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** \( \pi_t(x) \) or \( \pi_t \)

**Objective**: [Maximize sum of rewards]

\[
L_T(x_0) := \max \int_0^T c(x_t, \pi_t) \, dt + c(x_T) \text{ over } a_t \in \mathcal{A}, t \in \mathbb{R}_+
\]

**Value Function**

\[ c(x, \pi_t) \]
THE HAMILTON-JACOBI-BELLMAN EQUATION

( THE HJB EQUATION ).

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

\[ 0 = \min_{a \in A} \left\{ c(x, a) + \partial_t L_t(x) + f(x, a) \partial_x L_t(x) - \alpha L_t(x) \right\} \]  

(HJB)

is called the Hamilton-Jacobi-Bellman equation. It is the continuous time analogue of the Bellman equation.
A HEURISTIC DERIVATION OF THE HJB EQUATION.

We use Taylor approximations: [4.107.17]:

\[ x_{t+\delta} - x_t \approx f(x_t, a_t) \]  

\[ C_t(x_0, a) \approx \sum_{t \in \{0, \delta, \ldots, T-t\}} (1 - \alpha \delta)^\frac{t}{\delta} c(x_t, a_t) \delta + (1 - \alpha \delta)^\frac{t}{\delta} c(x_T) \]  

The Bellman equation for the dynamic program with costs (4) & Dynamic (4) is

\[ L_t(x) = \min_a \left\{ c(x, a) \delta + (1 - \alpha \delta) L_{t+\delta} (x + \delta f(x, a)) \right\} / \delta \]
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:

\[
0 = \min_a \left\{ c(x,a) + \partial_t L_t(x) + f(x,a) \partial_x L_t(x) - \alpha L_t(x) \right\}
\]
\textbf{Theorem:} Suppose $C_t(x, \pi)$ for a policy $\pi$ satisfies the HJB equation $\forall x, t$ then $\pi$ is optimal.

\textbf{Proof:} Let $\bar{x}_t, \bar{\pi}_t$ be from some other policy.

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

Integrating gives

\[
C_0(x_0, \pi) - e^{x^T} C_T(x) \leq \int_0^T e^{-at} C(\bar{x}_t, \bar{\pi}_t) \, dt
\]

\[
\therefore \quad C_0(x_0, \pi) \leq C_0(x_0, \bar{\pi})
\]
**Def 62** (LQ problem). We consider a dynamic program of the form

Minimize \[\int_0^T \left[ x_t Q x_t + a_t R a_t \right] dt + x_T Q_T x_T \] \hspace{1cm} \text{(LQ)}

subject to \[\frac{dx_t}{dt} = Ax_t + B a_t, \quad t \in \mathbb{R}_+ \]

over \[a_t \in \mathbb{R}^m, \quad t \in \mathbb{R}_+. \]

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 an Linear-Quadratic problem (LQ problem).

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

\[\dot{\Lambda}(t) = -Q - \Lambda(t)A - A^T \Lambda(t) + \Lambda(t)BR^{-1}B^T \Lambda(t) \quad \text{and} \quad \Lambda(T) = Q_T. \] \hspace{1cm} \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

$$D = \min_a \left\{ x^TQx + a^TRa + \partial_t L_t(x) + (Ax + Ra)^T \partial_x L_t(x) \right\}$$

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

$$D = \min_a \left\{ a^TRa + 2x^T\Lambda(t)Ba \right\} + x^TQx + x^T \dot{\Lambda}(t)x + 2x^T\Lambda(t)Ax$$

$$\Rightarrow a^* = -R^{-1}B^T \Lambda(t)x \quad \leftarrow \text{[as above]}$$
Finally subst. back $a^*$ gives

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

which implies the Riccati equation must hold $\square$. 