AA 203
Optimal and Learning-Based Control
The Hamilton-Jacobi-Bellman equation

Spencer M. Richards

Autonomous Systems Laboratory, Stanford University

April 26, 2023
(last updated May 3, 2023)

Agenda

1. The Bellman equation as a sufficient optimality condition
2. Continuous-time dynamic programming and the HJB equation
3. LQR control in continuous-time
4. Non-smooth value functions

Agenda

1. The Bellman equation as a sufficient optimality condition
2. Continuous-time dynamic programming and the HJB equation
3. LQR control in continuous-time
4. Non-smooth value functions

Review: Principle of optimality and dynamic programming

Consider the discrete-time OCP

$$\begin{aligned} \underset{\{u_t\}_{t=0}^{T-1}}{\text{minimize}} \;\; & \ell_T(x_T) + \sum_{t=0}^{T-1} \ell(t, x_t, u_t) \\ \text{subject to} \;\; & x_{t+1} = f(t, x_t, u_t),\ \forall t \in \{0, 1, \dots, T-1\} \\ & u_t \in \mathcal{U},\ \forall t \in \{0, 1, \dots, T-1\} \end{aligned}$$

Define the tail sub-problem *cost-to-go*

$$J(t, x, \{u_t\}_{k=t}^{T-1}) := \ell_T(x_T) + \sum_{k=t}^{T-1} \ell(k, x_k, u_k)$$

where $x_{k+1} = f(k, x_k, u_k)$ with initial condition $x_t = x$ is assumed implicitly.

Define the *value function* at $t \in \{0, 1, \dots, T\}$ and $x_t \in \mathbb{R}^n$ by

$$V(t, x_t) := \inf_{\{u_k\}_{k=t}^{T-1} \subseteq \mathcal{U}} J(t, x_t, \{u_k\}_{k=t}^{T-1}), \quad V(T, x_T) = \ell_T(x_T).$$

Previously we used "$J_t^*(x_t)$", but this notation will translate better to continuous-time later.

Review: Principle of optimality and dynamic programming

Suppose $\{u_t^*\}_{t=0}^{T-1} \subseteq \mathcal{U}$ is *globally optimal* for this OCP, i.e.,

$$J(0, x_0, \{u_t^*\}_{t=0}^{T-1}) = V(0, x_0).$$

Then the truncation $\{u_k^*\}_{k=t}^{T-1}$ is *globally optimal* for the corresponding tail sub-problem, i.e.,

$$J(t, x_t, \{u_k^*\}_{k=t}^{T-1}) = V(t, x_t).$$

From this, we must have the Bellman equation

$$V(t, x) = \inf_{u \in \mathcal{U}} \Big(\ell(t, x, u) + V(t+1, f(t, x, u)) \Big)$$

with the boundary condition $V(T, x) = \ell_T(x)$, for all $t \in \{0, 1, \ldots, T-1\}$ and $x \in \mathbb{R}^n$.

So the Bellman equation above is a *necessary* condition for *global* optimality of $\{u_t^*\}_{t=0}^{T-1}$.

The Bellman equation as a sufficient optimality condition

Suppose $\hat{V} : \{0, 1, \ldots, T\} \times \mathbb{R}^n \to \mathbb{R}$ is a function that satisfies the Bellman equation

$$\hat{V}(t, x) = \inf_{u \in \mathcal{U}} \Big(\ell(t, x, u) + \hat{V}(t+1, f(t, x, u)) \Big), \quad \hat{V}(T, x) = \ell_T(x),$$

for all $t \in \{0, 1, \ldots, T-1\}$ and $x \in \mathbb{R}^n$.

Suppose $(\hat{x}, \hat{u})$ satisfy $\hat{x}_{t+1} = f(t, \hat{x}_t, \hat{u}_t)$ with initial condition $\hat{x}_0$ and

$$\ell(t, \hat{x}_t, \hat{u}_t) + \hat{V}(t+1, f(t, \hat{x}_t, \hat{u}_t)) = \inf_{u \in \mathcal{U}} \Big(\ell(t, \hat{x}_t, u) + \hat{V}(t+1, f(t, \hat{x}_t, u)) \Big),$$

for all $t \in \{0, 1, \ldots, T-1\}$.

