Academic Report for DURF on Optimal Control Theory

Kehan Luo, Rongyao Li
Email: kl4747@nyu.edu, rl4785@nyu.edu

NYU Shanghai
Instructor: Prof. Vahagn Nersesyan
Email: vn2111@nyu.edu

May 2024 - September 2024

The Basic Problem: Controlled Dynamics

Consider an ordinary differential equation having the form

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) \quad (t > 0), \\ \mathbf{x}(0) = x^0, \end{cases}$$

We are here given the initial point $x^0 \in \mathbb{R}^n$ and the function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$. The unknown is the curve $\mathbf{x} : [0, \infty) \to \mathbb{R}^n$, which we interpret as the dynamical evolution of the **state** of some 'system'.

We generalize a bit and suppose now that $\mathbf{f}$ depends also upon some 'control' parameters belonging to a set $A \subset \mathbb{R}^m$, so that $\mathbf{f} : \mathbb{R}^n \times A \to \mathbb{R}^n$. Then if we select some value $a \in A$ and consider the corresponding dynamics:

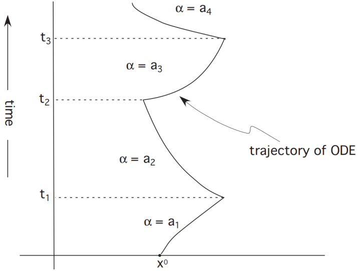
$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), a) \quad (t > 0), \\ \mathbf{x}(0) = x^0, \end{cases}$$

We obtain the evolution of our system when the parameter is constantly set to the value a.

The Basic Problem: Controlled Dynamics

For instance, we define the function $\boldsymbol{\alpha} : [0, \infty) \to A$ as follows:

$$\boldsymbol{\alpha}(t) = \begin{cases} a_1 & 0 \leq t \leq t_1 \\ a_2 & t_1 < t \leq t_2 \\ a_3 & t_2 < t \leq t_3 \quad \text{etc.} \end{cases}$$



More generally, we call a function $\boldsymbol{\alpha} : [0, \infty) \to A$ a **control**. Corresponding to each control, we consider the ODE

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\alpha}(t)) & (t > 0), \\ \mathbf{x}(0) = x^0, \end{cases}$$

and regard the trajectory $\mathbf{x}(\cdot)$ as the corresponding **response** of the system.

The Basic Problem: Payoffs and Optimization

Our overall task will be to determine what is the 'best' control for our system. For this, we need to specify a specific **payoff/reward** criterion. Let us define the payoff functional

$$\text{(P)} \qquad P[\boldsymbol{\alpha}(\cdot)] := \int_0^T r(\mathbf{x}(t), \boldsymbol{\alpha}(t))\, dt + g(\mathbf{x}(T)),$$

where $\mathbf{x}(\cdot)$ solves the (ODE) for the control $\boldsymbol{\alpha}(\cdot)$.

Here $r : \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are given, and we call r the **running payoff** and g the **terminal payoff**. The terminal time $T > 0$ is given as well.

Our aim is to find a control $\boldsymbol{\alpha}^*(\cdot)$, which maximizes the payoff. In other words, we want

$$P\left[\boldsymbol{\alpha}^*(\cdot)\right] \geq P[\boldsymbol{\alpha}(\cdot)]$$

for all controls $\boldsymbol{\alpha}(\cdot) \in \mathcal{A}$. Such a control $\boldsymbol{\alpha}^*(\cdot)$ is called **optimal**.

Controllability Question

Given the initial point x^0 and a 'target' set $S \subset \mathbb{R}^n$, does there exist a control steering the system to S in finite time?

Definitions.

We define the **reachable set** for time t to be

$\mathcal{C}(t) =$ set of initial points x^0 for which there exists a control such that $\mathbf{x}(t) = 0$,

and the **overall reachable set**

$\mathcal{C} =$ set of initial points x^0 for which there exists a control such that $\mathbf{x}(t) = 0$ for some finite time t. Note that $\mathcal{C} = \bigcup_{t \geq 0} \mathcal{C}(t)$.

We assume that our ODE is linear in both the state $\mathbf{x}(\cdot)$ and the control $\boldsymbol{\alpha}(\cdot)$, and consequently has the form

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(t) = M\mathbf{x}(t) + N\boldsymbol{\alpha}(t) & (t > 0), \\ \mathbf{x}(0) = x^0, \end{cases}$$

where $M \in \mathbb{M}^{n \times n}$ and $N \in \mathbb{M}^{n \times m}$. We assume the set A of control parameters is a cube in $\mathbb{R}^m$:

$$A = [-1, 1]^m = \{a \in \mathbb{R}^m \mid |a_i| \leq 1,\ i = 1, \ldots, m\}.$$

Controllability

Definitions.

(i) We say a set S is **symmetric** if $x \in S$ implies $-x \in S$.

(ii) The set S is **convex** if $x, \hat{x} \in S$ and $0 \leq \lambda \leq 1$ imply $\lambda x + (1-\lambda)\hat{x} \in S$.

(iii) The **controllability matrix** is

$$G = G(M, N) := \underbrace{\left[N, MN, M^2N, \ldots, M^{n-1}N\right]}_{\text{an } n \times (\text{mn}) \text{ matrix}}.$$

(iv) We write $\mathcal{C}^\circ$ for the interior of the set $\mathcal{C}$.

Theorem (Structure of Reachable Set):

(i) The reachable set $\mathcal{C}$ is symmetric and convex.

(ii) If $x^0 \in \mathcal{C}(\bar{t})$, then $x^0 \in \mathcal{C}(t)$ for all times $t \geq \bar{t}$.

Theorem (Controllability Matrix):

We have rank $G = n$ if and only if $0 \in \mathcal{C}^\circ$.

Criterion for Controllability

Theorem:

Let A be the cube $[-1, 1]^n$ in $\mathbb{R}^n$. Suppose as well that $\operatorname{rank} G = n$, and $\operatorname{Re} \lambda < 0$ for each eigenvalue λ of the matrix M. Then the system (ODE) is controllable.

Theorem (Improved Criterion):

Assume $\operatorname{rank} G = n$ and $\operatorname{Re} \lambda \leq 0$ for each eigenvalue λ of the matrix M, and assume further that every Jordan block of M corresponding to an eigenvalue λ with $\operatorname{Re} \lambda = 0$ is a 1×1 block. Then the system is controllable.

