbf{x}, \mathbf{u})$ is continuously differentiable wrt to $\mathbf{x}$ and continuous wrt $\mathbf{u}$
- **Existence and uniqueness**: for any admissible policy π and initial state $\mathbf{x}(\tau) \in \mathcal{X}$, $\tau \in [0, T]$, the **noise-free** system, $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))$, has a **unique state trajectory** $\mathbf{x}(t)$, $t \in [\tau, T]$.
- Stage cost $\ell(\mathbf{x}, \mathbf{u})$ is continuously differentiable wrt $\mathbf{x}$ and continuous wrt $\mathbf{u}$
- Terminal cost $\mathfrak{q}(\mathbf{x})$ is continuously differentiable wrt $\mathbf{x}$

Example: Existence and Uniqueness

- **Example**: Existence in not guaranteed

$$\dot{x}(t) = x(t)^2,\ x(0) = 1$$

$$\text{A solution does not exist for}\, T \geq 1 : x(t) = \frac{1}{1-t}$$

- **Example**: Uniqueness in not guaranteed

$$\dot{x}(t) = x(t)^{\frac{1}{3}},\ x(0) = 0$$

$$\text{Infinite number of solutions :}\ \begin{aligned} &x(t) = 0,\ \forall t \\ &x(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq \tau \\ \left(\frac{2}{3}(t-\tau)\right)^{3/2} & \text{for } t > \tau \end{cases} \end{aligned}$$

Special Case: Calculus of Variations

- Let $C^1([a,b],\mathbb{R}^m)$ be the set of continuously differentiable functions from $[a,b]$ to $\mathbb{R}^m$
- **Calculus of Variations**: find a curve $\mathbf{y}(x)$ for $x \in [a,b]$ from $\mathbf{y}_0$ to $\mathbf{y}_f$ that minimizes a cumulative cost function:

$$\begin{aligned} \min_{\mathbf{y} \in C^1([a,b],\mathbb{R}^m)} \quad & \mathfrak{q}(\mathbf{y}(b)) + \int_a^b \ell(\mathbf{y}(x), \dot{\mathbf{y}}(x))dx \\ \text{s.t.} \quad & \mathbf{y}(a) = \mathbf{y}_0,\ \mathbf{y}(b) = \mathbf{y}_f \end{aligned}$$

- The cost may be curve length or travel time for a particle accelerated by gravity (Brachistochrone Problem)
- Special case of continuous-time deterministic optimal control:
  - **fully-actuated** system: $\dot{\mathbf{x}} = \mathbf{u}$
  - **notation**: $t \leftarrow x$, $\mathbf{x}(t) \leftarrow \mathbf{y}(x)$, $\mathbf{u}(t) = \dot{\mathbf{x}}(t) \leftarrow \dot{\mathbf{y}}(x)$

Sufficient Condition for Optimality

- **Optimal value function**:

$$V^*(t, \mathbf{x}) \leq V^\pi(t, \mathbf{x}), \quad \forall \pi \in PC^0([0, T], \mathcal{U}),\ \mathbf{x} \in \mathcal{X}$$

Sufficient Optimality Condition: HJB PDE

Suppose that $V(t, \mathbf{x})$ is continuously differentiable in t and $\mathbf{x}$ and solves the Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE):

$$V(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}$$

$$-\frac{\partial V(t, \mathbf{x})}{\partial t} = \min_{\mathbf{u} \in \mathcal{U}} \left[\ell(\mathbf{x}, \mathbf{u}) + \nabla_{\mathbf{x}} V(t, \mathbf{x})^\top f(\mathbf{x}, \mathbf{u}) + \frac{1}{2} \operatorname{tr}\left(\Sigma(\mathbf{x}, \mathbf{u}) \left[\nabla_{\mathbf{x}}^2 V(t, \mathbf{x}) \right] \right) \right]$$

for all $t \in [0, T]$ and $\mathbf{x} \in \mathcal{X}$ and where $\Sigma(\mathbf{x}, \mathbf{u}) := C(\mathbf{x}, \mathbf{u}) C^\top(\mathbf{x}, \mathbf{u})$.

Then, under the assumptions on Slide 9, $V(t, \mathbf{x})$ is the unique solution of the HJB PDE and is equal to the optimal value function $V^*(t, \mathbf{x})$ of the continuous-time stochastic optimal control problem.

The policy $\pi^*(t, \mathbf{x})$ that attains the minimum in the HJB PDE for all t and $\mathbf{x}$ is an optimal policy.

Existence and Uniqueness of HJB PDE Solutions

- The HJB PDE is the continuous-time analog of the Bellman Equation
- The HJB PDE has at most one classical solution – a function which satisfies the PDE everywhere
- When the optimal value function is not smooth, the HJB PDE does not have a classical solution. It has a unique viscosity solution which is the optimal value function.
- Approximation of the HJB PDE based on MDP discretization is guaranteed to converge to the unique viscosity solution
- Most continuous function approximation schemes (which scale better) are unable to represent non-smooth value functions
- All examples of non-smooth value functions seem to be deterministic, i.e., noise smooths the optimal value function

HJB PDE Derivation

- A discrete-time approximation of the continuous-time optimal control problem can be used to derive the HJB PDE from the DP algorithm
- Motion model: $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}) + C(\mathbf{x}, \mathbf{u})\boldsymbol{\omega}$ with $\mathbf{x}(0) = \mathbf{x}_0$
- **Euler Discretization** of the SDE with time step τ:
  - Discretize $[0, T]$ into N pieces of width $\tau := \frac{T}{N}$
  - Define $\mathbf{x}_k := \mathbf{x}(k\tau)$ and $\mathbf{u}_k := \mathbf{u}(k\tau)$ for $k = 0, \ldots, N$
  - Discretized motion model:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \tau f(\mathbf{x}_k, \mathbf{u}_k) + C(\mathbf{x}_k, \mathbf{u}_k)\boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \tau I) \\ &= \mathbf{x}_k + \mathbf{d}_k, \quad \mathbf{d}_k \sim \mathcal{N}(\tau f(\mathbf{x}_k, \mathbf{u}_k), \tau\Sigma(\mathbf{x}_k, \mathbf{u}_k)) \end{aligned}$$

    where $\Sigma(\mathbf{x}, \mathbf{u}) = C(\mathbf{x}, \mathbf{u})C^\top(\mathbf{x}, \mathbf{u})$ as before
  - Gaussian motion model: $p_f(\mathbf{x}' \mid \mathbf{x}, \mathbf{u}) = \phi(\mathbf{x}'; \mathbf{x} + \tau f(\mathbf{x}, \mathbf{u}), \tau\Sigma(\mathbf{x}, \mathbf{u}))$, where ϕ is the Gaussian probability density function
  - Discretized stage cost: $\tau\ell(\mathbf{x}, \mathbf{u})$

HJB PDE Derivation

- Consider the Bellman Equation of the discrete-time problem and take the limit as $\tau \to 0$ to obtain a "continuous-time Bellman Equation"
- **Bellman Equation**: finite-horizon problem with $t := k\tau$

$$V(t, \mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} \left\{ \tau \ell(\mathbf{x}, \mathbf{u}) + \mathbb{E}_{\mathbf{x}' \sim p_f(\cdot | \mathbf{x}, \mathbf{u})} \left[V(t + \tau, \mathbf{x}') \right] \right\}$$

- Note that $\mathbf{x}' = \mathbf{x} + \mathbf{d}$ where $\mathbf{d} \sim \mathcal{N}(\tau f(\mathbf{x}, \mathbf{u}), \tau \Sigma(\mathbf{x}, \mathbf{u}))$
- Taylor-series expansion of $V(t + \tau, \mathbf{x}')$ around $(t, \mathbf{x})$:

$$\begin{aligned} V(t + \tau, \mathbf{x} + \mathbf{d}) = & V(t, \mathbf{x}) + \tau \frac{\partial V}{\partial t}(t, \mathbf{x}) + o(\tau^2) \\ & + \left[\nabla_{\mathbf{x}} V(t, \mathbf{x}) \right]^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top \left[\nabla_{\mathbf{x}}^2 V(t, \mathbf{x}) \right] \mathbf{d} + o(\mathbf{d}^3) \end{aligned}$$