Then we can write

$$\hat{V}(t, \hat{x}_t) - \hat{V}(t+1, f(t, \hat{x}_t, \hat{u}_t)) = \ell(t, \hat{x}_t, \hat{u}_t).$$

The Bellman equation as a sufficient optimality condition

Then we can write

$$
\begin{aligned}
\hat{V}(t,\hat{x}_t) - \hat{V}(t+1, f(t,\hat{x}_t,\hat{u}_t)) &= \ell(t,\hat{x}_t,\hat{u}_t) \\
\hat{V}(t,\hat{x}_t) - \hat{V}(t+1,\hat{x}_{t+1}) &= \ell(t,\hat{x}_t,\hat{u}_t) \\
\sum_{t=0}^{T-1}\Big(\hat{V}(t,\hat{x}_t) - \hat{V}(t+1,\hat{x}_{t+1})\Big) &= \sum_{t=0}^{T-1}\ell(t,\hat{x}_t,\hat{u}_t) \\
\hat{V}(0,\hat{x}_0) - \hat{V}(T,\hat{x}_T) &= \sum_{t=0}^{T-1}\ell(t,\hat{x}_t,\hat{u}_t) \\
\implies \hat{V}(0,\hat{x}_0) &= \ell_T(\hat{x}_T) + \sum_{t=0}^{T-1}\ell(t,\hat{x}_t,\hat{u}_t) \\
&= J(0,\hat{x}_0,\{\hat{u}_t\}_{t=0}^{T-1})
\end{aligned}
$$

So $\hat{V}(0,\hat{x}_0)$ is the cost-to-go for $\{\hat{u}_t\}_{t=0}^{T-1}$.

The Bellman equation as a sufficient optimality condition

Now consider any other (x, u) satisfying $x_{t+1} = f(t, x_t, u_t)$ and $x_0 = \hat{x}_0$. Then

$$\begin{aligned}\hat{V}(t, x_t) &= \inf_{u \in \mathcal{U}} \Big(\ell(t, x_t, u) + \hat{V}(t+1, f(t, x_t, u)) \Big) \\ &\leq \ell(t, x_t, u_t) + \hat{V}(t+1, f(t, x_t, u_t))\end{aligned}$$

A similar summation argument gives us

$$\hat{V}(0, x_0) \leq J(0, x_0, \{u_t\}_{t=0}^{T-1}).$$

Since $x_0 = \hat{x}_0$, we have

$$\hat{V}(0, \hat{x}_0) \leq J(0, \hat{x}_0, \{u_t\}_{t=0}^{T-1}).$$

Overall, $\hat{u}$ yields the cost $\hat{V}(0, \hat{x}_0)$ and no other admissible u can produce a smaller cost.

While we chose $t = 0$ as the initial time, this was arbitrary since the Bellman equation is assumed to hold for all $t \in \{0, 1, \ldots, T\}$ and $x \in \mathbb{R}^n$. So $\hat{V}(t, x)$ is the optimal cost-to-go, i.e., $\hat{V}(t, x) = V(t, x)$ for all $t \in \{0, 1, \ldots, T\}$ and $x \in \mathbb{R}^n$, and $\hat{u}$ is a *globally optimal* control.

Agenda

1. The Bellman equation as a sufficient optimality condition
2. Continuous-time dynamic programming and the HJB equation
3. LQR control in continuous-time
4. Non-smooth value functions

Continuous-time dynamic programming

Consider the continuous-time OCP

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & \ell_T(x(T)) + \int_0^T \ell(t, x(t), u(t))\, dt \\ \text{subject to} \quad & \dot{x}(t) = f(t, x(t), u(t)), \ \forall t \in [0, T] \\ & u(t) \in \mathcal{U}, \ \forall t \in [0, T] \end{aligned}$$

Define the tail sub-problem *cost-to-go*

$$J(t, x, u_{[t,T]}) := \ell_T(x(T)) + \int_t^T \ell(s, x(s), u(s))\, ds$$

where $\dot{x}(s) = f(s, x(s), u(s))$ with initial condition $x(t) = x$ is assumed implicitly.

