

8. CONTINUOUS TIME DYNAMIC PROGRAMMING.

ABSTRACT DEFINITION:

STATE $x \in \mathcal{X}$

ACTION $a \in \mathcal{A}$

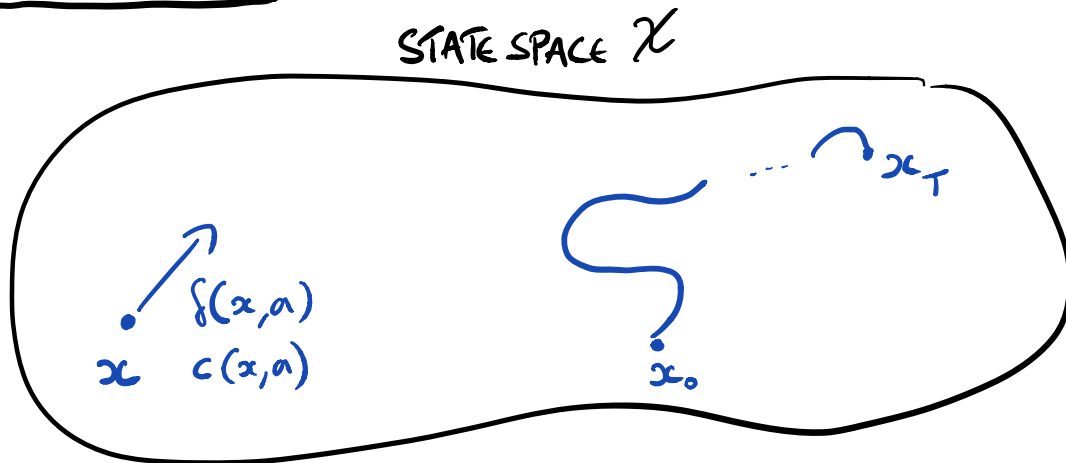
COST $c(x, a)$

CHANGE IN STATE

$$\frac{dx}{dt} = f(x, a)$$

POLICY $\pi(x)$ OR π_t

OBJECTIVE: [MAXIMIZE SUM OF REWARDS]



$$L_T(x_0) := \max \underbrace{\int_0^{T-\omega t} e^{-\alpha t} c(x_t, \pi_t) dt + e^{-\alpha T} c(x_T)}_{C_T(x, \pi)} \text{ over } a_t \in \mathcal{A}, t \in \mathbb{R}_+$$

↑
VALUE FUNCTION

THE HAMILTON-JACOBI-BELLMAN EQUATION (THE HJB EQUATION).

Def 60 (Hamilton-Jacobi-Bellman Equation). For a continuous-time dynamic program , the equation

$$0 = \min_{a \in \mathcal{A}} \{c(x, a) + \partial_t L_t(x) + f(x, a) \partial_x L_t(x) - \alpha L_t(x)\} \quad (\text{HJB})$$

is called the Hamilton-Jacobi-Bellman equation. It is the continuous time analogue of the Bellman equation

A HEURISTIC DERIVATION OF THE HJB EQN.

WE USE TAYLOR APPROXIMATIONS [A LOT!]:

$$x_{t+\delta} - x_t \approx f(x_t, a_t) \quad (+)$$

&

$$C_T(x_0, a) \approx \sum_{t \in \{0, \delta, \dots, T-\delta\}} (1-\alpha\delta)^{\frac{t}{\delta}} c(x_t, a_t) \delta + (1-\alpha\delta)^{\frac{t}{\delta}} c(x_T) \quad (H)$$

THE BELLMAN EQUATION FOR THE DYNAMIC PROGRAM WITH COSTS (+) & DYNAMIC (H) IS

$$\underline{L_t(x)} = \min_a \left\{ c(x, a) \delta + \underline{(1-\alpha\delta) L_{t+\delta}(x + \delta f(x, a))} \right\} \div \delta$$

NOTICE THAT

$$\frac{(1-\alpha\delta) L_{t+\delta}(x + \delta f(x,a)) - L_t(x)}{\delta}$$

$$\xrightarrow{\delta \rightarrow 0} \partial_t L_t(x) + f(x,a) \partial_x L_t(x) - \alpha L_t(x)$$

SO THE BELLMAN EQUATION BECOMES:

$$0 = \min_a \left\{ c(x,a) + \partial_t L_t(x) + f(x,a) \partial_x L_t(x) - \alpha L_t(x) \right\}$$

THEOREM: SUPPOSE $C_t(x, \pi)$ FOR A POLICY π SATISFIES THE HJB EQUATION $\forall x, t$ THEN π IS OPTIMAL.