HJB PDE Derivation

- Note that $\mathbb{E}\left[\mathbf{d}^\top M \mathbf{d}\right] = \boldsymbol{\mu}^\top M \boldsymbol{\mu} + \mathrm{tr}(\Sigma M)$ for $\mathbf{d} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ so that:

$$\begin{aligned}\mathbb{E}_{\mathbf{x}' \sim p_f(\cdot|\mathbf{x},\mathbf{u})}\left[V(t+\tau, \mathbf{x}')\right] = V(t,\mathbf{x}) + \tau \frac{\partial V}{\partial t}(t,\mathbf{x}) + o(\tau^2) \\ + \tau \left[\nabla_{\mathbf{x}} V(t,\mathbf{x})\right]^\top f(\mathbf{x},\mathbf{u}) + \frac{\tau}{2}\mathrm{tr}\left(\Sigma(\mathbf{x},\mathbf{u})\left[\nabla_{\mathbf{x}}^2 V(t,\mathbf{x})\right]\right)\end{aligned}$$

- Substituting in the Bellman Equation and simplifying, we get:

$$0 = \min_{\mathbf{u} \in \mathcal{U}} \left\{ \ell(\mathbf{x},\mathbf{u}) + \frac{\partial V}{\partial t}(t,\mathbf{x}) + \left[\nabla_{\mathbf{x}} V(t,\mathbf{x})\right]^\top f(\mathbf{x},\mathbf{u}) + \frac{1}{2}\mathrm{tr}\left(\Sigma(\mathbf{x},\mathbf{u})\left[\nabla_{\mathbf{x}}^2 V(t,\mathbf{x})\right]\right) + \frac{o(\tau^2)}{\tau} \right\}$$

- Taking the limit as $\tau \to 0$ (assuming it can be exchanged with $\min_{\mathbf{u} \in \mathcal{U}}$) leads to the HJB PDE:

$$-\frac{\partial V}{\partial t}(t,\mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} \left\{ \ell(\mathbf{x},\mathbf{u}) + \left[\nabla_{\mathbf{x}} V(t,\mathbf{x})\right]^\top f(\mathbf{x},\mathbf{u}) + \frac{1}{2}\mathrm{tr}\left(\Sigma(\mathbf{x},\mathbf{u})\left[\nabla_{\mathbf{x}}^2 V(t,\mathbf{x})\right]\right) \right\}$$

Example 1: Guessing a Solution for the HJB PDE

- System: $\dot{x}(t) = u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Cost: $\ell(x, u) = 0$ and $\mathfrak{q}(x) = \frac{1}{2}x^2$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$
- Since we only care about the square of the terminal state, we can construct a candidate optimal policy that drives the state towards 0 as quickly as possible and maintains it there:

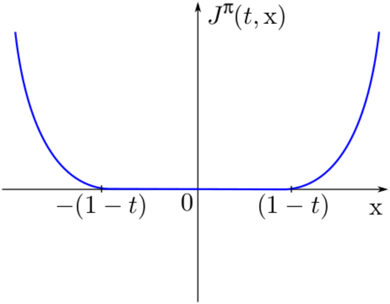
$$\pi(t, x) = -sgn(x) := \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$$

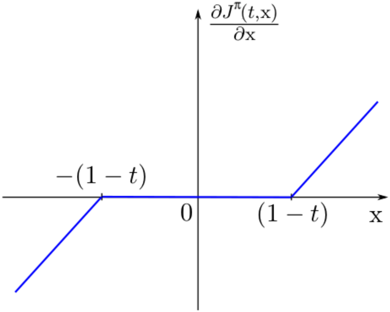
- The value in not smooth: $V^{\pi}(t, x) = \frac{1}{2}\left(\max\left\{0, |x| - (1 - t)\right\}\right)^2$
- We will verify that this function satisfies the HJB and is therefore indeed the optimal value function

Example 1: Partial Derivative wrt x

- Value function and its partial derivative wrt x for fixed t:

$$V^{\pi}(t,x) = \frac{1}{2}\left(\max\{0, |x| - (1-t)\}\right)^2 \qquad \frac{\partial V^{\pi}(t,x)}{\partial x} = sgn(x)\max\{0, |x| - (1-t)\}$$

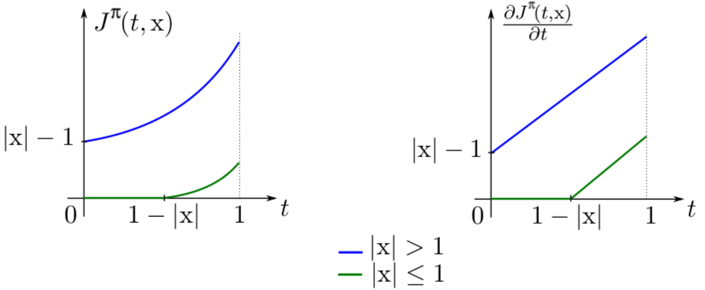




Example 1: Partial Derivative wrt t

- Value function and its partial derivative wrt t for fixed x:

$$V^{\pi}(t,x) = \frac{1}{2}\left(\max\{0, |x| - (1-t)\}\right)^2 \qquad \frac{\partial V^{\pi}(t,x)}{\partial t} = \max\{0, |x| - (1-t)\}$$



Example 1: Guessing a Solution for the HJB PDE

- Boundary condition: $V^{\pi}(1, x) = \frac{1}{2}x^2 = \mathfrak{q}(x)$
- The minimum in the HJB PDE is obtained by $u = -sgn(x)$:

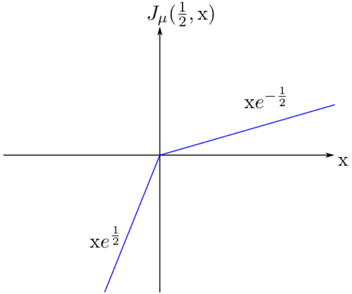
$$\min_{|u|\leq 1}\left(\frac{\partial V^{\pi}(t,x)}{\partial t} + \frac{\partial V^{\pi}(t,x)}{\partial x}u\right) = \min_{|u|\leq 1}\left((1 + sgn(x)u)\left(\max\{0, |x| - (1-t)\}\right)\right) = 0$$

- Conclusion: $V^{\pi}(t, x) = V^{*}(t, x)$ and $\pi^{*}(t, x) = -sgn(x)$ is an optimal policy

Example 2: HJB PDE without a Classical Solution

- System: $\dot{x}(t) = x(t)u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Cost: $\ell(x, u) = 0$ and $\mathfrak{q}(x) = x$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$
- Optimal policy: $\pi(t, x) = \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$
- Optimal value function:

$$V^\pi(t, x) = \begin{cases} e^{t-1}x & x > 0 \\ 0 & x = 0 \\ e^{1-t}x & x < 0 \end{cases}$$



- The value function is not differentiable wrt x at $x = 0$ and hence does not satisfy the HJB PDE in the classical sense

Inf-Horizon Continuous-Time Stochastic Optimal Control

- $V^{\pi}(\mathbf{x}) := \mathbb{E}\left[\int_0^{\infty} \underbrace{e^{-\frac{t}{\gamma}}}_{\text{discount}} \ell(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))dt\right]$ with $\gamma \in [0, \infty)$

HJB PDEs for the Optimal Value Function

Hamiltonian: $$H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \ell(\mathbf{x}, \mathbf{u}) + \mathbf{p}^{\top} f(\mathbf{x}, \mathbf{u}) + \frac{1}{2} \operatorname{tr}\left(C(\mathbf{x}, \mathbf{u}) C^{\top}(\mathbf{x}, \mathbf{u})[\nabla_{\mathbf{x}} \mathbf{p}]\right)$$

Finite Horizon: $$-\frac{\partial V^*}{\partial t}(t, \mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}, \mathbf{u}, \nabla_{\mathbf{x}} V^*(t, \mathbf{x})), \quad V^*(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x})$$

First Exit: $$0 = \min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}, \mathbf{u}, \nabla_{\mathbf{x}} V^*(\mathbf{x})), \quad V^*(\mathbf{x}) = \mathfrak{q}(\mathbf{x}), \;\; \forall \mathbf{x} \in \mathcal{T}$$

Discounted: $$\frac{1}{\gamma} V^*(\mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}, \mathbf{u}, \nabla_{\mathbf{x}} V^*(\mathbf{x}))$$

Tractable Problems

- **Control-affine motion model**: $\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u} + C(\mathbf{x})\omega$
- **Stage cost quadratic in u**: $\ell(\mathbf{x}, \mathbf{u}) = q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top R(\mathbf{x})\mathbf{u}$, $R(\mathbf{x}) \succ 0$
- The Hamiltonian can be minimized analytically wrt $\mathbf{u}$ (suppressing the dependence on $\mathbf{x}$ for clarity):

$$H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = q + \frac{1}{2}\mathbf{u}^\top R\mathbf{u} + \mathbf{p}^\top(\mathbf{a} + B\mathbf{u}) + \frac{1}{2}\operatorname{tr}(CC^\top \mathbf{p}_{\mathbf{x}})$$

$$\nabla_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = R\mathbf{u} + B^\top \mathbf{p} \qquad \nabla^2_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = R \succ 0$$