Criterion for Controllability (Continued)

Structure of Proof for Improved Criterion:

We proceed by contradiction.

1. Assumption and Setup: Assume $C \neq \mathbb{R}^n$. By the convexity of C, there exists a vector $b \neq 0$ and a real number μ such that

$$b \cdot x_0 \leq \mu \quad \text{for all } x_0 \in C.$$

2. Contradiction by Constructing $\mathbf{x_0}$: Given $b \neq 0$, $\mu \in \mathbb{R}$, our intention is to find $x_0 \in C$ so that the inequality $b \cdot x_0 \leq \mu$ fails. Recall $x_0 \in C$ if and only if there exist a time $t > 0$ and a control $\alpha(\cdot) \in A$ such that $x_0 = -\int_0^t X^{-1}(s) N\alpha(s)\, ds$. Thus, $b \cdot x_0 = -\int_0^t b^T X^{-1}(s) N\alpha(s)\, ds$. Define $v(s) := b^T X^{-1}(s) N$.

3. Non-zero Vector v(s): We assert that $v(s) \not\equiv 0$. Suppose instead that $v(s) \equiv 0$. Differentiating $b^T X^{-1}(s) N$ with respect to s and setting $s = 0$, we find $b^T M^k N = 0$for $k = 0, 1, 2, \ldots$ This implies b is orthogonal to the columns of G, so rank $G < n$, contradicting our hypothesis. Therefore, $v(s) \not\equiv 0$.

4. Control Construction: Define the control $\alpha(\cdot)$ as follows:
$\alpha(s) = \begin{cases} -\frac{v(s)}{|v(s)|} & \text{if } v(s) \neq 0, \\ 0 & \text{if } v(s) = 0. \end{cases}$ Then
$b \cdot x_0 = -\int_0^t v(s)\alpha(s)\, ds = \int_0^t |v(s)|\, ds.$
5. Integral of $|v(s)|$ Diverges: We want to find a time $t > 0$ such that $\int_0^t |v(s)|\, ds > \mu$. In fact, we assert that

$$\int_0^\infty |v(s)|\, ds = +\infty.$$

Define $\phi(t) := \int_t^\infty v(s)\, ds$. Since $p(M) = 0$ (Cayley-Hamilton Theorem), it follows that $\phi(t)$ satisfies an $(n+1)$th order ODE. Given $\phi(t) \not\equiv 0$, and considering the characteristic roots μ_i with $\operatorname{Re} \mu_i \geq 0$, we see $\phi(t)$ cannot decay to zero unless $\int_0^\infty |v(s)|\, ds = +\infty$.

6. Contradiction and Conclusion: Consequently, given any μ, there exists $t > 0$ such that $b \cdot x_0 = \int_0^t |v(s)|\, ds > \mu$, a contradiction to the initial assumption $b \cdot x_0 \leq \mu$. Therefore, $C = \mathbb{R}^n$, proving controllability.

Observability

We again consider the linear system of ODE

$$(\mathrm{ODE}) \qquad \begin{cases} \dot{\mathbf{x}}(t) = M\mathbf{x}(t) \\ \mathbf{x}(0) = x^0 \end{cases}$$

where $M \in \mathbb{M}^{n\times n}$. In this section we address the **observability** problem, modeled as follows. We suppose that we can observe

$$(\mathrm{O}) \qquad \mathbf{y}(t) := N\mathbf{x}(t) \quad (t \geq 0)$$

for a given matrix $N \in \mathbb{M}^{m\times n}$. Consequently, $\mathbf{y}(t) \in \mathbb{R}^m$. The interesting situation is when $m << n$ and we interpret $\mathbf{y}(\cdot)$ as low-dimensional "observations" or "measurements" of the high-dimensional dynamics $\mathbf{x}(\cdot)$.

Observability Question

Given the observations $\mathbf{y}(\cdot)$, can we in principle reconstruct $\mathbf{x}(\cdot)$? In particular, do observations of $\mathbf{y}(\cdot)$ provide enough information for us to deduce the initial value x^0 for (ODE)?

Definitions.

The pair $(\mathrm{ODE}), (\mathrm{O})$ is called observable if the knowledge of $\mathbf{y}(\cdot)$ on any time interval $[0, t]$ allows us to compute x^0.

More precisely, (ODE), (O) is **observable** if for all solutions $\mathbf{x}_1(\cdot), \mathbf{x}_2(\cdot), N\mathbf{x}_1(\cdot) \equiv N\mathbf{x}_2(\cdot)$ on a time interval $[0, t]$ implies $\mathbf{x}_1(0) = \mathbf{x}_2(0)$.

Theorem (Observability and Controllability):

The system $\begin{cases} \dot{\mathbf{x}}(t) = M\mathbf{x}(t) \\ \mathbf{y}(t) = N\mathbf{x}(t) \end{cases}$ is observable if and only if the system

$$\dot{\mathbf{z}}(t) = M^T\mathbf{z}(t) + N^T\alpha(t), \quad A = \mathbb{R}^m$$

is controllable, meaning that $\mathcal{C} = \mathbb{R}^n$.

Bang-Bang Principle

For this section, we will take A to be the cube $[-1,1]^m$ in $\mathbb{R}^m$.

Definition.

A control $\boldsymbol{\alpha}(\cdot) \in \mathcal{A}$ is called bang-bang if for each time $t \geq 0$ and each index $i = 1, \ldots, m$, we have $\left|\alpha^i(t)\right| = 1$, where

$$\boldsymbol{\alpha}(t) = \begin{pmatrix} \alpha^1(t) \\ \vdots \\ \alpha^m(t) \end{pmatrix}.$$

Theorem (Bang-Bang Principle):

Let $t > 0$ and suppose $x^0 \in \mathcal{C}(t)$, for the system

$$\dot{\mathbf{x}}(t) = M\mathbf{x}(t) + N\boldsymbol{\alpha}(t)$$

Then there exists a bang-bang control $\boldsymbol{\alpha}(\cdot)$ which steers x^0 to 0 at time t.

Pontryagin Maximum Principle (PMP)

Assume we are given a smooth function $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,
$L = L(x, v)$, L is called the **Lagrangian**.
Let $T > 0, x^0, x^1 \in \mathbb{R}^n$ be given.

