15. The Hamilton-Jacobi-Bellman equation

Let $S \subseteq \mathbb{R}^d$ and $A \subseteq \mathbb{R}^p$. A <u>continuous-time controllable dynamical system</u>, with state-space S and action-space A is a map

$$b : \mathbb{R}^+ \times S \times A \longrightarrow \mathbb{R}^d.$$

If we choose action a when in state x at time t, then $b(t,x,a)$ will have the interpretation as our instantaneous velocity.

We assume that b is continuous, and is differentiable in $x \in S$ with bounded derivative.

Proposition

Let $b : [0,T] \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be continuous and suppose that, for some $K < \infty$, for all $t \in [0,T]$ and $x, y \in \mathbb{R}^d$,

$$|b(t,x) - b(t,y)| \leq K|x-y|.$$

Then, for all $x_0 \in \mathbb{R}^d$, there is a unique differentiable function $x : [0,T] \longrightarrow \mathbb{R}^d$ such that $x(0) = x_0$ and

$$\dot{x}(t) = b(t, x(t)), \qquad 0 \leq t \leq T.$$

There is a proof, not examinable in this course, in the printed notes, §18.

A control is a continuous map

$$u : \mathbb{R}^+ \longrightarrow A.$$

Set

$$b^u(t,x) = b(t,x,u_t).$$

Assume $S = \mathbb{R}^d$. Then $b^u : \mathbb{R}^+ \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is continuous and (by the mean value theorem)

$$|b^u(t,x) - b^u(t,y)| \leq K|x-y|,$$

where

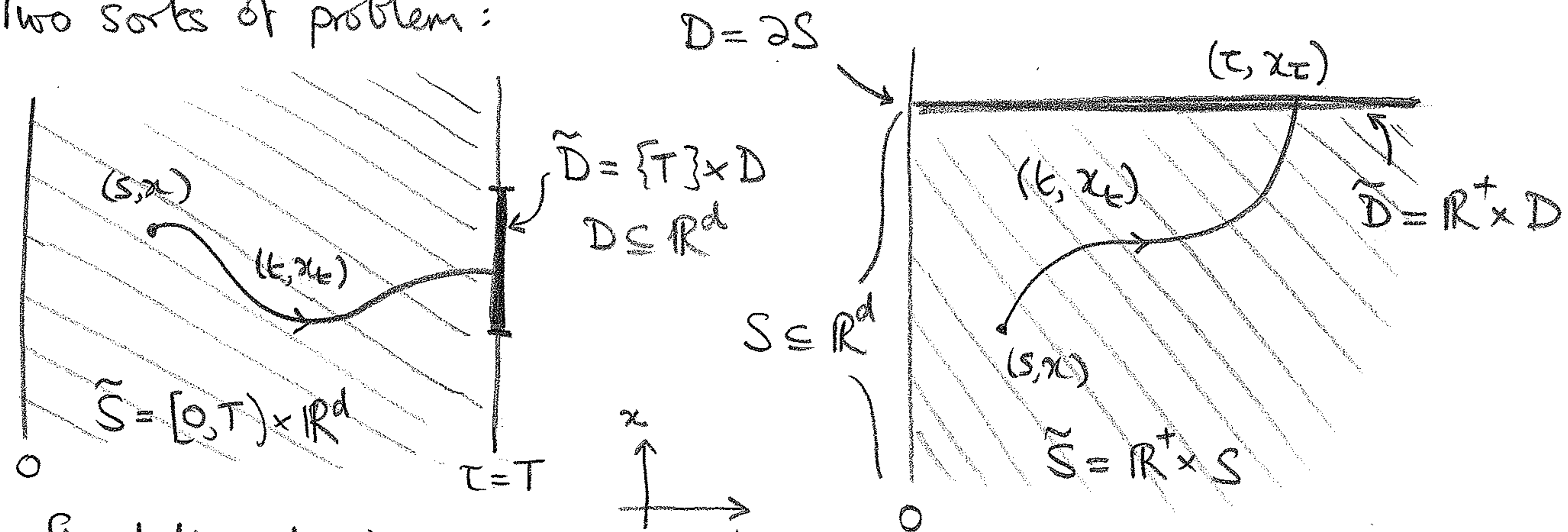
$$K = \sup_{t,x,a} \left| \frac{\partial b}{\partial x}(t,x,a) \right|.$$

So, given a continuous-time controllable dynamical system b, with state-space $\mathbb{R}^d$, for any starting point x_0, a choice of control u determines a unique <u>controlled path</u> $(x_t)_{t \geq 0}$ by the differential equation

$$\dot{x}_t = b(t, x_t, u_t), \quad t \geq 0$$

- We shall often solve the differential equations explicitly.
- The solution of differential equations is "local"; while $x_t \in S \subseteq \mathbb{R}^d$, only the values of b on S matter. So provided we can see that solutions do not leave S, then it will be enough to have b defined (and Lipschitz) on S.
- Continuity, of b on $\mathbb{R}^+ \times A$ and of u, is a convenient simple assumption, but can be weakened.