- Optimal policy for $t \in [0, T]$ and $\mathbf{x} \in \mathcal{X}$:

$$\pi^*(t, \mathbf{x}) = \arg\min_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, V_{\mathbf{x}}(t, \mathbf{x})) = -R^{-1}(\mathbf{x})B^\top(\mathbf{x})V_{\mathbf{x}}(t, \mathbf{x})$$

- The HJB PDE becomes a second-order quadratic PDE, no longer involving the min operator:

$$V(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x}),$$

$$-V_t(t, \mathbf{x}) = q + \mathbf{a}^\top V_{\mathbf{x}}(t, \mathbf{x}) + \frac{1}{2}\operatorname{tr}(CC^\top V_{\mathbf{xx}}(t, \mathbf{x})) - \frac{1}{2}V_{\mathbf{x}}(t, \mathbf{x})^\top BR^{-1}B^\top V_{\mathbf{x}}(t, \mathbf{x})$$

Example: Pendulum

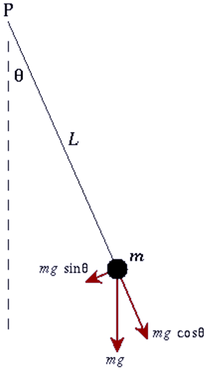
- Pendulum dynamics (Newton's second law for rotational systems):

$$mL^2\ddot{\theta} = u - mgL\sin\theta + noise$$

- Noise: $\sigma\omega(t)$ with $\omega(t) \sim \mathcal{GP}(0, \delta(t-t'))$

- State-space form with $\mathbf{x} = (x_1, x_2) = (\theta, \dot{\theta})$:

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ k\sin(x_1) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(u + \sigma\omega)$$



- Stage cost: $\ell(\mathbf{x}, u) = q(\mathbf{x}) + \frac{r}{2}u^2$

- Optimal value and policy for a discounted problem formulation:

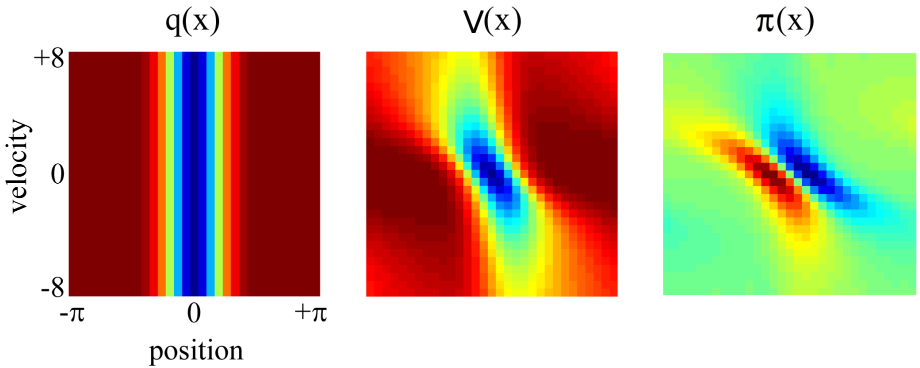
$$\pi^*(\mathbf{x}) = -\frac{1}{r}V^*_{x_2}(\mathbf{x})$$

$$\frac{1}{\gamma}V^*(\mathbf{x}) = q(\mathbf{x}) + x_2 V^*_{x_1}(\mathbf{x}) + k\sin(x_1)V^*_{x_2}(\mathbf{x}) + \frac{\sigma^2}{2}V^*_{x_2x_2}(\mathbf{x}) - \frac{1}{2r}(V^*_{x_2}(\mathbf{x}))^2$$

Example: Pendulum

- Parameters: $k = \sigma = r = 1$, $\gamma = 0.3$, $q(\theta, \dot{\theta}) = 1 - \exp(-2\theta^2)$
- Discretize the state space, approximate derivatives via finite differences, and iterate:

$$V^{(i+1)}(\mathbf{x}) = V^{(i)}(\mathbf{x}) + \alpha \left(\gamma \min_u H(\mathbf{x}, u, \nabla_{\mathbf{x}} V^{(i)}(\cdot)) - V^{(i)}(\mathbf{x}) \right), \quad \alpha = 0.01$$



Outline

Continuous-Time Optimal Control

Continuous-Time PMP

Continuous-Time LQR

Continuous-Time Deterministic Optimal Control

- **Problem statement**:

$$\begin{aligned} \min_{\pi} \quad & V^{\pi}(0, \mathbf{x}_0) := \mathfrak{q}(\mathbf{x}(T)) + \int_0^T \ell(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))dt \\ \text{s.t.} \quad & \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \;\; \mathbf{x}(0) = \mathbf{x}_0, \\ & \mathbf{x}(t) \in \mathcal{X}, \\ & \pi(t, \mathbf{x}(t)) \in PC^0([0, T], \mathcal{U}) \end{aligned}$$

- **Admissible policies**: $PC^0([0, T], \mathcal{U})$ is the set of piecewise continuous functions from $[0, T]$ to $\mathcal{U}$
- **Optimal value function**: $V^*(t, \mathbf{x}) = \min_\pi V^\pi(t, \mathbf{x})$

Relationship to Mechanics

- **Costate** $\mathbf{p}(t)$ is the gradient (sensitivity) of the optimal value function $V^*(t, \mathbf{x}(t))$ with respect to the state $\mathbf{x}(t)$.
- **Hamiltonian**: captures the total energy of the system:

$$H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \ell(\mathbf{x}, \mathbf{u}) + \mathbf{p}^\top f(\mathbf{x}, \mathbf{u})$$

- **Hamilton's principle of least action**: trajectories of mechanical systems minimize the action integral $\int_0^T \ell(\mathbf{x}(t), \dot{\mathbf{x}}(t))dt$, where the Lagrangian $\ell(\mathbf{x}, \dot{\mathbf{x}}) := K(\dot{\mathbf{x}}) - U(\mathbf{x})$ is the difference between kinetic and potential energy
- If the stage cost is the Lagrangian of a mechanical system, the Hamiltonian is the (negative) total energy (kinetic plus potential)

Lagrangian Mechanics

- Consider a point mass m with position $\mathbf{x}$ and velocity $\dot{\mathbf{x}}$
- Kinetic energy $K(\dot{\mathbf{x}}) := \frac{1}{2}m\|\dot{\mathbf{x}}\|_2^2$ and momentum $\mathbf{p} := m\dot{\mathbf{x}}$
- Potential energy $U(\mathbf{x})$ and conservative force $F = -\frac{\partial U(\mathbf{x})}{\partial \mathbf{x}}$
- Newtonian equations of motion: $F = m\ddot{\mathbf{x}}$
- Note that $-\frac{\partial U(\mathbf{x})}{\partial \mathbf{x}} = F = m\ddot{\mathbf{x}} = \frac{d}{dt}\mathbf{p} = \frac{d}{dt}\left(\frac{\partial K(\dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}\right)$
- Note that $\frac{\partial U(\mathbf{x})}{\partial \dot{\mathbf{x}}} = 0$ and $\frac{\partial K(\dot{\mathbf{x}})}{\partial \mathbf{x}} = 0$
- Lagrangian: $\ell(\mathbf{x}, \dot{\mathbf{x}}) := K(\dot{\mathbf{x}}) - U(\mathbf{x})$
- Euler-Lagrange equation: $\frac{d}{dt}\left(\frac{\partial \ell(\mathbf{x},\dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}\right) - \frac{\partial \ell(\mathbf{x},\dot{\mathbf{x}})}{\partial \mathbf{x}} = 0$

Conservation of Energy

- Total energy $E(\mathbf{x}, \dot{\mathbf{x}}) = K(\dot{\mathbf{x}}) + U(\mathbf{x}) = 2K(\dot{\mathbf{x}}) - \ell(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{p}^\top \dot{\mathbf{x}} - \ell(\mathbf{x}, \dot{\mathbf{x}})$
- Note that:

$$\frac{d}{dt}\left(\mathbf{p}^\top \dot{\mathbf{x}}\right) = \frac{d}{dt}\left(\frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}^\top \dot{\mathbf{x}}\right) = \left(\frac{d}{dt}\frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}\right)^\top \dot{\mathbf{x}} + \frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}^\top \ddot{\mathbf{x}}$$

$$\frac{d}{dt}\ell(\mathbf{x}, \dot{\mathbf{x}}) = \frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \mathbf{x}}^\top \dot{\mathbf{x}} + \frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}^\top \ddot{\mathbf{x}} + \frac{\partial}{\partial t}\ell(\mathbf{x}, \dot{\mathbf{x}})$$