Definitions.

(i) For the given curve $\mathbf{x}(\cdot)$, define

$$\mathbf{p}(t) := \nabla_v L(\mathbf{x}(t), \dot{\mathbf{x}}(t)) \quad (0 \leq t \leq T).$$

We call $\mathbf{p}(\cdot)$ the **generalized momentum**.

Hypothesis: Assume that for all $x, p \in \mathbb{R}^n$, we can solve the equation $p = \nabla_v L(x, v)$ for $v = \mathbf{v}(x, p)$.

(ii) Define the dynamical systems **Hamiltonian** $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by the formula

$$H(x, p) = p \cdot \mathbf{v}(x, p) - L(x, \mathbf{v}(x, p))$$

where $\mathbf{v}$ is defined above.

Pontryagin Maximum Principle (PMP): Fixed time, Free endpoint Problem

Now let us recall the basic problem. That is, **fixed time, free endpoint problem**.

We are given $A \subseteq \mathbb{R}^m$ and also $\mathbf{f} : \mathbb{R}^n \times A \to \mathbb{R}^n, x^0 \in \mathbb{R}^n$. We as before denote the set of admissible controls by

$$\mathcal{A} = \{\boldsymbol{\alpha}(\cdot) : [0, \infty) \to A \mid \boldsymbol{\alpha}(\cdot) \text{ is measurable } \}$$

Then given $\boldsymbol{\alpha}(\cdot) \in \mathcal{A}$, we solve for the corresponding evolution of our system:

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\alpha}(t)) \quad (t \geq 0) \\ \mathbf{x}(0) = x^0 \end{cases}$$

We also introduce the payoff functional

$$\text{(P)} \qquad P[\boldsymbol{\alpha}(\cdot)] = \int_0^T r(\mathbf{x}(t), \boldsymbol{\alpha}(t))dt + g(\mathbf{x}(T)),$$

where the terminal time $T > 0$, running payoff $r : \mathbb{R}^n \times A \to \mathbb{R}$ and terminal payoff $g : \mathbb{R}^n \to \mathbb{R}$ are given.

Pontryagin Maximum Principle (PMP): Fixed time, Free endpoint Problem

The aim is to find a control $\boldsymbol{\alpha}^*(\cdot)$ such that

$$P\left[\boldsymbol{\alpha}^*(\cdot)\right] = \max_{\boldsymbol{\alpha}(\cdot)\in\mathcal{A}} P[\boldsymbol{\alpha}(\cdot)]$$

The Pontryagin Maximum Principle (PMP) asserts the existence of a function $\mathbf{p}^*(\cdot)$, which together with the optimal trajectory $\mathbf{x}^*(\cdot)$ satisfies an analog of Hamilton's ODE.

For this, we will need an appropriate Hamiltonian defined below.

Definition.
The **control theory Hamiltonian** is the function

$$H(x, p, a) := \mathbf{f}(x, a) \cdot p + r(x, a) \quad (x, p \in \mathbb{R}^n, a \in A)$$

Pontryagin Maximum Principle (PMP)

Theorem (PMP for fixed time, free endpoint problem):

Assume $\boldsymbol{\alpha}^*(\cdot)$ is optimal for $(\mathrm{ODE}), (\mathrm{P})$ and $\mathbf{x}^*(\cdot)$ is the corresponding trajectory.

Then there exists a function $\mathbf{p}^* : [0, T] \to \mathbb{R}^n$ such that

$$\begin{aligned} &(\mathrm{ODE}) \qquad \dot{\mathbf{x}}^*(t) = \nabla_p H\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right) \\ &(\mathrm{ADJ}) \qquad \dot{\mathbf{p}}^*(t) = -\nabla_x H\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right), \end{aligned}$$

and

$$(\mathrm{M}) \qquad H\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right) = \max_{a \in A} H\left(\mathbf{x}^*(t), \mathbf{p}^*(t), a\right) \quad (0 \le t \le T).$$

In addition,

the mapping $t \mapsto H\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right)$ is constant.

Finally, we have the terminal condition

$$(\mathrm{T}) \qquad \mathbf{p}^*(T) = \nabla g\left(\mathbf{x}^*(T)\right)$$

Pontryagin Maximum Principle (PMP): Fixed time, Free endpoint Problem

Remarks and Interpretations:

(i) The identities (ADJ) are the **adjoint equations** and (M) the **maximization principle**. Notice that (ODE) and (ADJ) resemble the structure of Hamilton's equations.

We also call (T) the **transversality condition**.

(ii) More precisely, (ODE) says that for $1 \leq i \leq n$, we have

$$\dot{x}^{i*}(t) = H_{p_i}\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right) = f^i\left(\mathbf{x}^*(t), \boldsymbol{\alpha}^*(t)\right),$$

which is just the original equation of motion. Likewise, (ADJ) says

$$\begin{aligned}\dot{p}^{i*}(t) &= -H_{x_i}\left(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)\right) \\ &= -\sum_{j=1}^{n} p^{j*}(t) f^j_{x_i}\left(\mathbf{x}^*(t), \boldsymbol{\alpha}^*(t)\right) - r_{x_i}\left(\mathbf{x}^*(t), \boldsymbol{\alpha}^*(t)\right)\end{aligned}$$

$\square$

Pontryagin Maximum Principle (PMP): Free time, Fixed endpoint Problem

Now let us discuss the **free time, fixed endpoint problem**.

As before, given a control $\boldsymbol{\alpha}(\cdot) \in \mathcal{A}$, we solve for the corresponding evolution of our system:

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\alpha}(t)) \quad (t \geq 0) \\ \mathbf{x}(0) = x^0 \end{cases}$$

Assume now that a target point $x^1 \in \mathbb{R}^n$ is given. We introduce the payoff functional

$$\text{(P)} \qquad P[\boldsymbol{\alpha}(\cdot)] = \int_0^\tau r(\mathbf{x}(t), \boldsymbol{\alpha}(t))dt$$

Here $r : \mathbb{R}^n \times A \to \mathbb{R}$ is the given running payoff, and $\tau = \tau[\boldsymbol{\alpha}(\cdot)] \leq \infty$ denotes the first time the solution of (ODE) hits the target point x^1.

