31

CLASSICAL OPTIMAL CONTROL THEORY

31.1 Basic problem

We start with the basic formulation of a basic optimal control problem, which aims to seek an optimal control policy that maximizes accumulated gain during a dynamical process.

Definition 31.1.1 (basic optimal control problem). *Given a dynamic system*

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

the basic optimal control problem is to maximize the performance measure

$$\max_{u(x,t)} J(x_0, u(x,t)) = h(x_{t_f}, t_f) + \int_{t_0}^{t_f} g(x(t), u(x,t), t) dt$$

The functional relationship $u^*(x,t) = f(x(t), t)$ that maximize J is called ***optimal control policy*** *.*

For different concrete types of performance measure function g, see [1, p. 30]. Another common optimal control problem is with respect to an infinite horizon process.

Definition 31.1.2 (optimal control problem infinite horizon under discount). *Given a dynamic system $\dot{x}(t) = a(x(t), u(t), t), x(0) = x_0$, the optimal control problem for infinite horizon is to maximize the performance measure*

$$\max_{u(x)} J(x_0, u(x_0)) = \int_{t_0}^{\infty} e^{-\gamma(t-t_0)} g(x(t), u(x), t) dt,$$

where $\gamma \in (0,1)$ is the discount factor, and the functional relationship $u^*(x) = f(x)$ that maximize J is called ***optimal control policy****.*

In the basic optimal control problem, the control policy is time dependent. For infinite horizon problem, the optimal control policy does not have time dependence. If the optimal control is a function of initial state x_0 and t, that is $u^*(t) = f(x_0, t)$, then the optimal control is **open-loop control.**

31.2 Controllability & observability

A fundamental property of dynamical systems in the context of optimal control is its **controllability**. Intuitively, a dynamical system is controllable we can use finite steps of control to reach any states.

Definition 31.2.1 (controllability for discrete-time linear system). *A n dimensional discrete-time system*

$$x(k+1) = Ax(k) + Bu(k)$$

is said to be ***completely controllable*** *if for $x(0) = 0$ and given x_1, there exists a finite index N and sequence of control inputs $u(0), u(1), ..., u(N-1)$ such that this input sequence will yield $x(N) = x_1$.*

Note that the choice of the initial condition $x(0) = 0$ will not lose generality, because for other initial condition we can always arrive at that state using finite steps. Given a linear system, we have linear algebra tool to examine its controllability, as we show below.

Theorem 31.2.1 (controllability criterion). *[2, p. 278]A discrete-time linear system is completed controllable if and ony if the $n \times nm$ controllability matrix*

$$M = [B, AB, ..., A^{n-1}B]$$

has rank n.

Proof. Suppose a sequence of inputs $u(0), u(1), ..., u(N-1)$ is applied to the system, with $x(0) = 0$. It follows

$$x(N) = A^{N-1}Bu(0) + A^{N-2}Bu(1) + \cdots + Bu(N-1).$$

From here, we can see points in the state space can be reached if and only if they can be expressed as linear combinations of columns of M. It can be showed that $N = n$ will suffice (see reference). □

Remark 31.2.1 (caution when u is constrained)**.** The above theorem assumes that admissible u is unconstrained. If u is constrained, the theorem will not apply.

31.3 Dynamic programming principle

31.3.1 Principle of optimality

Theorem 31.3.1 (principle of optimality for trajecotries). *[1, p. 54]Let $a-b-e$ be an optimal trajectory in the state space from a to e with associated cost $J_{abc}*$, then $b-e$ is the optimal path from b to e.*

Proof: Suppose there is another path $b-f-e$ with less cost than the cost of $b-e$, then the total cost for $a-b-e$ can be reduced, which is a contradiction.

31.3.2 The Hamilton-Jacobi-Bellman equation (finite horizon)

Although the goal of optimal control problem is to seek optimal control policy u that maximize process gain

$$\max_{u(x,t)} h(x_{t_f}, t_f) + \int_{t_0}^{t_f} g(x(t), u(x,t), t)dt,$$

it is usually intractable to directly solve for u. In the Hamilton-Jacobi-Bellman equation framework, we first derive the governing equation for value function $V(x(t), t)$, which is the maximum process gain if the system starts from $x(t)$ at time t. Then u can be solved via V, as we show in the following.