Define the *value function* at $t \in [0, T]$ and $x(t) \in \mathbb{R}^n$ by

$$V(t, x(t)) := \inf_{u_{[0,T]} \subseteq \mathcal{U}} J(t, x(t), u_{[0,T]}), \quad V(T, x(T)) = \ell_T(x(T)).$$

Continuous-time dynamic programming

Suppose $u^* : [0,T] \to \mathcal{U}$ is *globally optimal* for this OCP, i.e.,

$$J(0, x(0), u^*_{[0,T]}) = V(0, x(0)).$$

Then the truncation $u^*_{[t,T]} : [0,T] \to \mathcal{U}$ is *globally optimal* for the corresponding tail sub-problem, i.e.,

$$J(t, x(t), u^*_{[t,T]}) = V(t, x(t)).$$

From this, we must have the Bellman equation

$$V(t,x) = \inf_{u_{[t,t+\varepsilon]} \in \mathcal{U}} \left(\int_t^{t+\varepsilon} \ell(s, x(s), u(s))\, ds + V(t+\varepsilon, x(t+\varepsilon)) \right)$$

with the boundary condition $V(T,x) = \ell_T(x)$, for all $t \in [0,T)$, $\varepsilon \in (0, T-t]$, and $x \in \mathbb{R}^n$, where $x(s)$ for $s \in [t, t+\varepsilon]$ is the state trajectory corresponding to $u_{[t,t+\varepsilon]}$ with initial condition $x(t) = x$.

That is, the Bellman equation above is a *necessary* condition for *global* optimality of u^*.

From this, we must have the Bellman equation

$$V(t,x) = \inf_{u_{[t,t+\varepsilon]}\in\mathcal{U}} \left(\int_t^{t+\varepsilon} \ell(s,x(s),u(s))\,ds + V(t+\varepsilon, x(t+\varepsilon)) \right)$$

with the boundary condition $V(T,x) = \ell_T(x)$, for all $t \in [0,T)$, $\varepsilon \in (0, T-t]$, and $x \in \mathbb{R}^n$, where $x(s)$ for $s \in [t, t+\varepsilon]$ is the state trajectory corresponding to $u_{[t,t+\varepsilon]}$ with initial condition $x(t) = x$.

Assume V is $\mathcal{C}^1$-smooth with respect to t and x. Then

$$V(t+\varepsilon, x(t+\varepsilon)) = V(t,x) + \frac{\partial V}{\partial t}(t,x)\varepsilon + \nabla_x V(t,x)^{\mathsf{T}} f(t,x,u(t))\varepsilon + o(\varepsilon)$$

and

$$\int_t^{t+\varepsilon} \ell(s,x(s),u(s))\,ds = \ell(t,x,u(t))\varepsilon + o(\varepsilon),$$

where $\lim_{\varepsilon\to 0} \frac{o(\varepsilon)}{\varepsilon} = 0$.

The Hamilton-Jacobi-Bellman equation

Assume V is $\mathcal{C}^1$-smooth with respect to t and x. Then

$$V(t+\varepsilon, x(t+\varepsilon)) = V(t,x) + \frac{\partial V}{\partial t}(t,x)\varepsilon + \nabla_x V(t,x)^\mathsf{T} f(t,x,u(t))\varepsilon + o(\varepsilon)$$

and

$$\int_t^{t+\varepsilon} \ell(s, x(s), u(s))\, ds = \ell(t,x,u(t))\varepsilon + o(\varepsilon),$$

where "$o(\varepsilon)$" is little-o notation encapsulating terms that satisfy $\lim_{\varepsilon\to 0} \frac{o(\varepsilon)}{\varepsilon} = 0$.

Substitute these into the Bellman equation to get

$$-\frac{\partial V}{\partial t}(t,x)\varepsilon = \inf_{u_{[t,t+\varepsilon]}\in\mathcal{U}} \Big(\ell(t,x,u(t))\varepsilon + \nabla_x V(t,x)^\mathsf{T} f(t,x,u(t))\varepsilon + o(\varepsilon) \Big)$$

Divide by ε and take the limit as $\varepsilon \to 0$ to get the *Hamilton-Jacobi-Bellman (HJB) equation*

$$-\frac{\partial V}{\partial t}(t,x) = \inf_{u\in\mathcal{U}} \Big(\ell(t,x,u) + \nabla_x V(t,x)^\mathsf{T} f(t,x,u) \Big)$$

with boundary condition $V(T,x) = \ell_T(x)$, for all $t \in [0,T)$ and $x \in \mathbb{R}^n$.