As before, the basic problem is to find an optimal control $\boldsymbol{\alpha}^*(\cdot)$ such that $P\left[\boldsymbol{\alpha}^*(\cdot)\right] = \max_{\boldsymbol{\alpha}(\cdot)\in\mathcal{A}} P[\boldsymbol{\alpha}(\cdot)]$

Define the Hamiltonian H as before.

Pontryagin Maximum Principle (PMP)

Theorem (PMP for free time, fixed endpoint problem):
Assume $\boldsymbol{\alpha}^*(\cdot)$ is optimal for $(\mathrm{ODE}), (\mathrm{P})$ and $\mathbf{x}^*(\cdot)$ is the corresponding trajectory.
Then there exists a function $\mathbf{p}^* : [0, \tau^*] \to \mathbb{R}^n$ such that

$$(\mathrm{ODE}) \qquad \dot{\mathbf{x}}^*(t) = \nabla_p H(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t))$$

$$(\mathrm{ADJ}) \qquad \dot{\mathbf{p}}^*(t) = -\nabla_x H(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t))$$

and

$$(\mathrm{M}) \quad H(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)) = \max_{a \in A} H(\mathbf{x}^*(t), \mathbf{p}^*(t), a) \quad (0 \le t \le \tau^*) \,.$$

Also,

$$H(\mathbf{x}^*(t), \mathbf{p}^*(t), \boldsymbol{\alpha}^*(t)) \equiv 0 \quad (0 \le t \le \tau^*)$$

Pontryagin Maximum Principle (PMP): Free time, Fixed endpoint Problem

Here τ^* denotes the first time the trajectory $x^*(\cdot)$ hits the target point x^1. We call $\mathbf{x}^*(\cdot)$ the **state** of the optimally controlled system and $\mathbf{p}^*(\cdot)$ the **costate**.

More precisely, we should define

$$H(x, p, q, a) = \mathbf{f}(x, a) \cdot p + r(x, a)q \quad (q \in \mathbb{R})$$

A more careful statement of the Maximum Principle says "there exists a constant $q \geq 0$ and a function $\mathbf{p}^* : [0, t^*] \to \mathbb{R}^n$ such that (ODE), (ADJ), and (M) hold".

If $q > 0$, we can renormalize to get $q = 1$, as we have done above. If $q = 0$, then H does not depend on running payoff r and in this case the Pontryagin Maximum Principle is not useful. This is a so-called "abnormal problem".

PMP for Linear Time-Optimal Control

Consider the linear system of ODE:

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(t) = M\mathbf{x}(t) + N\boldsymbol{\alpha}(t) \\ \mathbf{x}(0) = x^0, \end{cases}$$

for given matrices $M \in \mathbb{M}^{n\times n}$ and $N \in \mathbb{M}^{n\times m}$.

We will take A to be the cube $[-1,1]^m \subset \mathbb{R}^m$.

Define next $\quad \text{(P)} \qquad P[\boldsymbol{\alpha}(\cdot)] := -\int_0^\tau 1 ds = -\tau$

where $\tau = \tau(\boldsymbol{\alpha}(\cdot))$ denotes the first time the solution of our (ODE) hits the origin 0 . (If the trajectory never hits 0 , we set $\tau = \infty$.)

Optimal Time Problem:

We are given the starting point $x^0 \in \mathbb{R}^n$, and want to find an optimal control $\boldsymbol{\alpha}^*(\cdot)$ such that

$$P\left[\boldsymbol{\alpha}^*(\cdot)\right] = \max_{\boldsymbol{\alpha}(\cdot)\in\mathcal{A}} P[\boldsymbol{\alpha}(\cdot)]$$

Then

$\tau^* = -\mathcal{P}\left[\boldsymbol{\alpha}^*(\cdot)\right]$ is the minimum time to steer to the origin.

PMP for Linear Time-Optimal Control

Theorem (Existence of Time-Optimal Control):
Let $x^0 \in \mathbb{R}^n$. Then there exists an optimal bang-bang control $\boldsymbol{\alpha}^*(\cdot)$.

Theorem (PMP for Linear Time-Optimal Control):
There exists a nonzero vector h such that

$$(\mathrm{M}) \qquad h^T \mathbf{X}^{-1}(t) N \boldsymbol{\alpha}^*(t) = \max_{a \in A} \left\{ h^T \mathbf{X}^{-1}(t) N a \right\}$$

for each time $0 \leq t \leq \tau^*$.

Interpretation:
The significance of this assertion is that if we know h then the maximization principle (M) provides us with a formula for computing $\boldsymbol{\alpha}^*(\cdot)$, or at least extracting useful information.

The assertion (M) is a **special case** of the general Pontryagin Maximum Principle (PMP).

Dynamic Programming

We fix a terminal time $T > 0$ and then look at the controlled dynamics

$$(\mathrm{ODE}) \qquad \begin{cases} \dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), \boldsymbol{\alpha}(s)) & (0 < s < T) \\ \mathbf{x}(0) = x^0, \end{cases}$$

with the associated payoff functional

$$(\mathrm{P}) \qquad P[\boldsymbol{\alpha}(\cdot)] = \int_0^T r(\mathbf{x}(s), \boldsymbol{\alpha}(s))ds + g(\mathbf{x}(T)).$$

We embed this into a larger family of similar problems, by varying the starting times and starting points:

$$\begin{cases} \dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), \boldsymbol{\alpha}(s)) & (t \leq s \leq T) \\ \mathbf{x}(t) = x \end{cases}$$

with

$$P_{x,t}[\boldsymbol{\alpha}(\cdot)] = \int_t^T r(\mathbf{x}(s), \boldsymbol{\alpha}(s))ds + g(\mathbf{x}(T)).$$

Consider the above problems for all choices of starting times $0 \leq t \leq T$ and all initial points $x \in \mathbb{R}^n$.

Dynamic Programming

Definition.
For $x \in \mathbb{R}^n, 0 \leq t \leq T$, define the **value function** $v(x,t)$ to be the greatest payoff possible if we start at $x \in \mathbb{R}^n$ at time t.