Two sorts of problem:



<u>fixed time horizon</u> <u>indefinite time horizon</u>

$$\dot{x}_t = b(t, x_t, u_t), \quad x_s = x$$ <u>controlled path</u>

u is <u>feasible</u> if $x_T \in D$ u is <u>feasible</u> if $\tau = \inf\{t \geq 0 : x_t \in D\} < \infty$

In both cases we seek to minimize over feasible controls

$$V^u(s,x) = \int_s^\tau c(t, x_t, u_t)\,dt + C(\tau, x_\tau)$$

where $c : \tilde{S} \times A \to \mathbb{R}$, $C : \tilde{D} \to \mathbb{R}$ are given continuous bounded cost functions

Say (s,x) is feasible if there is a feasible control from (s,x).

Define

$$V(s,x) = \begin{cases} \inf_u V^u(s,x), & \text{if } (s,x) \text{ is feasible,} \\ \infty, & \text{otherwise.} \end{cases}$$

For a differentiable function $F : \tilde{S} \to \mathbb{R}$, the HJB equation is

$$\inf_a \{ c(t,x,a) + \underbrace{\dot{V}(t,x)}_{\partial/\partial t} + \underbrace{\nabla V(t,x,a)}_{\partial/\partial x} b(t,x,a) \} = 0$$

Non-rigorous argument that V satisfies HJB

Starting from (t,x) choose action a until $t+\delta$, then switch to an optimal control. Compare this with an optimal control from (t,x):

$$V(t,x) \leqslant c(t,x,a)\delta + V(t+\delta, x+b(t,x,a)\delta)$$

with equality up to $O(\delta^2)$ if a is itself optimal. Now

$$V(t+\delta, x+b(t,x,a)\delta) = V(t,x) + \dot{V}(t,x)\delta + \nabla V(t,x) b(t,x,a)\delta + O(\delta^2).$$

Substitute in the $\leqslant$, rearrange and divide by δ

$$c(t,x,a) + \dot{V}(t,x) + \nabla V(t,x) b(t,x,a) + O(\delta) \geqslant 0$$

with equality up to $O(\delta)$ is a is optimal.

Hence V satisfies HJB.

Proposition 15.1

Suppose we can find a function $F : \tilde{S} \cup \tilde{D} \longrightarrow \mathbb{R}$, differentiable with continuous derivative, such that, for some $(s, x) \in \tilde{S}$ and some feasible control u^*, we have

$$c(t, y, a) + \dot{F}(t, y) + \nabla F(t, y) b(t, y, a) \geq 0, \quad (t, y) \in \tilde{S}, \ a \in A$$

with equality when $t \in [s, \tau^*)$ and $(x, a) = (x_t^*, u_t^*)$. Suppose also that $F = C$ on $\tilde{D}$. Then $F(s, x) = V(s, x)$ and u^* is optimal for (s, x).

Proof It will suffice to consider the case $S=0$.

Fix a feasible control u and set

$$M_t = \int_0^t c(s, x_s, u_s)\,ds + F(t, x_t), \quad 0 \le t \le \tau.$$

Then M is continuous on $[0,\tau]$ and differentiable on $[0,\tau)$, with

$$\dot{M}_t = c(t, x_t, u_t) + \dot{F}(t, x_t) + \nabla F(t, x_t)\, b(t, x_t, u_t) \ge 0$$

and with equality in the case $u = u^*$. Therefore

$$F(0,x) = M_0 \le M_\tau = \int_0^\tau c(t, x_t, u_t)\,dt + C(\tau, x_\tau) = V^u(0,x)$$

with equality if $u = u^*$.

$\square$

- Often, the function V and the control u^*, so obtain will not have the regularity (continuously differentiable and continuous, respectively) that we assumed in Proposition 15.1. There are more general versions, more complicated to consider which may apply. We tend to gloss over this gap.

- The main hope is often to guess the form of $V(t,x)$ as a function of x, then minimize over a explicitly, to reduce the HJB problem to a differential equation in t.

How to find an optimal control

- Solve HJB equation

$$\begin{cases} \inf_a \{ c(t,x,a) + \dot{V}(t,x) + \nabla V(t,x) b(t,x,a) \} = 0 & \text{on } \tilde{S} \\ V = C & \text{on } \tilde{D} \end{cases}$$

and identify for each $(t,x) \in \tilde{S}$ a minimizing action $u(t,x)$.

- Solve the differential equation

$$\dot{x}_t = b^u(t, x_t), \quad t \geq s, \quad x_s = x$$

where $b^u(t,x) = b(t,x,u(t,x))$ and check $\tau < \infty$, $x_\tau \in D$.

- The control $u^*_t = u(t, x_t)$ then has the minimizing property required in Proposition 15.1.

Example – continuous-time LQ system

$$b(x,a) = Ax + Ba, \quad x \in \mathbb{R}^d, \quad a \in \mathbb{R}^p$$

$$\dot{x}_t = b(x_t, u_t), \quad x_s = x$$

$$c(x,a) = x^T R x + a^T Q a, \quad R, Q \text{ positive-definite}$$

Seek to minimize

$$V^u(s,x) = \int_s^T c(x_t, u_t)\,dt + x_T^T \Pi(T) x_T$$

for a given non-negative definite matrix $\Pi(T)$.

Example – maximizing discounted utility

T = lifetime (given), $\sqrt{a}$ = utility from consuming a

α = personal discounting rate, β = interest rate

maximize $\int_0^T e^{-\alpha s} \sqrt{u_s}\, ds$

subject to $\dot{x}_t = \beta x_t - u_t$, $x_t \geq 0$, $0 \leq t \leq T$