Theorem 31.3.2 (HJB for finite horizon process). *[1, p. 88] Let value function $V(x(t), t)$ be*

$$V(x(t), t) = \min_{u(\tau), t\leq\tau\leq t_f} [\int_t^{t_f} g(x(\tau), u(\tau), \tau)d\tau + h(x(t_f), t_f)].$$

Then the HJB equation is given as

$$0 = V_t + \min_{u(x,t)} [g(x(t), u(x,t), t) + V_x\dot{x}]$$

with boundary condition

$$V(x(t_f), t_f) = h(x(t_f), t_f).$$

With solved V, $u(x,t)$ is given by

$$u(x,t) = \arg\min_{u(x,t)} g(x(t), u(x,t), t) + V_x \dot{x}.$$

Proof. Let

$$V(x(t),t) = \min_{u(\tau), t\leq\tau\leq t_f} [\int_t^{t_f} g(x(\tau), u(\tau), \tau) d\tau + h(x(t_f), t_f)]$$

By subdividing the interval, we have

$$\begin{aligned} V(x(t),t) &= \min_{u(\tau), t\leq\tau\leq t_f} [\int_t^{t_f} g(x(\tau), u(\tau), \tau) d\tau + h(x(t_f), t_f)] \\ &= \min_{u(\tau), t\leq\tau\leq t_f} [\int_t^{t+dt} g(x(\tau), u(\tau), \tau) d\tau \\ &+ \int_{t+dt}^{t_f} g(x(\tau), u(\tau), \tau) d\tau + h(x(t_f), t_f)] \\ &= \min_{u(t)} [g(x(t), u(t), t) dt + V(x(t+dt), t+dt)] \\ &= \min_{u(t} [g(x(t), u(t), t) dt + V(x(t), t) + V_t dt + V_x \dot{x} dt] \end{aligned}$$

Then, we have the HJB equation

$$0 = V_t + \min_{u(t)} [g(x(t), u(t), t) + V_x \dot{x}]$$

with boundary condition

$$V(x(t_f), t_f) = h(x(t_f), t_f)$$

□

Remark 31.3.1. The function $V(x(t),t)$ is not a function of u since it is the already the minimum value.

31.3.3 The Hamilton-Jacobi-Bellman equation (infinite horizon)

In the infinite horizon setting, the value function $V(x(t),t)$ can be showed to be independent of t. Therefore, it can be written as $V(x(t))$.

Lemma 31.3.1 (time independence of value function). *Define the value function*

$$V(x(t),t) = \min_{u(\tau),t\leq\tau}[\int_t^\infty \exp(-\gamma(\tau-t))g(x(\tau),u(\tau),\tau)d\tau]$$

then V only depends on $x(t_0)$.

Proof.

$$\begin{aligned}
V(x(t),t) &= \min_{u(\tau),t\leq\tau}[\int_t^\infty \exp(-\gamma(\tau-t))g(x(\tau),u(\tau),\tau)d\tau] \\
&= \min_{u(s),0\leq s}[\int_0^\infty \exp(-\gamma s)g(x(s+t),u(s+t),s+t)ds] \\
&= \min_{u(s),0\leq 0}[\int_0^\infty \exp(-\gamma s)g(x(s),u(s),s)ds = V(x(0),0)
\end{aligned}$$

where we use variable substitution and the time invariance of g. □

Theorem 31.3.3 (HJB for infinite horizon process). *Let*

$$V(x(t),t) = \min_{u(\tau),t\leq\tau}[\int_t^\infty \exp(-\gamma(\tau-t))g(x(\tau),u(\tau),\tau)d\tau],$$

then HJB equation

$$0 = \min_{u(t)}[g(x(t),u(t),t) - \gamma V + V_x^T\dot{x}]$$

with boundary condition $V(x(t),t) = C, \forall x \in X$

Proof. Let

$$V(x(t),t) = \min_{u(\tau),t\leq\tau}[\int_t^\infty \exp(-\gamma(\tau-t))g(x(\tau),u(\tau),\tau)d\tau]$$