In other words,

$$v(x,t) := \sup_{\boldsymbol{\alpha}(\cdot) \in \mathcal{A}} P_{x,t}[\boldsymbol{\alpha}(\cdot)] \quad (x \in \mathbb{R}^n, 0 \leq t \leq T).$$

Notice then that

$$v(x,T) = g(x) \quad (x \in \mathbb{R}^n).$$

Dynamic Programming: Derivation of Hamilton–Jacobi–Bellman Equation

Our first task is to show that **the value function v satisfies a certain nonlinear partial differential equation**.

Our derivation will be based upon the reasonable principle that "it's better to be smart from the beginning, than to be stupid for a time and then become smart". We want to convert this philosophy of life into mathematics.

To simplify, we hereafter suppose that the set A of control parameter values is compact.

Dynamic Programming

Theorem (Hamilton–Jacobi–Bellman Equation):

Assume that the value function v is a C^1 function of the variables (x, t). Then v solves the nonlinear partial differential equation

$$v_t(x,t)+\max_{a\in A}\{\mathbf{f}(x,a)\cdot\nabla_x v(x,t)+r(x,a)\}=0 \quad (x\in\mathbb{R}^n, 0\le t< T),$$

with the terminal condition

$$v(x,T)=g(x) \quad (x\in\mathbb{R}^n).$$

Remark:

We call the PDE above the Hamilton-Jacobi-Bellman equation, and can rewrite it as

$$\text{(HJB)} \qquad v_t(x,t)+H(x,\nabla_x v)=0 \quad (x\in\mathbb{R}^n, 0\le t< T)$$

for the partial differential equations Hamiltonian

$$H(x,p):=\max_{a\in A}H(x,p,a)=\max_{a\in A}\{\mathbf{f}(x,a)\cdot p+r(x,a)\}$$

where $x, p\in\mathbb{R}^n$.

Dynamic Programming Method

Step 1: Solve the Hamilton-Jacobi-Bellman equation, and thereby compute the value function v.

Step 2: Use the value function v and the Hamilton-Jacobi-Bellman PDE to design an optimal feedback control $\boldsymbol{\alpha}^*(\cdot)$, as follows. Define for each point $x \in \mathbb{R}^n$ and each time $0 \leq t \leq T$, $\boldsymbol{\alpha}(x,t) = a \in A$ to be a parameter value where the maximum in (HJB) is attained.

In other words, we select $\boldsymbol{\alpha}(x,t)$ so that

$$v_t(x,t) + \mathbf{f}(x, \boldsymbol{\alpha}(x,t)) \cdot \nabla_x v(x,t) + r(x, \boldsymbol{\alpha}(x,t)) = 0.$$

Next we solve the following ODE, assuming $\boldsymbol{\alpha}(\cdot, t)$ is sufficiently regular to let us do so:

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}^*(s) = \mathbf{f}\left(\mathbf{x}^*(s), \boldsymbol{\alpha}\left(\mathbf{x}^*(s), s\right)\right) & (t \leq s \leq T) \\ \mathbf{x}(t) = x \end{cases}$$

Finally, define the **feedback control**

$$\boldsymbol{\alpha}^*(s) := \boldsymbol{\alpha}\left(\mathbf{x}^*(s), s\right)$$

Dynamic Programming Method

In summary, we design the optimal control this way:

If the state of system is x at time t, use the control which at time t takes on the parameter value $a \in A$ such that the minimum in (HJB) is obtained.

We demonstrate next that this construction does indeed provide us with an optimal control.

Theorem (Verification of Optimality):
The control $\boldsymbol{\alpha}^*(\cdot)$ defined by the construction

$$\boldsymbol{\alpha}^*(s) := \boldsymbol{\alpha}\left(\mathbf{x}^*(s), s\right)$$

is optimal.

Proof:
We have

$$P_{x,t}\left[\boldsymbol{\alpha}^*(\cdot)\right] = \int_t^T r\left(\mathbf{x}^*(s), \boldsymbol{\alpha}^*(s)\right) ds + g\left(\mathbf{x}^*(T)\right).$$

Dynamic Programming Method

Proof (Continued):

Furthermore according to the definition of $\boldsymbol{\alpha}(\cdot)$:

$$
\begin{aligned}
& P_{x,t}\left[\boldsymbol{\alpha}^*(\cdot)\right] \\
& = \int_t^T \left(-v_t\left(\mathbf{x}^*(s), s\right) - \mathbf{f}\left(\mathbf{x}^*(s), \boldsymbol{\alpha}^*(s)\right) \cdot \nabla_x v\left(\mathbf{x}^*(s), s\right)\right) ds + g\left(\mathbf{x}^*(T)\right) \\
& = -\int_t^T v_t\left(\mathbf{x}^*(s), s\right) + \nabla_x v\left(\mathbf{x}^*(s), s\right) \cdot \dot{\mathbf{x}}^*(s) ds + g\left(\mathbf{x}^*(T)\right) \\
& = -\int_t^T \frac{d}{ds} v\left(\mathbf{x}^*(s), s\right) ds + g\left(\mathbf{x}^*(T)\right) \\
& = -v\left(\mathbf{x}^*(T), T\right) + v\left(\mathbf{x}^*(t), t\right) + g\left(\mathbf{x}^*(T)\right) \\
& = -g\left(\mathbf{x}^*(T)\right) + v\left(\mathbf{x}^*(t), t\right) + g\left(\mathbf{x}^*(T)\right) = v(x, t) = \sup_{\boldsymbol{\alpha}(\cdot) \in \mathcal{A}} P_{x,t}[\boldsymbol{\alpha}(\cdot)]
\end{aligned}
$$

That is,

$$
P_{x,t}\left[\boldsymbol{\alpha}^*(\cdot)\right] = \sup_{\boldsymbol{\alpha}(\cdot) \in \mathcal{A}} P_{x,t}[\boldsymbol{\alpha}(\cdot)]
$$

and so $\boldsymbol{\alpha}^*(\cdot)$ is optimal, as asserted.

