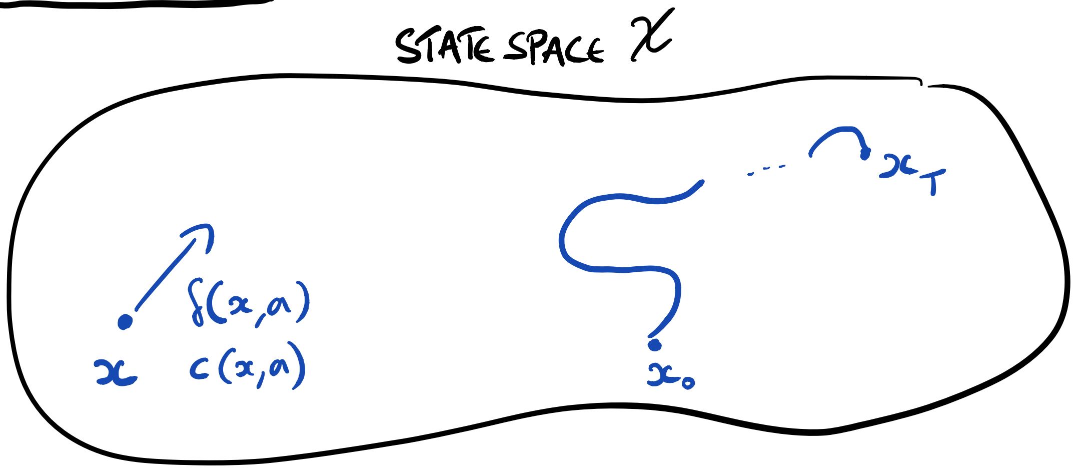
8. CONTINUOUS TIME DYNAMIC PROGRAMMING.

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]

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

VALUE FUNCTION

THE HAMILTON-JACOBI-BELLMAN EQUATION

(THE HJB EQUATION).

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

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

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

A HEURISTIC DERIVATION OF THE HJB EQN.

WE USE TAYLOR APPROXIMATIONS [A LOT!]:

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

&

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

THE BELLMAN EQUATION FOR THE DYNAMIC PROGRAM WITH COSTS (†) & DYNAMIC (‡) IS

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

NOTICE THAT

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

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

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.

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

INTEGRATING GIVES

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

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

LINEAR QUADRATIC REGULARIZATION (CONTINUOUS TIME)

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

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

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 an Linear-Quadratic problem (LQ problem).

Def 63 (Riccarti Equation). *The differential equation with*

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

is called the Riccarti equation.

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

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

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

PROOF: THE HJB EQN IS

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

WE "GUESS" $L_t(x) = x^\top \Lambda(t) x$

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

SUBSTITUTING GIVES

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

$$\Longrightarrow \quad a^* = -R^{-1}B^\top \Lambda(t) x$$ ← [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 RICCARTI EQUATION MUST HOLD $\square$.