- Conservation of energy using the Euler-Lagrange equation:

$$\frac{d}{dt}E(\mathbf{x}, \dot{\mathbf{x}}) = \frac{d}{dt}\left(\frac{\partial \ell(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}^\top \dot{\mathbf{x}}\right) - \frac{d}{dt}\ell(\mathbf{x}, \dot{\mathbf{x}}) = -\frac{\partial}{\partial t}\ell(\mathbf{x}, \dot{\mathbf{x}}) = 0$$

- In our formulation, the costate is the negative momentum and the Hamiltonian is the negative total energy

- **Optimal open-loop trajectories** (local minima) can be computed by solving a boundary-value ODE with initial **state** $\mathbf{x}(0) = \mathbf{x}_0$ and terminal **costate** $\mathbf{p}(T) = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}(T))$

Theorem: Pontryagin's Minimum Principle (PMP)

- Let $\mathbf{u}^*(t) : [0, T] \to \mathcal{U}$ be an optimal control trajectory
- Let $\mathbf{x}^*(t) : [0, T] \to \mathcal{X}$ be the associated state trajectory from $\mathbf{x}_0$
- Then, there exists a **costate trajectory** $\mathbf{p}^*(t) : [0, T] \to \mathcal{X}$ satisfying:
  1. **Canonical equations with boundary conditions**:

$$\begin{aligned} \dot{\mathbf{x}}^*(t) &= \nabla_{\mathbf{p}} H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t)), \quad && \mathbf{x}^*(0) = \mathbf{x}_0 \\ \dot{\mathbf{p}}^*(t) &= -\nabla_{\mathbf{x}} H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t)), \quad && \mathbf{p}^*(T) = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}^*(T)) \end{aligned}$$

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

$$\begin{aligned} \mathbf{u}^*(t) &\in \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t)), && \forall t \in [0, T] \\ H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t)) &= \mathit{constant}, && \forall t \in [0, T] \end{aligned}$$

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

HJB PDE vs PMP

- The HJB PDE provides a lot of information – the optimal value function and an optimal policy for all time and all states!
- Often, we only care about the optimal trajectory for a specific initial condition $\mathbf{x}_0$. Exploiting that we need less information, we can arrive at simpler conditions for optimality – the PMP
- 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 a candidate PMP policy is optimal
- The PMP requires solving an ODE with split boundary conditions (not easy but easier than the nonlinear HJB PDE!)
- The PMP does **not apply to infinite horizon problems**, so one has to use the HJB PDE in that case

Proof of PMP (Step 0: Preliminaries)

Lemma: ∇-min Exchange

Let $F(t, \mathbf{x}, \mathbf{u})$ be continuously differentiable in $t \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$ and let $\mathcal{U} \subseteq \mathbb{R}^m$ be a convex set. Assume $\pi^*(t, \mathbf{x}) = \arg\min_{\mathbf{u} \in \mathcal{U}} F(t, \mathbf{x}, \mathbf{u})$ exists and is continuously differentiable. Then, for all t and $\mathbf{x}$:

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

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

$$\frac{\partial}{\partial t} G(t, \mathbf{x}) = \frac{\partial}{\partial t} F(t, \mathbf{x}, \mathbf{u})\Big|_{\mathbf{u}=\pi^*(t,\mathbf{x})} + \underbrace{\frac{\partial}{\partial \mathbf{u}} F(t, \mathbf{x}, \mathbf{u})\Big|_{\mathbf{u}=\pi^*(t,\mathbf{x})}}_{=0 \text{ by 1st order optimality condition}} \frac{\partial \pi^*(t, \mathbf{x})}{\partial t}$$

A similar derivation can be used for the partial derivative wrt $\mathbf{x}$.

Proof of PMP (Step 1: HJB PDE gives $V^*(t, \mathbf{x})$)

- **Extra Assumptions**: $V^*(t, \mathbf{x})$ and $\pi^*(t, \mathbf{x})$ are continuously differentiable in t and $\mathbf{x}$ and $\mathcal{U}$ is convex. These assumptions can be avoided in a more general proof.

- With a continuously differentiable value function, the HJB PDE is also a necessary condition for optimality:

$$V^*(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}$$
$$0 = \min_{\mathbf{u} \in \mathcal{U}} \underbrace{\left(\ell(\mathbf{x}, \mathbf{u}) + \frac{\partial}{\partial t} V^*(t, \mathbf{x}) + \nabla_{\mathbf{x}} V^*(t, \mathbf{x})^\top f(\mathbf{x}, \mathbf{u}) \right)}_{:= F(t, \mathbf{x}, \mathbf{u})}, \quad \forall t \in [0, T], \mathbf{x} \in \mathcal{X}$$

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

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

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

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

$$\begin{aligned} 0 &= \nabla_{\mathbf{x}}\left(\min_{\mathbf{u} \in \mathcal{U}} F(t, \mathbf{x}, \mathbf{u})\right) \\ &= \nabla_{\mathbf{x}} \ell(\mathbf{x}, \mathbf{u}^*) + \nabla_{\mathbf{x}} \frac{\partial}{\partial t} V^*(t, \mathbf{x}) + [\nabla_{\mathbf{x}}^2 V^*(t, \mathbf{x})] f(\mathbf{x}, \mathbf{u}^*) + [\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{u}^*)]^\top \nabla_{\mathbf{x}} V^*(t, \mathbf{x}) \end{aligned}$$

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

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

$$\dot{\mathbf{x}}^*(t) = f(\mathbf{x}^*(t), \mathbf{u}^*(t)) = \nabla_{\mathbf{p}} H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}), \qquad \mathbf{x}^*(0) = \mathbf{x}_0$$

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

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

$$0 = \left.\frac{\partial^2 V^*(t,\mathbf{x})}{\partial t^2}\right|_{\mathbf{x}=\mathbf{x}^*(t)} + \left[\frac{\partial}{\partial t}\nabla_{\mathbf{x}} V^*(t,\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*(t)}\right]^\top \dot{\mathbf{x}}^*(t)$$

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

$$0 = \nabla_{\mathbf{x}}\ell(\mathbf{x},\mathbf{u}^*)|_{\mathbf{x}=\mathbf{x}^*(t)} + \frac{d}{dt}\left(\underbrace{\nabla_{\mathbf{x}} V^*(t,\mathbf{x})|_{\mathbf{x}=\mathbf{x}^*(t)}}_{=:\mathbf{p}^*(t)}\right)$$

$$+ [\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{u}^*)|_{\mathbf{x}=\mathbf{x}^*(t)}]^\top [\nabla_{\mathbf{x}} V^*(t,\mathbf{x})|_{\mathbf{x}=\mathbf{x}^*(t)}]$$

$$= \nabla_{\mathbf{x}}\ell(\mathbf{x},\mathbf{u}^*)|_{\mathbf{x}=\mathbf{x}^*(t)} + \dot{\mathbf{p}}^*(t) + [\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{u}^*)|_{\mathbf{x}=\mathbf{x}^*(t)}]^\top \mathbf{p}^*(t)$$

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

Proof of PMP (Step 4: Done)

- The boundary condition $V^*(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x})$ implies that $\nabla_{\mathbf{x}} V^*(T, \mathbf{x}) = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ and thus $\mathbf{p}^*(T) = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}^*(T))$

- From the HJB PDE we have:

$$-\frac{\partial}{\partial t} V^*(t, \mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}, \mathbf{u}, \nabla_{\mathbf{x}} V^*(t, \cdot))$$

which along the optimal trajectory $\mathbf{x}^*(t)$, $\mathbf{u}^*(t)$ becomes:

$$-r(t) = H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t)) = const$$

- Finally, note that

$$\begin{aligned}
\mathbf{u}^*(t) &= \arg\min_{\mathbf{u} \in \mathcal{U}} F(t, \mathbf{x}^*(t), \mathbf{u}) \\
&= \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \ell(\mathbf{x}^*(t), \mathbf{u}) + [\nabla_{\mathbf{x}} V^*(t, \mathbf{x})|_{\mathbf{x}=\mathbf{x}^*(t)}]^\top f(\mathbf{x}^*(t), \mathbf{u}) \right\} \\
&= \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \ell(\mathbf{x}^*(t), \mathbf{u}) + \mathbf{p}^*(t)^\top f(\mathbf{x}^*(t), \mathbf{u}) \right\} \\
&= \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t))
\end{aligned}$$

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 $f(x, u) = ux$, $\ell(x, u) = -(1 - u)x$ and $\mathfrak{q}(x) = 0$