The Hamilton-Jacobi-Bellman equation

The *Hamilton-Jacobi-Bellman (HJB) equation* is

$$-\frac{\partial V}{\partial t}(t,x) = \inf_{u \in \mathcal{U}} \left(\ell(t,x,u) + \nabla_x V(t,x)^\mathsf{T} f(t,x,u) \right)$$

with boundary condition $V(T,x) = \ell_T(x)$, for all $t \in [0,T)$ and $x \in \mathbb{R}^n$.

If we assume a globally optimal control exists, then we can replace "$\inf$" with "$\min$" above, and the HJB equation is a necessary condition for global optimality of $u^* : [0,T] \to \mathcal{U}$. A similar derivation to the discrete-time case can show that the HJB equation can be used to form *sufficient* optimality conditions.

Define the *Hamiltonian*

$$H(t,x,u,p) \coloneqq p^\mathsf{T} f(t,x,u) - \ell(t,x,u).$$

Then we can rewrite the HJB equation as

$$\frac{\partial V}{\partial t}(t,x) = \sup_{u \in \mathcal{U}} H(t,x,u,-\nabla_x V(t,x)).$$

HJB versus PMP

The *Hamilton-Jacobi-Bellman (HJB) equation* is

$$\frac{\partial V}{\partial t}(t,x) = \sup_{u \in \mathcal{U}} H(t, x, u, -\nabla_x V(t,x))\Big)$$

with boundary condition $V(T,x) = \ell_T(x)$, for all $t \in [0,T)$ and $x \in \mathbb{R}^n$.

In the PMP, we saw that an optimal control u^* must satisfy

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

This is an *open-loop* specification, since u^* depends on the state and co-state trajectories, which come from solving a BVP over the entire interval $[0,T]$.

With the HJB, we have that

$$u^*(t) = \arg\max_{u \in \mathcal{U}} H(t, x^*(t), u, -\nabla_x V(t, x^*(t))), \ \forall t \in [0,T].$$

This is a *closed-loop* specification, since if we know $V(t,x)$ everywhere, then $u^*(t)$ is completely determined by $x^*(t)$.

HJB versus PMP

In the PMP, we saw that an optimal control u^* must satisfy

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

This is an *open-loop* specification, since u^* depends on the state and co-state trajectories, which come from solving a BVP over the entire interval $[0, T]$.

With the HJB, we have that

$$u^*(t) = \arg\max_{u \in \mathcal{U}} H(t, x^*(t), u, -\nabla_x V(t, x^*(t))),\ \ \forall t \in [0, T].$$

This is a *closed-loop* specification, since if we know $V(t, x)$ everywhere, then $u^*(t)$ is completely determined by $x^*(t)$. However, computing $V(t, x)$ everywhere is much harder to do; it is the solution of a PDE, while the PMP required us to solve a system of ODEs.

Comparing the PMP and the HJB also gives us a new interpretation of the adjoint state as a *sensitivity*

$$p^*(t) = -\nabla_x V(t, x^*(t)).$$

Agenda

1. The Bellman equation as a sufficient optimality condition
2. Continuous-time dynamic programming and the HJB equation
3. LQR control in continuous-time
4. Non-smooth value functions

LQR control in continuous-time

Consider the continuous-time OCP

$$\begin{aligned} \underset{u}{\text{minimize}} \;\; & \frac{1}{2}x(t)^\mathsf{T} Q_T x(t) + \frac{1}{2}\int_0^T \big(x(t)^\mathsf{T} Q(t)x(t) + u(t)^\mathsf{T} R(t)u(t)\big)\, dt \\ \text{subject to} \;\; & \dot{x}(t) = A(t)x(t) + B(t)u(t), \;\; \forall t \in [0, T] \end{aligned}$$

where $Q_T \succeq 0$, $Q(t) \succeq 0$, and $R(t) \succ 0$ for all $t \in [0, T]$.