By subdividing the interval, we have

$$
\begin{aligned}
V(x(t),t) &= \min_{u(\tau),t\leq\tau}[\int_t^{t_f} \exp(-\gamma(\tau-t)g(x(\tau),u(\tau),\tau)d\tau] \\
&= \min_{u(\tau),t\leq\tau}[\int_t^{t+dt} \exp(-\gamma dt)g(x(\tau),u(\tau),\tau)d\tau \\
&+ \exp(-\gamma dt)\int_{t+dt}^{\infty} \exp(-\gamma(\tau-t-dt)g(x(\tau),u(\tau),\tau)d\tau] \\
&= \min_{u(t)}[g(x(t),u(t),t)dt + \exp(-\gamma dt)V(x(t+dt),t+dt)] \\
&= \min_{u(t)}[g(x(t),u(t),t)dt + \exp(-\gamma dt)V(x(t),t) + V_x\dot{x}dt]
\end{aligned}
$$

Then, we have the HJB equation

$$0 = \min_{u(t)}[g(x(t),u(t),t) - \gamma V + V_x\dot{x}]$$

with boundary condition $V(x(t),t) = C, \forall x \in X$ where we have used the time independence property of V, and $\exp(-\gamma dt) = 1 - \gamma dt$ □

Remark 31.3.2. If $\gamma = 0$, then there is no discount.

31.4 Deterministic linear quadratic control

31.4.1 Linear quadratic control (finite horizon)

Definition 31.4.1 (finite horizon linear quadratic control). *Consider the system state equation given as*

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

and we want to minimize

$$J = \frac{1}{2}x^T(t_f)Hx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} x^T(t)Qx(t) + u^T(t)R(t)u(t)dt$$

where H and Q are real symmetric positive semi-definite matrices, R is a real symmetric positive definite matrix.

Remark 31.4.1. Note that matrix R has to be positive definite to eliminate the situation that $u(t)$ blows up in order to minimize J.

Theorem 31.4.1 (HJB equation for finite horizon linear quadratic control). *Define the value function*

$$V(x(t),t) = \min_{u(\tau),t\leq\tau\leq t_f}[\frac{1}{2}x^T(t_f)Hx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} x^T(t)Qx(t) + u^T(t)R(t)u(t)dt]$$

The HJB equation is given as

$$0 = V_t + \frac{1}{2}x^TQx - \frac{1}{2}V_x^TBR^{-1}B^TV_x + V_x^TAx$$

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

Proof. Use Theorem 31.3.2, we have

$$0 = V_t + \min_{u(t)}[g(x(t),u(t),t) + V_x^T\dot{x}].$$

Note that $\dot{x} = Ax + Bu$, the minimize

$$\frac{1}{2}x^T(t)Qx(t) + \frac{1}{2}u^T(t)R(t)u(t) + V_x^T(Ax + Bu)$$

over u. The minimizer is given by $u^* = -R^{-1}B^TV_x$. Plug in u^* and we will get the result.

□

Remark 31.4.2 (solution to HJB). We can propose a solution with quadratic form $V(x(t),t) = \frac{1}{2}x^T H(t)x$ and solve the form of $H(t)$. Also see [1, p. 93] for details.

31.4.2 Linear quadratic control(infinite horizon)

Definition 31.4.2 (infinite horizon linear quadratic control). *Consider the system state equation given as*

$$\dot{x}(t) = Ax(t) + Bu(t)$$

and we want to minimize

$$J = \frac{1}{2}\int_{t_0}^{\infty} \exp(-\gamma t)[x^T(t)Qx(t) + u^T(t)R(t)u(t)]dt$$

where H and Q are real symmetric positive semi-definite matrices, R is a real symmetric positive definite matrix, and γ is the discount factor($\gamma = 0$ means no discount).

Remark 31.4.3. Note that R has to be positive definite to eliminate the situation that $u(t)$ blows up in order to minimize J.