PROOF: LET $\tilde{x}_t, \tilde{\pi}_t$ BE FROM SOME OTHER POLICY.

$$\begin{aligned} -\frac{d}{dt} (e^{-\alpha t} C_t(\tilde{x}_t, \Pi)) &= e^{-\alpha t} \{c_t(\tilde{x}_t, \tilde{\pi}_t) - [c_t(\tilde{x}_t, \tilde{\pi}_t) - \alpha C + f_t(\tilde{x}_t, \tilde{\pi}_t) \partial_x C + \partial_t C]\} \\ &\leq e^{-\alpha t} c_t(\tilde{x}_t, \tilde{\pi}_t) \end{aligned}$$

INTEGRATING GIVES

$$C_0(x_0, \pi) - \underbrace{e^{-\alpha T} C_T(x)}_{\text{sp}} \leq \int_0^T e^{-\alpha t} c(\tilde{x}_t, \tilde{\pi}_t) dt$$

$$\therefore C_0(x_0, \pi) \leq C_0(x_0, \tilde{\pi})$$

LINEAR QUADRATIC REGULARIZATION (CONTINUOUS TIME)

Def 62 (LQ problem). We consider a dynamic program of the form

$$\begin{array}{ll} \text{Minimize} & \int_0^T [x_t Q x_t + a_t R a_t] dt + x_T Q_T x_T \\ \text{subject to} & \frac{dx_t}{dt} = A x_t + B a_t, \quad t \in \mathbb{R}_+ \\ \text{over} & a_t \in \mathbb{R}^m, \quad t \in \mathbb{R}_+. \end{array} \quad (\text{LQ})$$

Here $x_t \in \mathbb{R}^n$ and $a_t \in \mathbb{R}^m$. A and B are matrices. Q and R symmetric positive definite matrices. This is a Linear-Quadratic problem (LQ problem).

Def 63 (Riccati Equation). The differential equation with

$$\dot{\Lambda}(t) = -Q - \Lambda(t)A - A^\top \Lambda(t) + \Lambda(t)BR^{-1}B^\top \Lambda(t) \quad \text{and} \quad \Lambda(T) = Q_T. \quad (\text{RicEq})$$

is called the Riccati equation.

Thrm 64. For each time t , the optimal action for the LQ problem is

$$a_t = -R^{-1}B^T \Lambda(t)x_t,$$

where $\Lambda(t)$ is the solution to the Riccati equation.

PROOF: The HJB EQN IS

$$0 = \min_a \{ x^T Q x + a^T R a + \partial_t L_t(x) + (Ax + Ra)^T \partial_x L_t(x) \}$$

We "GUESS" $L_t(x) = x^T \Lambda(t) x$

$$\therefore \partial_x L_t(x) = 2\Lambda(t)x \quad \& \quad \partial_t L_t(x) = x^T \dot{\Lambda}(t)x$$

SUBSTITUTING GIVES

$$0 = \min_a \{ a^T R a + 2x^T \Lambda(t) B a \} + x^T Q x + x^T \dot{\Lambda}(t)x + 2x^T \Lambda(t) A x$$

$$\Rightarrow a^* = -R^{-1}B^T \Lambda(t)x \quad \leftarrow [AS ABOVE]$$

PROOF (CONT.)

FINALLY SUBST. BACK a^* GIVES

$$0 = x^T \left\{ Q + \dot{\Lambda}(t) + \Lambda(t)A + A^T \Lambda(t) - \Lambda(t)BR^{-1}B^T \Lambda(t) \right\} x$$

WHICH IMPLIES THE RICCATI EQUATION MUST HOLD

□.