Connections between Dynamic Programming and PMP

Return now to our usual control theory problem, with dynamics

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), \boldsymbol{\alpha}(s)) \quad (t \leq s \leq T) \\ \mathbf{x}(t) = x \end{cases}$$

and payoff

$$\text{(P)} \qquad P_{x,t}[\boldsymbol{\alpha}(\cdot)] = \int_t^T r(\mathbf{x}(s), \boldsymbol{\alpha}(s))ds + g(\mathbf{x}(T)).$$

As above, the value function is

$$v(x,t) = \sup_{\boldsymbol{\alpha}(\cdot)} P_{x,t}[\boldsymbol{\alpha}(\cdot)].$$

Connections between Dynamic Programming and PMP

The next theorem demonstrates that **the costate in the Pontryagin Maximum Principle is in fact the gradient in x of the value function v, taken along an optimal trajectory**.

Theorem (Costates and Gradients):
Assume $\alpha^*(\cdot), \mathbf{x}^*(\cdot)$ solve the control problem (ODE), (P).

If the value function v is C^2, then the costate $\mathbf{p}^*(\cdot)$ occuring in the Maximum Principle is given by

$$\mathbf{p}^*(s) = \nabla_x v\left(\mathbf{x}^*(s), s\right) \quad (t \leq s \leq T)$$

Structure of Proof:
1. Definition and Objective: Define $p(t) := \nabla_x v(x(t), t)$. The goal is to prove that $p(t)$ satisfies the conditions of the Pontryagin Maximum Principle, specifically the adjoint equation (ADJ) and the maximization condition (M).

Connections between Dynamic Programming and PMP

Structure of Proof (Continued):

2. Calculation of the Derivative: Compute the time derivative of $p_i(t)$:

$$\dot{p}_i(t) = \frac{d}{dt} v_{x_i}(x(t), t) = v_{x_i t}(x(t), t) + \sum_{j=1}^{n} v_{x_i x_j}(x(t), t)\dot{x}_j(t).$$

Substitute the Hamilton-Jacobi-Bellman (HJB) equation for v to simplify the expression.

3. Establishing the Adjoint Equation (ADJ): Use the fact that v solves the HJB equation to find that $\dot{p}(t) = -(\nabla_x f)p - \nabla_x r$. This matches the form of the adjoint equation, confirming (ADJ).

4. Verification of the Maximization Condition (M): Finally, verify that the Hamiltonian $H(x(t), p(t), a)$ is maximized at the control $a = \alpha(t)$, satisfying condition (M) of the Maximum Principle.

Connections between Dynamic Programming and PMP

Interpretations: The foregoing provides us with another way to look at transversality conditions:

(i) **Free endpoint problem**: Recall that for the free endpoint problem we have the condition

$$\mathbf{p}^*(T) = \nabla g\left(\mathbf{x}^*(T)\right)$$

for the payoff functional

$$\int_t^T r(\mathbf{x}(s), \alpha(s))ds + g(\mathbf{x}(T))$$

To understand this better, note $\mathbf{p}^*(s) = -\nabla v\left(\mathbf{x}^*(s), s\right)$. But $v(x, t) = g(x)$, and hence the foregoing implies

$$\mathbf{p}^*(T) = \nabla_x v\left(\mathbf{x}^*(T), T\right) = \nabla g\left(\mathbf{x}^*(T)\right).$$

Connections between Dynamic Programming and PMP

(ii) **Constrained initial and target sets**: Recall that for this problem we stated the transversality conditions that

$$\begin{cases} \mathbf{p}^*(0) \text{ is perpendicular to } T_0 \\ \mathbf{p}^*(\tau^*) \text{ is perpendicular to } T_1 \end{cases}$$

when τ^* denotes the first time the optimal trajectory hits the target set X_1. Now let v be the value function for this problem:

$$v(x) = \sup_{\boldsymbol{\alpha}(\cdot)} P_x[\boldsymbol{\alpha}(\cdot)]$$

with the constraint that we start at $x^0 \in X_0$ and end at $x^1 \in X_1$.

But then v will be constant on the set X_0 and also constant on X_1. Since ∇v is perpendicular to any level surface, ∇v is therefore perpendicular to both ∂X_0 and ∂X_1. And since

$$\mathbf{p}^*(t) = \nabla v\left(\mathbf{x}^*(t)\right),$$

this means that $\begin{cases} \mathbf{p}^* \text{ is perpendicular to } \partial X_0 \text{ at } t = 0 \\ \mathbf{p}^* \text{ is perpendicular to } \partial X_1 \text{ at } t = \tau^* \end{cases}$

Application 1: Classic Brachistochrone Problem

Given two points A and B, as shown, the graph of a function $y(\cdot)$ joining these points can be interpreted as a wire path along which a ball of unit mass slides without friction under the influence of gravity. The goal is to design the slide to minimize the time it takes for the ball to slide from A to B.

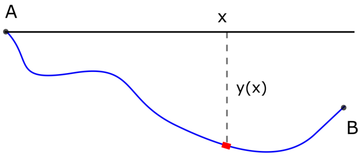


Figure 1: Illustration of the Brachistochrone problem.

Assumptions

1. For simplicity, we assume $A = (0, 0)$, $B = (x_b, y_b)$ where $y_b \leq 0$, and $y(x) \leq 0$ for all $0 \leq x \leq x_b$.
2. No other forces except gravity are acting on the ball, meaning the ball's total energy (kinetic energy + potential energy) is constant.
3. The slide we are designing must be continuous.

Solution

By assumption 2, we have

$$\frac{v^2}{2} + gy = 0 \tag{1}$$

on the interval $[0, b]$, where v is the velocity and g is the gravitational acceleration. The constant is 0, since $v(0) = y(0) = 0$.
Thus, we obtain

$$v = \sqrt{-2gy} \tag{2}$$

The time for the bead to slide from A to B is then

$$\int_0^b \frac{ds}{v} = \int_0^b \left(\frac{\left(1 + (y')^2\right)}{-2gy} \right)^{\frac{1}{2}} dx \tag{3}$$

The Lagrangian Approach

We seek a path $y_0(\cdot)$ from A to B that minimizes

$$I[y(\cdot)] = \int_0^b \left(\frac{\left(1 + (y')^2\right)}{-y} \right)^{\frac{1}{2}} dx \tag{4}$$

Now, we have the Lagrangian:

$$L = \left(\frac{1+z^2}{-y} \right)^{\frac{1}{2}}, \quad \frac{\partial L}{\partial z} = -\left(\frac{1+z^2}{-y} \right)^{-\frac{1}{2}} \frac{z}{y} \tag{5}$$

Euler-Lagrange Equation

Consequently,

$$
\begin{aligned}
y' \frac{\partial L}{\partial z}\left(y, y'\right) - L\left(y, y'\right) &= -\left(\frac{1+\left(y'\right)^2}{-y}\right)^{-\frac{1}{2}} \frac{\left(y'\right)^2}{y} \\
&\quad - \left(\frac{1+\left(y'\right)^2}{-y}\right)^{\frac{1}{2}} \\
&= \left(-y\left(1+\left(y'\right)^2\right)\right)^{-\frac{1}{2}}
\end{aligned} \tag{6}
$$

Since L does not depend on x, it follows from Theorem [ ] that

$$
y' \frac{\partial L}{\partial z}\left(y, y'\right) - L\left(y, y'\right) \tag{7}
$$

is constant. Therefore,

$$
y\left(1+\left(y'\right)^2\right) = C \tag{8}
$$

on the interval $[0, b]$ for some (negative) constant C.