Theorem 31.4.2. *The HJB equation for the infinite horizon linear quadratic control problem is given as*

$$\gamma V = \frac{1}{2}x^T Qx - \frac{1}{2}V_x^T BR^{-1}B^T V_x + V_x^T Ax$$

with boundary condition $V(x(t_0) = 0, t_0) = 0$.

Proof. Use Theorem 31.3.3, we have

$$\gamma V = \min_{u(t)}[g(x(t), u(t), t) + V_x^T \dot{x}].$$

Note that $\dot{x} = Ax + Bu$, the minimize

$$\frac{1}{2}x^T(t)Qx(t) + \frac{1}{2}u^T(t)R(t)u(t) + V_x^T(Ax + Bu)$$

over u. The minimizer is given by $u^* = -R^{-1}B^T V_x$. Plug in u^* and we will get the result. □

Remark 31.4.4 (solution methods).

- See [1, p. 213][3] for details on how to solve this nonlinear algebraic equations.

- For infinite horizon case will give a ordinary differential equation instead of a partial differential equation in finite horizon case.
- We can use finite difference method to solve this ODE.Note that in every interior node, we have a algebraic equation.

31.5 Continuous-time stochastic optimal control

31.5.1 HJB equation for general nonlinear systems

Definition 31.5.1 (general nonlinear system control). *[4, p. 421]*

- *We are given a continuous-time* $n-$*dimensional dynamic system*

$$\dot{x}(t) = f(x(t), u(t), t) + L(t)w(t), x(0) = x_0$$

where $L(t) \in \mathbb{R}^{n\times s}$*, and random disturbance* $w(t)$ *satisfying*

$$E[w(t)] = 0, E[w(t)w(\tau)^T] = W(t)\delta(t-\tau)$$

- *The goal is to minimize*

$$J = E[\phi(x(t_f), t_f) + \int_{t_0}^{t_f} \mathcal{L}(x(t), u(t), t)dt]$$

by choosing $u(t)$ *as the control input. The* ϕ *is the terminal cost and* $\mathcal{L}(x, u, t)$ *is the instantaneous cost function.*

Definition 31.5.2 (value function). *The value function* $V(x, t)$ *is defined over the state space and the time interval* $[t, t_f]$*,given as*

$$V(x(t), t) = \min_{u(t), t\in[t,t_f]} E[\int_{t}^{t_f} \mathcal{L}(x(\tau), u(\tau), \tau)d\tau]$$

Remark 31.5.1 (interpretation). The value function is a deterministic function and is the expected optimal cost for the system starting at $x(t)$ at time t.

Theorem 31.5.1 (Hamilton-Jacobi-Bellman (HJB) equation). *Under optimal control, the value function of the optimal trajectories must satisfy the following HJB equation given as:*

$$\partial_t V(x, t) = \min_{u(t)} \{\mathcal{L}(x, u, t) + \nabla_x V(x, t)^T f(x, u) + \frac{1}{2} Tr[\nabla_x^2 V(x, t) L(t) W(t) L(t)^T]\}$$

Proof.

$$\begin{aligned}
& V(x+\Delta x, t+\Delta t) \\
=& V(x,t) + \partial_t V(x,t)\Delta t + \nabla_x V(x,t)^T \Delta x + \frac{1}{2}\Delta x^T \nabla_x^2 V(x,t)\Delta x + o(\Delta t) \\
=& V + \partial_t V \Delta t + \nabla_x V^T (f+Lw)\Delta t + (f+Lw)^T \nabla_x^2 V(x,t)(f+Lw)(\Delta t)^2 + o(\Delta t) \\
=& V + \partial_t V \Delta t + \nabla_x V^T (f+Lw)\Delta t + (f+Lw)^T \nabla_x^2 V(x,t)(f+Lw)(\Delta t)^2 + o(\Delta t)
\end{aligned}$$

where we use $\Delta x = (f+Lw)\Delta t$, the trace of a scalar is the scalar itself and the cyclic rule of matrix trace [Lemma A.8.8]. □