Example: Resource Allocation for a Martian Base

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

$$
\begin{aligned}
\dot{x}^*(t) &= \nabla_p H(x^*, u^*, p^*) = x^*(t)u^*(t), \quad x^*(0) = x, \\
\dot{p}^*(t) &= -\nabla_x H(x^*, u^*, p^*) = (1 - u^*(t)) - p^*(t)u^*(t), \;\; p^*(T) = 0, \\
u^*(t) &= \operatorname*{arg\,min}_{0 \le u \le 1} H(x^*(t), u, p^*(t)) = \operatorname*{arg\,min}_{0 \le u \le 1} (x^*(t)(p^*(t)+1)u)
\end{aligned}
$$

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

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

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 $t \leq T - 1$:

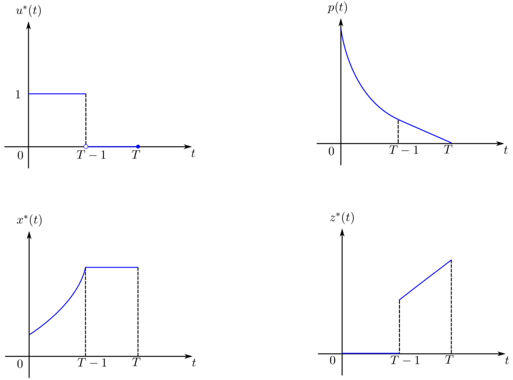
$$\dot{p}^*(t) = -p^*(t), \;\; p(T-1) = -1$$
$$\Rightarrow p^*(t) = e^{-[(T-1)-t]} p(T-1) \leq -1 \;\text{ for } t < T - 1$$

- Optimal control:

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

Example: Resource Allocation for a Martian Base

- Optimal trajectories for the Martian resource allocation problem:



- **Conclusions**:
  - All robots replicate themselves from $t = 0$ to $t = T - 1$ and then all robots 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 $\mathbf{x}(0) = \mathbf{x}_0$, a final state $\mathbf{x}(T) = \mathbf{x}_\tau$ is given.
- The terminal cost $\mathfrak{q}(\mathbf{x}(T))$ is not useful since $V^*(T, \mathbf{x}) = \infty$ if $\mathbf{x}(T) \neq \mathbf{x}_\tau$. The terminal boundary condition for the costate $\mathbf{p}(T) = \nabla_\mathbf{x} \mathfrak{q}(\mathbf{x}(T))$ does not hold but as compensation we have a different boundary condition $\mathbf{x}(T) = \mathbf{x}_\tau$.
- We still have $2n$ ODEs with $2n$ boundary conditions:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= f(\mathbf{x}(t), \mathbf{u}(t)), \qquad \mathbf{x}(0) = \mathbf{x}_0,\ \mathbf{x}(T) = \mathbf{x}_\tau \\ \dot{\mathbf{p}}(t) &= -\nabla_\mathbf{x} H(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)) \end{aligned}$$

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

$$\begin{aligned} \dot{\mathbf{x}}(t) &= f(\mathbf{x}(t), \mathbf{u}(t)), \qquad \mathbf{x}(0) = \mathbf{x}_0,\ \ \mathbf{x}_j(T) = \mathbf{x}_{\tau,j},\ \ \forall j \in I \\ \dot{\mathbf{p}}(t) &= -\nabla_\mathbf{x} H(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)), \qquad \mathbf{p}_j(T) = \frac{\partial}{\partial x_j} \mathfrak{q}(\mathbf{x}(T)),\ \ \forall j \notin I \end{aligned}$$

PMP with Fixed Terminal Set

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

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

- The costate boundary condition requires that $\mathbf{p}(T)$ is orthogonal to the tangent space $D = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla_\mathbf{x} h_j(\mathbf{x}(T))^\top \mathbf{d} = 0,\ j = 1, \ldots, n-k\}$:

$$\begin{aligned}
&\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \qquad \mathbf{x}(0) = \mathbf{x}_0, && h_j(\mathbf{x}(T)) = 0,\ j = 1, \ldots, n-k \\
&\dot{\mathbf{p}}(t) = -\nabla_\mathbf{x} H(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)), && \mathbf{p}(T) \in \mathbf{span}\{\nabla_\mathbf{x} h_j(\mathbf{x}(T)), \forall j\} \\
& && \text{or} \quad \mathbf{d}^\top \mathbf{p}(T) = 0,\ \forall \mathbf{d} \in D
\end{aligned}$$

PMP with Free Initial State

- Suppose that $\mathbf{x}_0$ is free and subject to optimization with additional cost term $\ell_0(\mathbf{x}_0)$
- The total cost becomes $\ell_0(\mathbf{x}_0) + V(0, \mathbf{x}_0)$ and the necessary condition for an optimal initial state $\mathbf{x}_0$ is:

$$\nabla_{\mathbf{x}} \ell_0(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0} + \underbrace{\nabla_{\mathbf{x}} V(0, \mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}}_{=\mathbf{p}(0)} = 0 \quad \Rightarrow \quad \mathbf{p}(0) = -\nabla_{\mathbf{x}} \ell_0(\mathbf{x}_0)$$

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

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t))$$
$$\dot{\mathbf{p}}(t) = -\nabla_{\mathbf{x}} H(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t)), \;\; \mathbf{p}(0) = -\nabla_{\mathbf{x}} \ell_0(\mathbf{x}_0), \;\; \mathbf{p}(T) = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}(T))$$

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

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

- Recall that on the optimal trajectory:

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

- 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:

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

PMP with Time-Varying System and Cost

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

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

- Convert the problem to a time-invariant one by making t part of the state, i.e., let $y(t) = t$ with dynamics:

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

- Augmented state $\mathbf{z}(t) := (\mathbf{x}(t), y(t))$ and system:

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

$$\bar{\ell}(\mathbf{z}, \mathbf{u}) := \ell(\mathbf{x}, \mathbf{u}, y) \quad \bar{\mathfrak{q}}(\mathbf{z}) := \mathfrak{q}(\mathbf{x})$$

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

$$\begin{aligned} H(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) &= \ell(\mathbf{x}, \mathbf{u}, t) + \mathbf{p}^\top f(\mathbf{x}, \mathbf{u}, t) \\ \dot{\mathbf{x}}^*(t) &= f(\mathbf{x}^*(t), \mathbf{u}^*(t), t), & \mathbf{x}^*(0) &= \mathbf{x}_0 \\ \dot{\mathbf{p}}^*(t) &= -\nabla_{\mathbf{x}} H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t), t), & \mathbf{p}^*(T) &= \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}^*(T)) \\ \mathbf{u}^*(t) &\in \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t), t) \end{aligned}$$

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

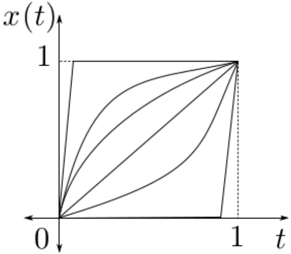
Singular Problems

- The minimum condition $\mathbf{u}(t) \in \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t), t)$ may be insufficient to determine $\mathbf{u}^*(t)$ for all t when $\mathbf{x}^*(t)$ and $\mathbf{p}^*(t)$ are such that $H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t), t)$ is independent of $\mathbf{u}$ over a nontrivial interval of time

- Optimal trajectories consist of portions where $\mathbf{u}^*(t)$ can be determined from the minimum condition (**regular arcs**) and where $\mathbf{u}^*(t)$ cannot be determined from the minimum condition since the Hamiltonian is independent of $\mathbf{u}$ (**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 $x(1) = 1$



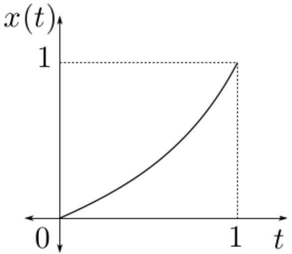
- Approach: use PMP to find a locally optimal open-loop policy

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:

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

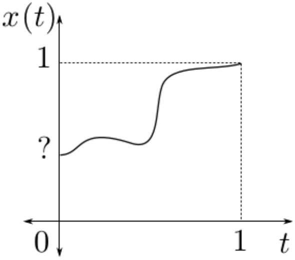
- Candidate trajectory: $\ddot{x}(t) = x(t) \quad \Rightarrow \quad x(t) = ae^t + be^{-t} = \frac{e^t - e^{-t}}{e - e^{-1}}$
  - $x(0) = 0 \quad \Rightarrow \quad a + b = 0$
  - $x(1) = 1 \quad \Rightarrow \quad ae + be^{-1} = 1$



- 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

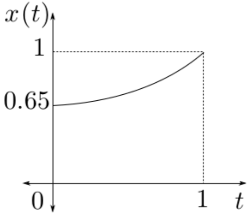
Example: Free Initial State

- 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:

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



  - $x(1) = 1 \quad \Rightarrow \quad ae + be^{-1} = 1$
  - $p(0) = 0 \quad \Rightarrow \quad -a + b = 0$
  - $x(0) \approx 0.65$
  - 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: $T = \textit{free}$
- Note: if we do not include 1 in the stage-cost (e.g., use the same cost as in the previous example), 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: $H(x(t), u(t), p(t)) = 1 + \frac{1}{2}(x(t)^2 + u(t)^2) + p(t)u(t)$
  - 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:

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

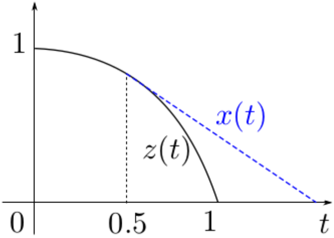
- Candidate trajectory: $\ddot{x}(t) = x(t) \quad \Rightarrow \quad x(t) = ae^t + be^{-t} = \frac{e^t - e^{-t}}{e^T - e^{-T}}$
  - $x(0) = 0 \quad \Rightarrow \quad a + b = 0$
  - $x(T) = 1 \quad \Rightarrow \quad ae^T + be^{-T} = 1$

- Free terminal time:

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

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:

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

  - 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 example, the singular arc can be determined from the costate ODE. For $p(t) = 0$:

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

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 $0 \leq t \leq t_s \leq \frac{1}{2}$: $\quad x(t) = z(t) \quad p(t) = 0$
- For $t_s < t \leq 1$:

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

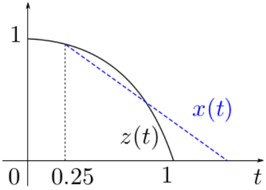
$$\dot{p}(t) = -(x(t) - z(t)) = t_s^2 - t_s - t^2 + t, \qquad p(t_s) = p(1) = 0$$

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

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

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

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



Outline

Continuous-Time Optimal Control

Continuous-Time PMP

Continuous-Time LQR

Globally Optimal Closed-Loop Control

- **Finite-horizon continuous-time deterministic optimal control**:

$$\begin{aligned}
\min_{\pi} \quad & V^{\pi}(0,\mathbf{x}_0) := \mathfrak{q}(\mathbf{x}(T)) + \int_0^T \ell(\mathbf{x}(t),\pi(t,\mathbf{x}(t)))dt \\
\text{s.t.} \quad & \dot{\mathbf{x}}(t) = f(\mathbf{x}(t),\mathbf{u}(t)), \;\; \mathbf{x}(0) = \mathbf{x}_0 \\
& \mathbf{x}(t) \in \mathcal{X}, \; \pi(t,\mathbf{x}(t)) \in PC^0([0,T],\mathcal{U})
\end{aligned}$$

- **Hamiltonian**: $H(\mathbf{x},\mathbf{u},\mathbf{p}) := \ell(\mathbf{x},\mathbf{u}) + \mathbf{p}^\top f(\mathbf{x},\mathbf{u})$

HJB PDE: Sufficient Condition for Optimality

If $V(t,\mathbf{x})$ satisfies the HJB PDE:

$$\begin{aligned}
V(T,\mathbf{x}) &= \mathfrak{q}(\mathbf{x}), && \forall \mathbf{x} \in \mathcal{X} \\
-\frac{\partial}{\partial t}V(t,\mathbf{x}) &= \min_{\mathbf{u}\in\mathcal{U}} H(\mathbf{x},\mathbf{u},\nabla_{\mathbf{x}}V(t,\mathbf{x})), && \forall \mathbf{x} \in \mathcal{X}, t \in [0,T]
\end{aligned}$$

then it is the optimal value function and the policy $\pi(t,\mathbf{x})$ that attains the minimum is an optimal policy.

Locally Optimal Open-Loop Control

- **Finite-horizon continuous-time deterministic optimal control**:

$$
\begin{aligned}
\min_{\pi} \quad & V^{\pi}(0, \mathbf{x}_0) := \mathfrak{q}(\mathbf{x}(T)) + \int_0^T \ell(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))dt \\
\text{s.t.} \quad & \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \;\; \mathbf{x}(0) = \mathbf{x}_0 \\
& \mathbf{x}(t) \in \mathcal{X}, \; \pi(t, \mathbf{x}(t)) \in PC^0([0, T], \mathcal{U})
\end{aligned}
$$

- **Hamiltonian**: $H(\mathbf{x}, \mathbf{u}, \mathbf{p}) := \ell(\mathbf{x}, \mathbf{u}) + \mathbf{p}^\top f(\mathbf{x}, \mathbf{u})$

PMP ODE: Necessary Condition for Optimality

If $(\mathbf{x}^*(t), \mathbf{u}^*(t))$ for $t \in [0, T]$ is a trajectory from an optimal policy $\pi^*(t, x)$, then it satisfies:

$$
\begin{aligned}
&\dot{\mathbf{x}}^*(t) = f(\mathbf{x}^*(t), \mathbf{u}^*(t)), && \mathbf{x}^*(0) = \mathbf{x}_0 \\
&\dot{\mathbf{p}}^*(t) = -\nabla_{\mathbf{x}}\ell(\mathbf{x}^*(t), \mathbf{u}^*(t)) - [\nabla_{\mathbf{x}} f(\mathbf{x}^*(t), \mathbf{u}^*(t))]^\top \mathbf{p}^*(t), && \mathbf{p}^*(T) = \nabla_{\mathbf{x}}\mathfrak{q}(\mathbf{x}^*(T)) \\
&\mathbf{u}^*(t) = \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}^*(t), \mathbf{u}, \mathbf{p}^*(t)), && \forall t \in [0, T] \\
&H(\mathbf{x}^*(t), \mathbf{u}^*(t), \mathbf{p}^*(t)) = \mathit{constant}, && \forall t \in [0, T]
\end{aligned}
$$

Tractable Problems

- Control-affine dynamics and quadratic-in-control cost:

$$\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u} \qquad \ell(\mathbf{x}, \mathbf{u}) = q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top R(\mathbf{x})\mathbf{u} \qquad R(\mathbf{x}) \succ 0$$

- **Hamiltonian**:

$$H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top R(\mathbf{x})\mathbf{u} + \mathbf{p}^\top(\mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u})$$
$$\nabla_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = R(\mathbf{x})\mathbf{u} + B(\mathbf{x})^\top \mathbf{p} \qquad \nabla^2_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = R(\mathbf{x})$$

- **HJB PDE**: obtains the globally optimal value function and policy:

$$\pi^*(t, \mathbf{x}) = \arg\min_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, V_{\mathbf{x}}(t, \mathbf{x})) = -R(\mathbf{x})^{-1} B(\mathbf{x})^\top V_{\mathbf{x}}(t, \mathbf{x}),$$
$$V(T, \mathbf{x}) = \mathfrak{q}(\mathbf{x}),$$
$$-V_t(t, \mathbf{x}) = q(\mathbf{x}) + \mathbf{a}(\mathbf{x})^\top V_{\mathbf{x}}(t, \mathbf{x}) - \frac{1}{2} V_{\mathbf{x}}(t, \mathbf{x})^\top B(\mathbf{x}) R(\mathbf{x})^{-1} B(\mathbf{x})^\top V_{\mathbf{x}}(t, \mathbf{x}).$$

Tractable Problems

- Control-affine dynamics and quadratic-in-control cost:

$$\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u} \qquad \ell(\mathbf{x},\mathbf{u}) = q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top R(\mathbf{x})\mathbf{u} \qquad R(\mathbf{x}) \succ 0$$

- **Hamiltonian**:

$$H(\mathbf{x},\mathbf{u},\mathbf{p}) = q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top R(\mathbf{x})\mathbf{u} + \mathbf{p}^\top(\mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u})$$

$$\nabla_{\mathbf{u}} H(\mathbf{x},\mathbf{u},\mathbf{p}) = R(\mathbf{x})\mathbf{u} + B(\mathbf{x})^\top \mathbf{p} \qquad \nabla^2_{\mathbf{u}} H(\mathbf{x},\mathbf{u},\mathbf{p}) = R(\mathbf{x})$$

- **PMP**: both necessary and sufficient for a local minimum:

$$\mathbf{u} = \arg\min_{\mathbf{u}} H(\mathbf{x},\mathbf{u},\mathbf{p}) = -R(\mathbf{x})^{-1} B(\mathbf{x})^\top \mathbf{p},$$

$$\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) - B(\mathbf{x})R^{-1}(\mathbf{x})B^\top(\mathbf{x})\mathbf{p}, \qquad \mathbf{x}(0) = \mathbf{x}_0$$

$$\dot{\mathbf{p}} = -\left(\mathbf{a}_{\mathbf{x}}(\mathbf{x}) + \nabla_{\mathbf{x}} B(\mathbf{x})\mathbf{u}\right)^\top \mathbf{p} - q_{\mathbf{x}}(\mathbf{x}) - \frac{1}{2}\nabla_{\mathbf{x}}[\mathbf{u}^\top R(\mathbf{x})\mathbf{u}], \quad \mathbf{p}(T) = \mathfrak{q}_{\mathbf{x}}(\mathbf{x}(T))$$