Solving the ODE

We can directly integrate the ODE (8) by letting $C = -\frac{1}{\alpha^2}$, to obtain a parametric solution:

$$
\begin{aligned}
x &= R(\phi - \sin\phi) \\
y &= -R(1 - \cos\phi)
\end{aligned}
\tag{9}
$$

where $R = \frac{1}{2\alpha^2}$. This forms a field of extremals, see Figure 2.

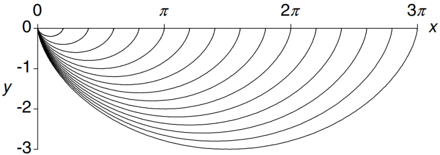


Figure 2: One-parameter family of cycloids.

Each extremal extends along the x-axis π times its maximum depth. There is one (and only one) extremal that passes through the boundary points $A = (0, 0)$ and $B = (x_b, y_b)$.

Remark

These extremals are cycloids, the curve traced by a point on the rim of a circle as it rolls horizontally. To see this, we define ξ as the angle between the tangent and vertical lines for any point on the extremal $y(\cdot)$.

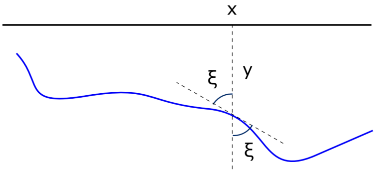


Figure 3: Angles and derivatives.

The angle ξ satisfies:

$$\sin \xi = \frac{1}{\left(1 + (y')^2\right)^{\frac{1}{2}}} \tag{10}$$

Combining with the ODE (8), we have:

$$\frac{\sin \xi}{(-y)^{\frac{1}{2}}} = d \tag{11}$$

for some constant d. We then verify that (11) is a property of cycloids. The key observation is that if a point $C = (x, y)$ on a rolling circle of diameter $d > 0$ generates a cycloid, and if A is the instantaneous point of contact of the circle with the line, then the vector AC is perpendicular to the velocity vector $\mathbf{v}$ of the rolling point C.

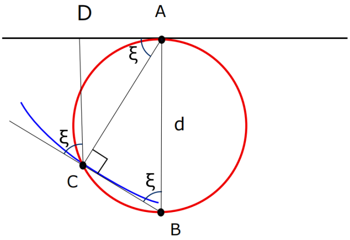


Figure 4: Geometry of cycloid

Thus **v**, which is tangent to the blue cycloid curve, is parallel (or on the same line) to CB, with B denoting the point directly opposite A on the circle. As $|AB| = d$, elementary geometry shows for the angles ξ that $|AC| = d \sin \xi$ and

$$-y = |DC| = |AC| \sin \xi = d \sin^2 \xi \tag{12}$$

This verifies (11).

$\square$

The cycloid is a tautochrone, meaning that if we release two balls from rest at different locations along the cycloid wire curve, they arrive at the lowest point at the same time. A proof of this can be found in [L, page 55].

Application 2: Moon Lander

This model asks us to bring a spacecraft to a soft landing on the lunar surface, using the least amount of fuel. In other words, we need to design a fuel usage strategy to obtain 0 velocity at the landing while minimizing the total amount of fuel we use.

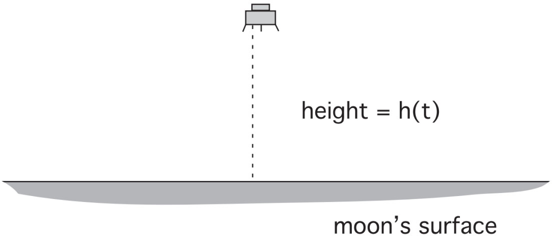


Figure 5: A spacecraft landing on the moon

Notations and Assumptions

$$\begin{aligned} h(t) &= \text{ height at time } t \\ v(t) &= \text{ velocity } = \dot{h}(t) \\ m(t) &= \text{ mass of spacecraft (changing as fuel is used up)} \\ \alpha(t) &= \text{ thrust at time } t \end{aligned}$$

1. For simplicity, we assume the thrust is constrained so that $0 \leq \alpha(t) \leq 1$, in other words, $A = [0, 1]$.
2. We naturally have the constraints that the height and mass be nonnegative: $h(t) \geq 0, m(t) \geq 0$.
3. We assume the initial conditions h_0 and m_0 are positive.
4. We assume there is no other forces than gravitational force that is affecting the spacecraft.

Problem Statement: Dynamics

Therefore, the dynamics are:

$$\begin{cases} \dot{h}(t) = v(t) \\ \dot{v}(t) = -g + \dfrac{\alpha(t)}{m(t)} \\ \dot{m}(t) = -k\alpha(t) \end{cases} \tag{13}$$

with the initial conditions:

$$\begin{cases} h(0) = h_0 > 0 \\ v(0) = v_0 \\ m(0) = m_0 > 0 \end{cases} \tag{14}$$

Problem Statement: Payoff Function

The goal is to maximize the remaining fuel $m(\tau)$, where $\tau = \tau[\alpha(\cdot)]$ is the first time $h(\tau) = v(\tau) = 0$. Since $\alpha = -\frac{\dot{m}}{k}$, our intention is equivalently to minimize the total applied thrust before landing; so as to maximize

$$P[\alpha(\cdot)] = -\int_0^\tau \alpha(t)dt$$

This is because

$$\int_0^\tau \alpha(t)dt = \frac{m_0 - m(\tau)}{k}$$

Solution: Introducing the maximum principle

In terms of the general notation, we have

$$\mathbf{x}(t) = \left(\begin{array}{c} h(t) \\ v(t) \\ m(t) \end{array} \right), \mathbf{f} = \left(\begin{array}{c} v \\ -g + a/m \\ -ka \end{array} \right)$$

