<4. Dynamic Programming and HJB Equation>

- 목차
  - Continuous-Time Dynamic System
    - 정의
    - 비용함수
    - 이산화 및 DP 알고리즘 적용
  - Taylor-series ⊕ Hamilton-Jacobi-Bellman equation
    - 식 유도
    - 정리 4.1 + 예시
    - 정리 4.2

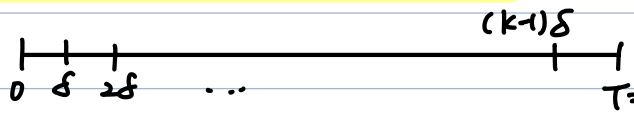
* Continuous-Time Dynamic System

- Notation
  - $s(t)$: state trajectory, $s(t) \in S$
  - $a(t)$: control trajectory, $a(t) \in A$
  - t : time, $0 \le t \le T$
- $s(0) = s_0$ 일 때, $\dot{s}(t) = f(s(t), a(t))$, f는 s에 대해 미분 가능 & a에 대해 연속

* Continuous-Time Optimal Control

- 비용함수 $h(s(T)) + \int_0^T g(s(t), a(t))\, dt$ 를 최소화하는 $a(t), s(t)$가 optimal control.
  - ~~Q. g는 보상 아닌가?~~
  - A. 여기서는 비용으로 사용. $s(t), a(t)$ 실행해서 발생한 비용 $g(s,a)$
- (ex) . t 시기에 $s(t)$를 생산하는 생산과는 $s(t)$의 일부 $a(t)$를 재투자해, $1-a(t)$를 storable good 생산에 할당 가능. Q. 왜 이렇게 나옴?
  - . 따라서 $s(t)$는 $\frac{ds(t)}{dt} = \gamma a(t) s(t)$ 로 표현 가능하며, γ는 상수.
  - . 생산과는 $0 \le a(t) \le 1$, $t \in [0, T]$ 라고 할 때 $\int_0^T (1-a(t)) s(t)\, dt$ 라는 전체 상품의 양을 (비율) 최대화하고 싶어함.

* Continuous time → Discrete



- $s_{k+1} = s_k + f(s_k, a_k)\delta$, $\text{cost} = h(s_k) + \sum_{k=0}^{k-1} g(s_k, a_k)\delta$
- $t = k\delta$ 일 때 최적의 cost-to-go 함수는 → 앞으로의 비용만 고려

$$\tilde{V}_k(s_k) = \min_{a \in A} \left[h(s_k) + \sum_{i=k}^{k-1} g(s_i, a_i)\delta \right] \quad \forall k = 0, 1, \ldots, k-1$$

$k\delta$ 시점부터 $(k-1)\delta$ 까지...

$$\tilde{V}_k(s_k) = h(s(T))$$

원래 비용함수 : $h(s(T)) + \int_0^T g(s(t), a(t))\, dt$

근데 $t = k\delta(T)$ 일 때는 오로지 유지비용만 들어감.

* DP 알고리즘 적용

DP는 $V_t = \sim + V_{t+1}$ 꼴로 표현

- $t = k\delta$ 일 때 비용함수는

$$\tilde{V}_k(s) = \min_{a \in A} \left[g(s,a)\delta + \tilde{V}_{k+1}(s + f(s,a)\delta) \right] \quad \forall k = 0, 1, \ldots, k-1$$

$$\tilde{V}_K(s) = h(s(T))$$

* 테일러 급수 적용 (1차)

- $T_f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{6}f^{(3)}(a)(x-a)^3 \ldots$

Q. 이 식이 어떻게 나옴?

- $\tilde{V}_{k+1}(s + f(s,a)\delta) = \tilde{V}_k(s) + \nabla_t \tilde{V}_k(s) \cdot \delta + \nabla_s \tilde{V}_k(s)' f(s,a)\delta + o(\delta)$
  - ↳ $(\underset{\sim}{t}\delta, s)$ 에 대해서 미분 진행.

- $\tilde{V}_k(s) = \min_{a \in A} \left[g(s,a)\delta + \tilde{V}_k(s) + \nabla_t \tilde{V}_k(s)\delta + \nabla_s \tilde{V}_k(s)' f(s,a)\delta + o(\delta) \right]$

* Hamilton – Jacobi – Bellman equation

- 이제 $\delta \to 0$ 으로 보내서 continuous-time 으로 만들어줌

$$\lim_{k \to \infty, \delta \to 0} \tilde{V}_k^*(s) = \tilde{V}_t^*(s)$$

- $0 = \min_{a \in A} \left[g(s,a) + \nabla_t V_t(s) + \nabla_s V_t(s)' f(s,a) \right], \quad V_T(s) = h(s)$

- 위의 식이 HJB 방정식이며, 모든 (t,s) 쌍에 대해 비용함수 $V_t(s)$ 가 충족되어야 하는 미분 방정식임.
- 하지만 $V_t(s)$ 가 미분 가능한지는 사전에 확인 불가.

* HJB- 정리 4.1 : Sufficient optimality condition

- $V_t(s)$가 HJB의 해이고, $a^*(t,s)$가 비용함수를 최소화 시킨다고 가정. → optimal control policy
- 초기 상태 $s(0)$가 주어졌을 때 state-control 경로를 $\{s^*(t) \mid t \in [0,T]\}$, $\{a^*(t) \mid t \in [0,T]\}$ 라 할 때, $V_t(s) = V^*_t(s)$ 이고, $\{a^*(t) \mid t \in [0,T]\}$가 최적의 control 경로임.
- HJB를 풀어서 $a^*(t,s)$ 얻는 것은 어려우므로, 후보 policy가 최적인지 아닌지를 HJB를 통해 증명해야 함.
- 만약 HJB로 V가 최적 해임을 보이면, $V = V^*$ 이며 $a^*(t,s)$가 각각의 (t,s)에 대해 V를 최소화하는 control policy 임.

* 벨만 방정식 예시

- $|a(t)| \le 1$ 이고, $\dot{s}(t) = a(t)$ 일 때 비용함수는 $\frac{1}{2}(s(T))^2$. → 역시 바로. Q. why?
- 모든 t,s에 대해서 HJB는 $0 = \min_{|a| \le 1} \left[\nabla_t V_t(s) + \nabla_s V_t(s)\, a \right]$, $V_T(s) = \frac{1}{2}s^2$

- An evident candidate for optimality
  - state를 0을 향해 빠르게 이동시키기
- 상응하는 control policy : $a(t,s) = -\text{sign}(s)$
- 〃 cost-to-go 함수 : $V_t(s) = \frac{1}{2}\left(\max\{0, |s| - (T-t)\}\right)^2$
- $V_t(s)$가 HJB 방정식 만족한다 할 때, $a^*(t,s) = -\text{sgn}(s)$가 RHS 최소 달성.

  ∴ state & control trajectories are optimal.

Q. 다 무슨 말인지 모르겠어...

* HJB-정리 4.2 : Pontryagin Minimum Principle : Necessary optimality condition

@. 여기도 예시 무슨말?

- $a^*(t)$가 optimal 하고 $S^*(0) = S_0$가 주어질 때,

$$\dot{s}^*(t) = f(S^*(t), a^*(t)), \quad 0 \le t \le T$$

-