Example: Pendulum

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ k \sin(x_1) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \quad \mathbf{x}(0) = \mathbf{x}_0$$

$$a_{\mathbf{x}}(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ k \cos(x_1) & 0 \end{bmatrix}$$

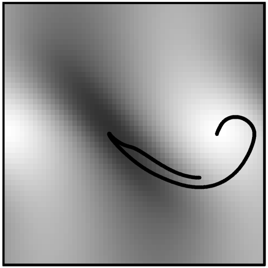
- Cost:

$$\ell(\mathbf{x}, u) = 1 - e^{-2x_1^2} + \frac{r}{2} u^2 \text{ and } \mathfrak{q}(\mathbf{x}) = 0$$

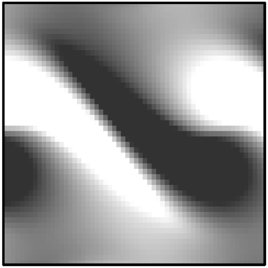
- PMP locally optimal trajectories:

$$\begin{aligned} u(t) &= -r^{-1} p_2(t), & t &\in [0, T] \\ \dot{x}_1 &= x_2, & x_1(0) &= 0 \\ \dot{x}_2 &= k \sin(x_1) - r^{-1} p_2, & x_2(0) &= 0 \\ \dot{p}_1 &= -4e^{-2x_1^2} x_1 - p_2, & p_1(T) &= 0 \\ \dot{p}_2 &= -k \cos(x_1) p_1, & p_2(T) &= 0 \end{aligned}$$

- Optimal value from HJB:



- Optimal policy from HJB:



Linear Quadratic Regulator

- Key assumptions that allowed minimizing the Hamiltonian analytically:
  - The system dynamics are linear in the control $\mathbf{u}$
  - The stage-cost is quadratic in the control $\mathbf{u}$

- **Linear Quadratic Regulator** (LQR): deterministic time-invariant linear system needs to minimize a quadratic cost over a finite horizon:

$$\min_{\pi} \quad V^{\pi}(0, \mathbf{x}_0) := \underbrace{\frac{1}{2}\mathbf{x}(T)^{\top}\mathbb{Q}\mathbf{x}(T)}_{\mathfrak{q}(\mathbf{x}(T))} + \int_0^T \underbrace{\frac{1}{2}\mathbf{x}(t)^{\top}Q\mathbf{x}(t) + \frac{1}{2}\mathbf{u}(t)^{\top}R\mathbf{u}(t)}_{\ell(\mathbf{x}(t),\mathbf{u}(t))}\,dt$$

$$\text{s.t. } \dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}, \quad \mathbf{x}(0) = \mathbf{x}_0,$$

$$\mathbf{x}(t) \in \mathbb{R}^n, \; \mathbf{u}(t) = \pi(t, \mathbf{x}(t)) \in \mathbb{R}^m$$

where $Q = Q^{\top} \succeq 0$, $\mathbb{Q} = \mathbb{Q}^{\top} \succeq 0$, and $R = R^{\top} \succ 0$

Linear ODE System

- Linear time-invariant ODE System:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t), \qquad \mathbf{x}(t_0) = x_0$$

- Transition matrix for LTI ODE system: $\Phi(t,s) = e^{A(t-s)}$
  - $\Phi(t,t) = I$
  - $\Phi^{-1}(t,s) = \Phi(s,t)$
  - $\Phi(t,s) = \Phi(t,t_0)\Phi(t_0,s)$
  - $\Phi(t_1 + t_2, s) = \Phi(t_1,s)\Phi(t_2,s) = \Phi(t_2,s)\Phi(t_1,s)$
  - $\frac{d}{dt}\Phi(t,s) = A\Phi(t,s)$

- Solution to LTI ODE system:

$$\mathbf{x}(t) = \Phi(t,t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \Phi(t,s)B\mathbf{u}(s)ds$$

LQR via the PMP

- Hamiltonian: $H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \frac{1}{2}\mathbf{u}^\top R\mathbf{u} + \mathbf{p}^\top A\mathbf{x} + \mathbf{p}^\top B\mathbf{u}$
- Canonical equations with boundary conditions:

$$\begin{aligned} \dot{\mathbf{x}} &= \nabla_p H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = A\mathbf{x} + B\mathbf{u}, & \mathbf{x}(0) &= \mathbf{x}_0 \\ \dot{\mathbf{p}} &= -\nabla_x H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = -Q\mathbf{x} - A^\top \mathbf{p}, & \mathbf{p}(T) &= \mathbb{Q}\mathbf{x}(T) \end{aligned}$$

- **PMP**:

$$\begin{aligned} \nabla_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) &= R\mathbf{u} + B^\top \mathbf{p} = 0 & \Rightarrow \quad & \mathbf{u}(t) = -R^{-1}B^\top \mathbf{p}(t) \\ \nabla^2_{\mathbf{u}} H(\mathbf{x}, \mathbf{u}, \mathbf{p}) &= R \succ 0 & \Rightarrow \quad & \mathbf{u}(t) \text{ is a minimum} \end{aligned}$$

- **Hamiltonian matrix**: the canonical equations can be simplified to a linear time-invariant (LTI) system with two-point boundary conditions:

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^\top \\ -Q & -A^\top \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix}, \quad \begin{aligned} \mathbf{x}(0) &= \mathbf{x}_0 \\ \mathbf{p}(T) &= \mathbb{Q}\mathbf{x}(T) \end{aligned}$$

LQR via the PMP

- **Claim**: There exists a matrix $M(t) = M(t)^T \succeq 0$ such that $\mathbf{p}(t) = M(t)\mathbf{x}(t)$ for all $t \in [0, T]$

- Solve the LTI system described by the Hamiltonian matrix backwards in time:

$$\begin{bmatrix} \mathbf{x}(t) \\ \mathbf{p}(t) \end{bmatrix} = \underbrace{e^{\begin{bmatrix} A & -BR^{-1}B^\top \\ -Q & -A^\top \end{bmatrix}(t-T)}}_{\Phi(t,T)} \begin{bmatrix} \mathbf{x}(T) \\ \mathbb{Q}\mathbf{x}(T) \end{bmatrix}$$

$$\mathbf{x}(t) = (\Phi_{11}(t, T) + \Phi_{12}(t, T)\mathbb{Q})\mathbf{x}(T)$$

$$\mathbf{p}(t) = (\Phi_{21}(t, T) + \Phi_{22}(t, T)\mathbb{Q})\mathbf{x}(T)$$

- Since $D(t, T) := \Phi_{11}(t, T) + \Phi_{12}(t, T)\mathbb{Q}$ is invertible for $t \in [0, T]$:

$$\mathbf{p}(t) = \underbrace{(\Phi_{21}(t, T) + \Phi_{22}(t, T)\mathbb{Q})D^{-1}(t, T)}_{=:M(t)}\mathbf{x}(t), \quad \forall t \in [0, T]$$

LQR via the PMP

- From $\mathbf{x}(0) = D(0,T)\mathbf{x}(T)$, we obtain an **open-loop control policy**:

$$\mathbf{u}(t) = -R^{-1}B^\top(\Phi_{21}(t,T) + \Phi_{22}(t,T)\mathbb{Q})D(0,T)^{-1}\mathbf{x}_0$$

- From $\mathbf{p}(t) = M(t)\mathbf{x}(t)$, however, we can also obtain a **closed-loop control policy**:

$$\mathbf{u}(t) = -R^{-1}B^\top M(t)\mathbf{x}(t)$$

- We can obtain a better description of $M(t)$ by differentiating $\mathbf{p}(t) = M(t)\mathbf{x}(t)$ and using the canonical equations:

$$\begin{aligned}
\dot{\mathbf{p}}(t) &= \dot{M}(t)\mathbf{x}(t) + M(t)\dot{\mathbf{x}}(t) \\
-Q\mathbf{x}(t) - A^\top\mathbf{p}(t) &= \dot{M}(t)\mathbf{x}(t) + M(t)A\mathbf{x}(t) - M(t)BR^{-1}B^\top\mathbf{p}(t) \\
-\dot{M}(t)\mathbf{x}(t) &= Q\mathbf{x}(t) + A^\top M(t)\mathbf{x}(t) + M(t)A\mathbf{x}(t) - M(t)BR^{-1}B^\top M(t)\mathbf{x}(t)
\end{aligned}$$

  which needs to hold for all $\mathbf{x}(t)$ and $t \in [0,T]$ and satisfy the boundary condition $\mathbf{p}(T) = M(T)\mathbf{x}(T) = \mathbb{Q}\mathbf{x}(T)$

