ECE276B: Planning & Learning in Robotics
Lecture 14: Continuous-time Optimal Control

Instructor:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Zhichao Li: zhl355@eng.ucsd.edu

Jinzhao Li: jil016@eng.ucsd.edu

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Continuous-time System Dynamics

- **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):

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

  defined by functions $\mathbf{f} : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^n$ 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 $\mathbf{g}(t)$

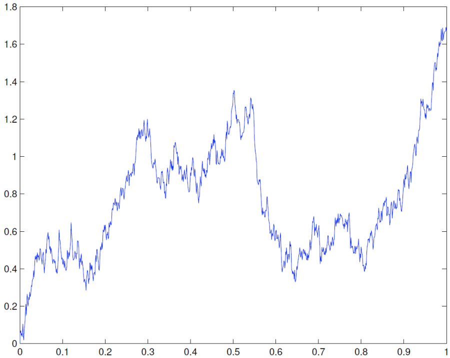
Brownian Motion

- 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 $\boldsymbol{\beta}(t) \sim \mathcal{GP}(\mathbf{0}, \min\{t, t'\}\, Q)$

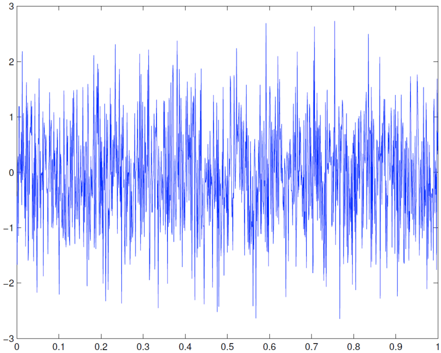
White Noise

- **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 path of $\boldsymbol{\omega}(t)$ is 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 formal derivative of Brownian motion: $d\boldsymbol{\beta}(t) = \boldsymbol{\omega}(t)dt$, where $\boldsymbol{\beta}(t) \sim \mathcal{GP}(\mathbf{0}, \min\{t, t'\}\, Q)$
- White noise is used to model the 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

- Infinite-dimensional dynamic constrained optimization:

$$\min_{\pi} V^{\pi}(0, \mathbf{x}_0) := \mathbb{E}\left\{ \int_0^T \underbrace{\ell(\mathbf{x}(t), \pi(t, \mathbf{x}(t)))}_{\text{stage cost}} dt + \underbrace{\mathfrak{q}(\mathbf{x}(T))}_{\text{terminal cost}} \;\middle|\; \mathbf{x}(0) = \mathbf{x}_0 \right\}$$

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

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

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

- **Problem variations**:
  - $\mathbf{x}(0)$ 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 or free for optimization
  - Additional state and control constraints can be imposed via $\mathcal{X}$ and $\mathcal{U}$

Assumptions

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

Examples: Existence and Uniqueness

- **Example**: Existence in not guaranteed in general

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

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

- **Example**: Uniqueness in not guaranteed in general

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

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

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 $\mathbf{y}(x)$ from $\mathbf{y}_0$ to $\mathbf{y}_f$ that minimizes a certain objective such as curve length or travel time for a particle accelerated by gravity (Brachistochrone Problem)

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

- Special case of continuous-time deterministic optimal control:
  - **fully-actuated** system: $\dot{\mathbf{x}} = \mathbf{u}$
  - **notation**: $\mathbf{x}(t) \leftarrow \mathbf{y}(x)$, $\mathbf{u}(t) = \dot{\mathbf{x}}(t) \leftarrow \dot{\mathbf{y}}(x)$

Optimal Value Function

- **Optimal policy**: $\mathbf{u}^*(t) := \pi^*(t, \mathbf{x}(t))$
- **Optimal value function**:

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

HJB PDE

The Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE) is satisfied for all time-state pairs $(t, \mathbf{x})$ by the optimal value function $V^*(t, \mathbf{x})$:

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

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

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

- The HJB PDE is the continuous-time analog of the Bellman Equation

HJB PDE Derivation

- A discrete-time approximation of the cont.-time optimal control problem can be used to derive the HJB PDE from the DP algorithm
- **Motion model**: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) + C(\mathbf{x}, \mathbf{u})\boldsymbol{\omega}$ 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 system dynamics**:

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

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

HJB PDE Derivation

- **Idea**: apply the Bellman Equation to the now discrete-time problem and take the limit as $\tau \to 0$ to obtain a "continuous-time Bellman Equation"

- **Bellman Equation**: finite-horizon problem with $t := k\tau$

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

- 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:

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

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

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

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

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

Infinite-Horizon 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: $H[\mathbf{x}, \mathbf{u}, \mathbf{p}(\cdot)] = \ell(\mathbf{x}, \mathbf{u}) + \mathbf{p}(\mathbf{x})^\top \mathbf{f}(\mathbf{x}, \mathbf{u}) + \frac{1}{2}\operatorname{tr}(\Sigma(\mathbf{x}, \mathbf{u})[\nabla_\mathbf{x} \mathbf{p}(\mathbf{x})])$