The HJB equation for this problem is

$$-\frac{\partial V}{\partial t}(t, x) = \inf_{u \in \mathbb{R}^m} \left(\frac{1}{2}x^\mathsf{T} Q(t)x + \frac{1}{2}u^\mathsf{T} R(t)u + \nabla_x V(t, x)^\mathsf{T}(A(t)x + B(t)u) \right)$$

with boundary condition $V(T, x) = \frac{1}{2}x^\mathsf{T} Q_T x$.

Take the derivative with respect to u and set it equal to zero to get

$$u = -R(t)^{-1} B(t)^\mathsf{T} \nabla_x V(t, x)$$

LQR control in continuous-time

The HJB equation for this problem is

$$-\frac{\partial V}{\partial t}(t,x) = \inf_{u\in\mathbb{R}^m}\left(\frac{1}{2}x^\mathsf{T}Q(t)x + \frac{1}{2}u^\mathsf{T}R(t)u + \nabla_x V(t,x)^\mathsf{T}(A(t)x + B(t)u)\right)$$

with boundary condition $V(T,x) = \frac{1}{2}x^\mathsf{T}Q_T x$.

Take the derivative with respect to u and set it equal to zero to get

$$u = -R(t)^{-1}B(t)^\mathsf{T}\nabla_x V(t,x)$$

Substitute this back into the HJB equation and rearrange to get

$$\frac{\partial V}{\partial t}(t,x) = \frac{1}{2}\nabla_x V(t,x)^\mathsf{T}B(t)R(t)^{-1}B(t)^\mathsf{T}\nabla_x V(t,x) - x^\mathsf{T}A(t)^\mathsf{T}\nabla_x V(t,x) - \frac{1}{2}x^\mathsf{T}Q(t)x,$$

which must hold for all (t,x) with boundary condition $V(T,x) = \frac{1}{2}x^\mathsf{T}Q_T x$.

LQR control in continuous-time

Substitute this back into the HJB equation and rearrange to get

$$\frac{\partial V}{\partial t}(t,x) = \frac{1}{2}\nabla_x V(t,x)^\top B(t)R(t)^{-1}B(t)^\top \nabla_x V(t,x) - x^\top A(t)^\top \nabla_x V(t,x) - \frac{1}{2}x^\top Q(t)x,$$

which must hold for all (t,x) with boundary condition $V(T,x) = \frac{1}{2}x^\top Q_T x$.

Based on the boundary condition, let us make the *ansatz* $V(t,x) = \frac{1}{2}x^\top P(t)x$, where $P(t)$ is symmetric positive-definite. Then the HJB equation becomes

$$\begin{aligned}\frac{1}{2}x\dot{P}(t)x &= \frac{1}{2}x^\top P(t)B(t)R(t)^{-1}B(t)^\top P(t)x - x^\top A(t)^\top P(t)x - \frac{1}{2}x^\top Q(t)x \\ &= \frac{1}{2}x^\top\Big(P(t)B(t)R(t)^{-1}B(t)^\top P(t) - P(t)A(t) - A(t)^\top P(t) - Q(t)\Big)x\end{aligned}$$

This must hold for all (t,x), so $P(t)$ must satisfy the *continuous-time Riccati equation*

$$\dot{P}(t) = P(t)B(t)R(t)^{-1}B(t)^\top P(t) - P(t)A(t) - A(t)^\top P(t) - Q(t),$$

which is an ODE that can be solved backwards in time from $P(T) = Q_T$.

LQR control in continuous-time

The value function is $V(t,x) = \frac{1}{2}x^\top P(t)x$, where $P(t) \succ 0$ must satisfy the *continuous-time Riccati equation*

$$\dot{P}(t) = P(t)B(t)R(t)^{-1}B(t)^\top P(t) - P(t)A(t) - A(t)^\top P(t) - Q(t),$$

which is an ODE that can be solved backwards in time from $P(T) = Q_T$.

The optimal control is

$$u^* = -R(t)^{-1}B(t)^\top \nabla_x V(t,x) = \underbrace{-R(t)^{-1}B(t)^\top P(t)}_{=:K(t)} x,$$

which is a *linear feedback policy*.