31.5.2 Linear Gaussian quadratic system

Definition 31.5.3 (linear Gaussian quadratic control). *[4, p. 421]*

- *We are given a continuous-time $n-$dimensional dynamic system*

$$\dot{x}(t) = Fx + Gu + Lw$$

where $L(t) \in \mathbb{R}^{n\times s}$, and random disturbance $w(t)$ satisfying

$$E[w(t)] = 0, E[w(t)w(\tau)^T] = W(t)\delta(t-\tau)$$

- *The goal is to minimize*

$$J = \frac{1}{2}E[x^T(t_f)S_f x(t_f) + \int_{t_0}^{t_f} [x(t)^T \; u(t)^T] \begin{bmatrix} Q(t) & M(t) \\ M(t)^T & R(t) \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}$$

by choosing $u(t)$ as the control input. The $R(t), Q(t)$ are symmetric matrices and $R(t)$ is required to be positive definite.

Theorem 31.5.2 (Hamilton-Jacobi-Bellman (HJB) equation). *Under optimal control, the value function of the optimal trajectories must satisfy the following HJB equation given as:*

$$\partial_t V(x,t) = -\min_{u(t)} \frac{1}{2}\{x^T Q x + 2x^T M u + u^T R u + x^T S(Fx + Gu) + Tr(SLWL^T)\}$$

Proof. (use Theorem 31.5.1]. □

31.6 Stochastic dynamic programming

31.6.1 Discrete-time Stochastic dynamic programming: finite horizon

Definition 31.6.1 (basic problem of finite horizon). *[5, p. 12]*

- *We are given a discrete-time dynamic system*

$$x_{k+1} = f_k(x_k, u_k, w_k)$$

where the state x_k is an element of a space S_k, the control u_k is an element in the control space C_k, and random disturbance w_k is an element of a space D_k.

- *A control policy π is consisting of a sequence of functions*

$$\pi = \{\mu_0, \mu_1, ..., \mu_N\}$$

where $\mu_k : S_k \to C_k$ is a function maps states x_k to $u_k = \mu_k(x_k)$.

- *For given reward function $g_k, k = 0, 1, ..., N$, the expected cost of π starting at x_0 is*

$$J_\pi(x_0) = E[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)]$$

where the expectation is taken over the joint distribution of all w_k and x_k.

- *The goal is to find an optimal control policy π^* such that*

$$J_{\pi^*}(x_0) = \min_\pi J_\pi(x_0)$$

Theorem 31.6.1 (Principle of Optimality). *[5, p. 18] Let $\pi^* = \{\mu_0^*, \mu_1^*, ..., \mu_N^*\}$ be a optimal policy for the basic problem, and assume that when using π^*, a given state x_i has positive probability. Then the truncated policy $\{\mu_i^*, \mu_{i+1}^*, ..., \mu_N^*\}$ is optimal for the subproblem starting at x_i*

$$E[g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k)]$$

Lemma 31.6.1 (dynamic programming algorithm for basic problem of finite horizon). *The optimal cost function J^* and its associated optimal control policy $\pi^* = \{\mu_0^*, \mu_1^*, ..., \mu_{N-1}^*\}$ can be calculated using the following backward induction procedures:*

$$J_N^*(x_N) = g_N(x_N)$$

$$J_k^*(x_k) = \min_{\mu_k(x_k)} E_{w_k}[g_k(x_k, u_k, w_k) + J_{k+1}^*(f_k(x_k, \mu_k(x_k), w_k))], k = 0, 1, ..., N-1$$

Proof. Directly from principle of optimality. □

Remark 31.6.1 (interpretation). The lemma provides a way to calculate the optimal control policy.

Lemma 31.6.2 (monotonicity property of dynamic programming I). *If we change the final cost g_N to an uniformly larger cost function g_N'(i.e. $g_N'(x) \geq g_N(x), \forall x$), then all optimal cost function J_k^* will be uniformly increasing(at least not decreasing).*

Similar situation holds when g_N is changed to an uniformly smaller one.