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

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

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

Existence and Uniqueness of HJB PDE Solutions

- The HJB PDE has at most one classical solution – a function which satisfies the PDE everywhere
- If a classical solution exists then it is the optimal value function
- The HJB PDE may not have a classical solution, in which case the optimal value function is not smooth (e.g., bang-bang control)
- The HJB PDE always has a unique viscosity solution which is the optimal value function
- Approximation schemes based on MDP discretization are guaranteed to converge to the unique viscosity solution
- Most continuous function approximation schemes (which scale better) are unable to represent non-smooth solutions
- All examples of non-smoothness seem to be deterministic, i.e., noise tends to smooth the optimal value function

Example 1: Guessing a Solution for the HJB PDE

- System: $\dot{x}(t) = u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $\ell(x, u) = 0$ 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:

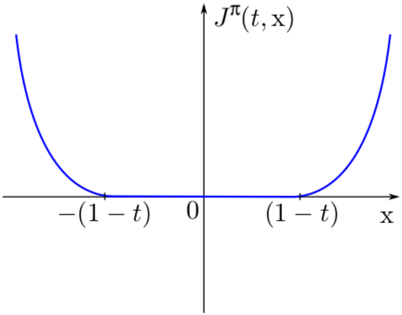
$$\pi(t, x) = -sgn(x) := \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$$

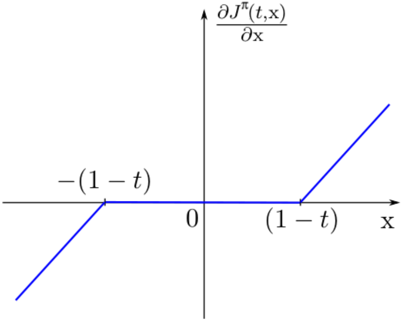
- The value in not smooth: $V^{\pi}(t, x) = \frac{1}{2}\left(\max\{0, |x| - (1 - t)\}\right)^2$
- We will verify that this function satisfies the HJB and is therefore indeed the optimal value function

Example 1: Partial Derivative wrt x

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

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

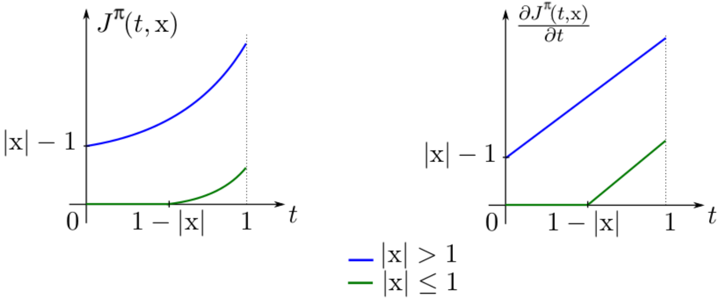




Example 1: Partial Derivative wrt t

- 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
- Solving the HJB PDE in general is non-trivial

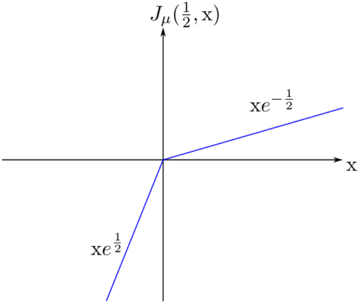
Example 2: HJB PDE without a Classical Solution

- System: $\dot{x}(t) = x(t)u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $\ell(x, u) = 0$ 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

Optimality Conditions

- The HJB PDE is not a necessary condition for optimality of the continuous-time optimal control problem but it is sufficient

Theorem: HJB PDE Sufficiency

Suppose that $V(t, \mathbf{x})$ is continuously differentiable in t and $\mathbf{x}$ and solves the HJB PDE:

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

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

for all $\mathbf{x} \in \mathcal{X}$ and $0 \le t \le T$. Suppose also that a policy $\pi^*(t, \mathbf{x})$ attains the minimum in the HJB PDE above for all t and $\mathbf{x}$ and is piecewise-continuous in t. Then, under the assumptions on Slide 7, $V(t, \mathbf{x})$ is the unique solution of the HJB PDE and is equal to the optimal value function $V^*(t, \mathbf{x})$, while $\pi^*(t, \mathbf{x})$ is an optimal policy.

Tractable Problems

- **Control-affine system dynamics**: $\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u} + C(\mathbf{x})\boldsymbol{\omega}$
- **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):

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

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

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

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

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

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

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

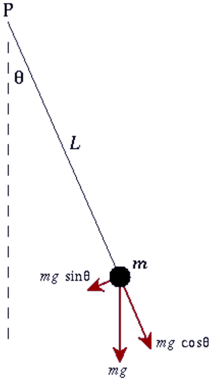
Example: Pendulum

- **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:

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