Recall that in the discrete-time case we had to solve the *discrete-time Riccati equation*

$$P_t = Q_t + A_t^\top P_{t+1}A_t - A_t^\top P_{t+1}B_t(R_t + B_t^\top P_{t+1}B_t)^{-1}B_t^\top P_{t+1}A_t$$

recursively from the boundary condition $P_T = Q_T$. The optimal control input in this case was $u^* = K_t x$ with $K_t \coloneqq -(R_t + B_t^\top P_{t+1}B_t)^{-1}B_t^\top P_{t+1}A_t$.

Agenda

Non-smooth value functions

Recall that in deriving the HJB equation we assumed that $V(t,x)$ is $\mathcal{C}^1$-smooth with respect to t and x. However, this is often not true, particularly for problems with bounded control and a terminal cost.

As an example, consider the scalar problem

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & x(T) \\ \text{subject to} \quad & \dot{x}(t) = x(t)u(t), \ \forall t \in [0,T] \\ & u(t) \in [-1,1], \ \forall t \in [0,T] \end{aligned}$$

By inspection,

$$u^* = \begin{cases} 1, & x < 0 \\ ?, & x = 0 \\ -1, & x > 0 \end{cases} \implies \dot{x}^* = \begin{cases} x, & x < 0 \\ 0, & x = 0 \\ -x, & x > 0 \end{cases} \implies x^*(t) = \begin{cases} e^{t-t_0} x_0, & x_0 < 0 \\ 0, & x_0 = 0 \\ e^{-(t-t_0)} x_0, & x_0 > 0 \end{cases}$$

Non-smooth value functions

As an example, consider the scalar problem

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & x(T) \\ \text{subject to} \quad & \dot{x}(t) = x(t)u(t), \ \forall t \in [0, T] \\ & u(t) \in [-1, 1], \ \forall t \in [0, T] \end{aligned}$$

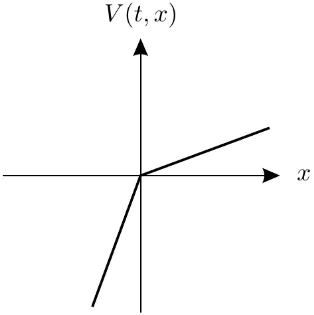
The value function is

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

which is not $\mathcal{C}^1$-smooth, but it does satisfy the HJB equation

$$-\frac{\partial V}{\partial t}(t,x) = \inf_{u \in [-1,1]} \nabla_x V(t,x) x u = -|\nabla_x V(t,x) x|$$

away from $x = 0$, with boundary condition $V(T, x) = x$.



$V(t,x)$ for fixed t.

Non-smooth value functions

It turns out that the HJB equation

$$-\frac{\partial V}{\partial t}(t,x) = \inf_{u\in\mathcal{U}} \Big(\ell(t,x,u) + \nabla_x V(t,x)^\mathsf{T} f(t,x,u) \Big)$$

with boundary condition $V(T,x) = \ell_T(x)$ can have non-smooth solutions. However, we must reinterpret what we mean by a solution.

We say V is a *viscosity solution* to the HJB if for each (t,x) we have

$$-\frac{\partial \overline{\varphi}}{\partial t}(t,x) - \inf_{u\in\mathcal{U}} \Big(\ell(t,x,u) + \nabla_x \overline{\varphi}(t,x)^\mathsf{T} f(t,x,u) \Big) \leq 0$$

$$-\frac{\partial \underline{\varphi}}{\partial t}(t,x) - \inf_{u\in\mathcal{U}} \Big(\ell(t,x,u) + \nabla_x \underline{\varphi}(t,x)^\mathsf{T} f(t,x,u) \Big) \geq 0$$

for all $\mathcal{C}^1$-smooth test functions $\overline{\phi}$ and $\underline{\phi}$ such that $\overline{\varphi} - V$ has a local minimum at (t,x) and $\underline{\varphi} - V$ has a local maximum at (t,x). More details can be found in (Liberzon, 2012, §5.3).

With appropriate technical assumptions on f, ℓ, ℓ_T, and $\mathcal{U}$, the value function V is the unique viscosity solution of the HJB equation and it is *locally Lipschitz*.

Next class

System identification and adaptive control

References

D. Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, 2012.