Proof. Obviously $J_N^{*'} = g_N'$ will uniformly increase. For other k with induction,

$$J_k^{*'} = \min E[g_k + J_{k+1}^{*'}] \geq \min E[g_k + J_{k+1}^*] = J_k^*$$

□

Lemma 31.6.3 (monotonicity property of dynamic programming II). *[5, p. 60] Consider the basic problem with all functions and sets being time-invariant($S_k = S, g_k = g, f_k = f, ...$). If in the dynamic programming algorithm we have*

$$J_{N-1}^*(x) \leq J_N^*(x), \forall x \in S$$

then

$$J_k^*(x) \leq J_{k+1}^*(x), \forall x \in S, \forall k.$$

Similarly, if

$$J_{N-1}^*(x) \geq J_N^*(x), \forall x \in S$$

then

$$J_k^*(x) \geq J_{k+1}^*(x), \forall x \in S, \forall k.$$

31.6.2 Discrete-time stochastic dynamic programming: infinite horizon

31.6.2.1 *Fundamentals*

Definition 31.6.2 (basic problem of infinite horizon). *[6, p. 3]*

- *We are given a **stationary** discrete-time dynamic system*

$$x_{k+1} = f(x_k, u_k, w_k)$$

where the state x_k is an element of a space S, the control u is an element in the control space C, and random disturbance w_k is an element of a space D.
- *A **stationary** control policy π is consisting of a sequence of functions*

$$\pi = \{\mu, \mu, ...\}$$

where $\mu : S \to C$ is a function maps states x_k to $u_k = \mu(x_k)$.
- *For a given cost function $g, k = 0, 1, ..., N$, the expected cost of π starting at x_0 is*

$$J_\pi(x_0) = \lim_{N\to\infty} E_{w_k, k=1,...,N}[\sum_{k=0}^{N} \alpha^k g(x_k, \mu(x_k), w_k)]$$

where $\alpha \in [0, 1)$ is the discount factor, the expectation is taken over the joint distribution of all w_k and x_k.
- *The goal is to find an optimal control policy π^* such that*

$$J_{\pi^*}(x_0) = \min_\pi J_\pi(x_0)$$

Remark 31.6.2 (what stationarity means?).

- Compared to finite horizontal problem, infinite horizon problem requires the dynamical system to be time invariant.
- If $f(x_k, u_k, w_k)$ is state dependent but not time dependent, then the dynamic system is still time-invariant. For example, we can have $f(x_k, u_k, w_k) = A(x_k)x_k + B(x_k)u_k + L(x_k)w_k$, or write as $f(x, u, w) = A(x)x + B(x)u + L(x)w$

Definition 31.6.3 (dynamic programming operator).

- $(TJ)(x) = \min_{u\in U(x)} E[g(x, u, w) + \alpha J(f(x, u, w))]$
- $(T_\mu J)(x) = E[g(x, u, w) + \alpha J(f(x, \mu(x), w))]$

31.6.2.2 *Convergence analysis*

Lemma 31.6.4 (Monotonicity lemma). *[6, p. 9] For any functions $J, J' : X \to \mathbb{R}$ such that for all $x \in X$,*

$$J(x) \leq J'(x)$$

and any stationary policy $\mu : X \to U$, we have

$$(T^k J)(x) \leq (T^k J')(x)$$

and

$$(T^k J)(x) \leq (T^k J')(x)$$

for all $x \in X$ and all $k = 1, 2, ...$

Proof. For $k = 1$, we can show its correctness. For other k use induction. □

Lemma 31.6.5 (constant shift lemma). *[6, p. 9] For every k, function $J : X \to \mathbb{R}$, stationary policy μ, scalar $r \in \mathbb{R}$, and $x \in X$, we have*

$$(T^k(J + r))(x) = (T^k J)(x) + \alpha^k r$$
$$(T^k_\mu(J + r))(x) = (T^k_\mu J)(x) + \alpha^k r$$

Proof. For $k = 1$, we can show that

$$(T(J + r))(x) = (T^k J)(x) + \alpha r$$
$$(T_\mu(J + r))(x) = (T^k_\mu J)(x) + \alpha r$$

Then we can use induction for other k. □

Theorem 31.6.2 (dynamic programming operator as a contraction mapping). *[5, p. 18] The following two operators defined as the space of bounded functions of $J : X \to \mathbb{R}$*

- $(TJ)(x) = \min_{u \in U(x)} E[g(x, u, w) + \alpha J(f(x, u, w))]$
- $(T_\mu J)(x) = E[g(x, u, w) + \alpha J(f(x, \mu(x), w))]$

are contracting mappings with respect to the sup-norm/max-norm. Note that the expectation is taken respect to distribution of w.