LQR via the PMP (Summary)

- A unique candidate satisfies the necessary conditions of the PMP for optimality:

$$
\begin{aligned}
\mathbf{u}(t) &= -R^{-1}B^\top \mathbf{p}(t) \\
&= -R^{-1}B^\top(\Phi_{21}(t,T) + \Phi_{22}(t,T)\mathbb{Q})D(0,T)^{-1}\mathbf{x}_0 && \text{(open-loop)} \\
&= -R^{-1}B^\top M(t)\mathbf{x}(t) && \text{(closed-loop)}
\end{aligned}
$$

- The candidate policy is linear in the state and the matrix $M(t)$ satisfies a quadratic **Riccati differential equation** (RDE):

$$
-\dot{M}(t) = Q + A^\top M(t) + M(t)A - M(t)BR^{-1}B^\top M(t), \quad M(T) = \mathbb{Q}
$$

- The HJB PDE is needed to decide whether $\mathbf{u}(t)$ is globally optimal

LQR via the HJB PDE

- Hamiltonian: $H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \frac{1}{2}\mathbf{u}^\top R\mathbf{u} + \mathbf{p}^\top A\mathbf{x} + \mathbf{p}^\top B\mathbf{u}$

- HJB PDE for $t \in [0, T]$ and $\mathbf{x} \in \mathcal{X}$:

$$\begin{aligned}
\pi^*(t, \mathbf{x}) &= \arg\min_{\mathbf{u} \in \mathcal{U}} H(\mathbf{x}, \mathbf{u}, V_x(t, \mathbf{x})) = -R^{-1}B^\top V_x(t, \mathbf{x}), \\
-V_t(t, \mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{x}^\top A^\top V_x(t, \mathbf{x}) - \frac{1}{2}V_x(t, \mathbf{x})^\top BR^{-1}B^\top V_x(t, \mathbf{x}), \\
V(T, \mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top \mathbb{Q}\mathbf{x}
\end{aligned}$$

- Guess a solution to the HJB PDE based on the intuition from the PMP:

$$\begin{aligned}
\pi(t, \mathbf{x}) &= -R^{-1}B^\top M(t)\mathbf{x} \\
V(t, \mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top M(t)\mathbf{x} \\
V_t(t, \mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top \dot{M}(t)\mathbf{x} \\
V_x(t, \mathbf{x}) &= M(t)\mathbf{x}
\end{aligned}$$

LQR via the HJB PDE

- Substituting the candidate $V(t, \mathbf{x})$ into the HJB PDE leads to the same **RDE** as before and we know that $M(t)$ satisfies it!

$$\begin{aligned} \frac{1}{2}\mathbf{x}^\top M(T)\mathbf{x} &= \frac{1}{2}\mathbf{x}^\top \mathbb{Q}\mathbf{x} \\ -\frac{1}{2}\mathbf{x}^\top \dot{M}(t)\mathbf{x} &= \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{x}^\top A^\top M(t)\mathbf{x} - \frac{1}{2}\mathbf{x}^\top M(t)BR^{-1}B^\top M(t)\mathbf{x} \end{aligned}$$

- **Conclusion**: since $M(t)$ satisfies the RDE, $V(t, \mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M(t)\mathbf{x}$ is the unique solution to the HJB PDE and is the optimal value function for the LQR problem with associated optimal policy $\pi(t, \mathbf{x}) = -R^{-1}B^\top M(t)\mathbf{x}$

Continuous-Time Finite-Horizon LQG

- **Linear Quadratic Gaussian** (LQG) regulation problem:

$$\min_{\pi} \quad V^{\pi}(0, \mathbf{x}_0) = \frac{1}{2}\mathbb{E}\left\{ e^{-\frac{T}{\gamma}} \mathbf{x}(T)^\top \mathbb{Q} \mathbf{x}(T) + \int_0^T e^{-\frac{t}{\gamma}} \begin{bmatrix} \mathbf{x}^\top(t) & \mathbf{u}^\top(t) \end{bmatrix} \begin{bmatrix} Q & P^\top \\ P & R \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{u}(t) \end{bmatrix} dt \right\}$$

$$\text{s.t.} \;\; \dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} + C\boldsymbol{\omega}, \;\; \mathbf{x}(0) = \mathbf{x}_0,$$
$$\mathbf{x}(t) \in \mathbb{R}^n, \; \mathbf{u}(t) = \pi(t, \mathbf{x}(t)) \in \mathbb{R}^m$$

- **Discount factor**: $\gamma \in [0, \infty]$

- **Optimal value**: $V^*(t, \mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M(t)\mathbf{x} + m(t)$

- **Optimal policy**: $\pi^*(t, \mathbf{x}) = -R^{-1}(P + B^\top M(t))\mathbf{x}$

- **Riccati Equation**:

$$-\dot{M}(t) = Q + A^\top M(t) + M(t)A - (P + B^\top M(t))^\top R^{-1}(P + B^\top M(t)) - \frac{1}{\gamma}M(t), \qquad M(T) = \mathbb{Q}$$

$$-\dot{m} = \frac{1}{2}\operatorname{tr}(CC^\top M(t)) - \frac{1}{\gamma}m(t), \qquad m(T) = 0$$

- $M(t)$ is independent of the noise amplitude C, which implies that the optimal policy $\pi^*(t, \mathbf{x})$ is **the same for the stochastic LQG and deterministic LQR problems**!

Continuous-Time Infinite-Horizon LQG

- **Linear Quadratic Gaussian** (LQG) regulation problem:

$$\min_{\pi} \quad V^{\pi}(\mathbf{x}_0) := \frac{1}{2}\mathbb{E}\left\{\int_0^{\infty} e^{-\frac{t}{\gamma}} \begin{bmatrix}\mathbf{x}^\top(t) & \mathbf{u}^\top(t)\end{bmatrix} \begin{bmatrix} Q & P^\top \\ P & R \end{bmatrix} \begin{bmatrix}\mathbf{x}(t) \\ \mathbf{u}(t)\end{bmatrix} dt\right\}$$

$$\text{s.t.} \quad \dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} + C\boldsymbol{\omega}, \;\; \mathbf{x}(0) = \mathbf{x}_0$$

$$\mathbf{x}(t) \in \mathbb{R}^n, \; \mathbf{u}(t) = \pi(\mathbf{x}(t)) \in \mathbb{R}^m$$

- **Discount factor**: $\gamma \in [0, \infty)$

- **Optimal value**: $V^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M\mathbf{x} + m$

- **Optimal policy**: $\pi^*(\mathbf{x}) = -R^{-1}(P + B^\top M)\mathbf{x}$

- **Riccati Equation** ('care' in Matlab):

$$\frac{1}{\gamma}M = Q + A^\top M + MA - (P + B^\top M)^T R^{-1}(P + B^\top M)$$

$$m = \frac{\gamma}{2}\operatorname{tr}(CC^\top M)$$

- M is independent of the noise amplitude C, which implies that the optimal policy $\pi^*(\mathbf{x})$ is **the same for LQG and LQR**!

Relation Between Continuous-Time and Discrete-Time LQR

- The continuous-time system:

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}$$
$$\ell(\mathbf{x}, \mathbf{u}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \frac{1}{2}\mathbf{u}^\top R\mathbf{u}$$

  can be discretized with time step τ:

$$\mathbf{x}_{t+1} = (I + \tau A)\mathbf{x}_t + \tau B\mathbf{u}_t$$
$$\tau\ell(\mathbf{x}, \mathbf{u}) = \frac{\tau}{2}\mathbf{x}^\top Q\mathbf{x} + \frac{\tau}{2}\mathbf{u}^\top R\mathbf{u}$$

- In the limit as $\tau \to 0$, the discrete-time Riccati equation reduces to the continuous one:

$$\begin{aligned} M &= \tau Q + (I + \tau A)^\top M(I + \tau A) \\ &\qquad - (I + \tau A)^\top M\tau B(\tau R + \tau B^\top M\tau B)^{-1}\tau B^\top M(I + \tau A) \\ M &= \tau Q + M + \tau A^\top M + \tau MA - \tau MB(R + \tau B^\top MB)^{-1}B^\top M + o(\tau^2) \\ 0 &= Q + A^\top M + MA - MB(R + \tau B^\top MB)^{-1}B^\top M + \frac{1}{\tau}o(\tau^2) \end{aligned}$$