Proof. Denote

$$c = \max_{x \in X} \left| J(x) - J'(x) \right|,$$

so that for all $x \in X$, we have

$$J(x) - c \leq J'(x) \leq J(x) + c$$

Apply T and use Monotonicity and constant shift lemma, we have

$$TJ - \alpha c \leq TJ' \leq J + \alpha c, \forall x \in X$$

Therefore

$$\left|TJ - TJ'\right| \leq \alpha c$$

and

$$\max\left|TJ - TJ'\right| \leq \alpha \max\left|J - J'\right|$$

□

Corollary 31.6.2.1 (convergence rate). *[6, p. 18] For any two bounded functions $J, J' : X \to \mathbb{R}$, we have*

$$\max_{x \in X}\left|(T^k J)(x) - (T^k J')(x)\right| \leq \alpha^k \max_{x \in X}\left|(J)(x) - (J')(x)\right|$$

Corollary 31.6.2.2 (convergence rate). *[6, p. 18] For any two bounded functions $J, J' : X \to \mathbb{R}$ and any stationary policy μ, we have*

$$\max_{x \in X}\left|(T_\mu^k J)(x) - (T_\mu^k J')(x)\right| \leq \alpha^k \max_{x \in X}\left|(J)(x) - (J')(x)\right|$$

Remark 31.6.3 (interpretation of convergence)**.**

- Any initial J is guaranteed to converge.
- The convergence rate depends on the initial distance between J and J^*, and the discount factor. In the extreme case of $\alpha = 0$, convergence is one single step.

31.7 Notes on bibliography

For introductory treatment on classical control theory, see [2][1]. For application of optimal control theory in finance, see [7][8][9][5]. For advanced treatment on this topic, see [10]. For an introduction to calculus of variations, see [1]. For treatment of linear state space control, see [11]. For certainty equivalence, see [5, p. 160]. For dynamic programming theory, see [12]. For reinforcement learning, see [13].

BIBLIOGRAPHY

1. Kirk, D. E. *Optimal control theory: an introduction* (Courier Corporation, 2012).
2. Luenberger, D. *Introduction to dynamic systems: theory, models, and applications* (Wiley, 1979).
3. Wikipedia. *Algebraic Riccati equation — Wikipedia, The Free Encyclopedia* [Online; accessed 1-August-2016]. 2016.
4. Stengel, R. F. *Optimal control and estimation* (Courier Corporation, 2012).
5. Bertsekas, D. *Dynamic Programming and Optimal Control* ISBN: 9781886529083 (Athena Scientific, 2012).
6. Bertsekas, D. *Dynamic Programming and Optimal Control Athena Scientific optimization and computation series* **v. 2.** ISBN: 9781886529441 (Athena Scientific, 2012).
7. Miranda, M. J. & Fackler, P. L. *Applied computational economics and finance* (MIT press, 2004).
8. Chang, F.-R. *Stochastic optimization in continuous time* (Cambridge University Press, 2004).
9. Pham, H. *Continuous-time stochastic control and optimization with financial applications* (Springer Science & Business Media, 2009).
10. Fleming, W. H. & Soner, H. M. *Controlled Markov processes and viscosity solutions* (Springer Science & Business Media, 2006).
11. Williams, R. L., Lawrence, D. A., *et al. Linear state-space control systems* (John Wiley & Sons, 2007).
12. Bertsekas, D. P. *Abstract dynamic programming* (Athena Scientific, 2018).
13. Wiering, M. & Van Otterlo, M. Reinforcement learning. *Adaptation, Learning, and Optimization* **12** (2012).