Hence the Hamiltonian is

$$\begin{aligned} H(x, p, a) &= \mathbf{f} \cdot p + r \\ &= (v, -g + a/m, -ka) \cdot (p_1, p_2, p_3) - a \\ &= -a + p_1 v + p_2 \left(-g + \frac{a}{m} \right) + p_3(-ka) \end{aligned}$$

and the adjoint dynamics (ADJ) are:

$$\begin{cases} \dot{p}^1(t) = 0 \\ \dot{p}^2(t) = -p^1(t) \\ \dot{p}^3(t) = \frac{p^2(t)\alpha(t)}{m(t)^2} \end{cases} \tag{15}$$

Maximization Condition and Optimal Control

The maximization condition (M) reads:

$$
\begin{aligned}
H(\mathbf{x}(t), \mathbf{p}(t), \alpha(t)) &= \max_{0 \le a \le 1} H(\mathbf{x}(t), \mathbf{p}(t), a) \\
&= \max_{0 \le a \le 1} \left\{ -a + p^1(t)v(t) + p^2(t)\left[-g + \frac{a}{m(t)}\right] + p^3(t)(-ka) \right\} \\
&= p^1(t)v(t) - p^2(t)g + \max_{0 \le a \le 1} \left\{ a \left(-1 + \frac{p^2(t)}{m(t)} - kp^3(t) \right) \right\}
\end{aligned}
\tag{16}
$$

Thus the optimal control law is given by the rule:

$$
\alpha(t) = \begin{cases} 1 & \text{if} \quad 1 - \frac{p^2(t)}{m(t)} + kp^3(t) < 0 \\ 0 & \text{if} \quad 1 - \frac{p^2(t)}{m(t)} + kp^3(t) > 0 \end{cases} \tag{17}
$$

Constructing the Feedback Control

By Pontryagin's Maximum Principle, we derive the value of the feedback control as (17). However, we do not know how many switches between 0 and 1 do we need to design an optimal control, as well as at what time should we switch. Let us start by guessing that we first leave the rocket engine off (i.e., set $\alpha \equiv 0$) and turn the engine on only at the end. Denote by τ the first time that $h(\tau) = v(\tau) = 0$, meaning that we have landed. We guess that there exists a switching time $t^* < \tau$ when we turn the engines on at full power (i.e., set $\alpha \equiv 1$). Consequently:

$$\alpha(t) = \begin{cases} 0 & \text{for} \quad 0 \leq t \leq t^* \\ 1 & \text{for} \quad t^* \leq t \leq \tau \end{cases} \tag{18}$$

Justification of the Guess

To justify our guess about the structure of the optimal control, let us now find the costate $\mathbf{p}(\cdot)$ so that $\alpha(\cdot)$ and $\mathbf{x}(\cdot)$ described above satisfy (ODE 13), (ADJ 15), (M 16). To do this, we will have have to figure out appropriate initial conditions:

$$p^1(0) = \lambda_1, p^2(0) = \lambda_2, p^3(0) = \lambda_3$$

We solve (ADJ 15) for $\alpha(\cdot)$ as above, and get:

$$\begin{cases} p^1(t) \equiv \lambda_1 & (0 \leq t \leq \tau) \\ p^2(t) = \lambda_2 - \lambda_1 t & (0 \leq t \leq \tau) \\ p^3(t) = \begin{cases} \lambda_3 & (0 \leq t \leq t^*) \\ \lambda_3 + \int_{t^*}^{t} \frac{\lambda_2 - \lambda_1 s}{(m_0 + k(t^* - s))^2}\, ds & (t^* \leq t \leq \tau) \end{cases} \end{cases} \tag{19}$$

Define

$$r(t) := 1 - \frac{p^2(t)}{m(t)} + p^3(t)k$$

then

$$\dot{r} = -\frac{\dot{p}^2}{m} + \frac{p^2 \dot{m}}{m^2} + \dot{p}^3 k = \frac{\lambda_1}{m} + \frac{p^2}{m^2}(-k\alpha) + \left(\frac{p^2\alpha}{m^2}\right) k = \frac{\lambda_1}{m(t)} \quad (20)$$

As we are only to verify the existence of $\mathbf{p}$, we are free to choose $\lambda_1 < 0$, such that r is decreasing.
We calculate

$$r(t^*) = 1 - \frac{(\lambda_2 - \lambda_1 t^*)}{m_0} + \lambda_3 k \quad (21)$$

and then adjust λ_2, λ_3 so that $r(t^*) = 0$. Then r is nonincreasing, $r(t^*) = 0$, and consequently $r > 0$ on $[0, t^*)$, $r < 0$ on $(t^*, \tau]$. But $(M\,16)$ says

$$\alpha(t) = \begin{cases} 1 & \text{if } r(t) < 0 \\ 0 & \text{if } r(t) > 0 \end{cases}$$

Thus an optimal control changes just once from 0 to 1 ; and so our guess of $\alpha(\cdot)$ does indeed satisfy the Pontryagin Maximum Principle. $\square$

We now attempt to figure out the form of the solution and check if it accords with the Maximum Principle.

Form of Solutions

For times $t^* \leq t \leq \tau$, our ODE becomes:

$$\begin{cases} \dot{h}(t) = v(t) \\ \dot{v}(t) = -g + \frac{1}{m(t)} \quad (\text{for } t^* \leq t \leq \tau) \\ \dot{m}(t) = -k \end{cases} \tag{22}$$

with $h(\tau) = 0$, $v(\tau) = 0$, and $m(t^*) = m_0$. We solve these dynamics:

$$\begin{cases} m(t) = m_0 + k(t^* - t) \\ v(t) = g(\tau - t) + \frac{1}{k} \log\left[\frac{m_0 + k(t^* - \tau)}{m_0 + k(t^* - t)}\right] \\ h(t) = \text{complicated formula} \end{cases} \tag{23}$$

Now, setting $t = t^*$:

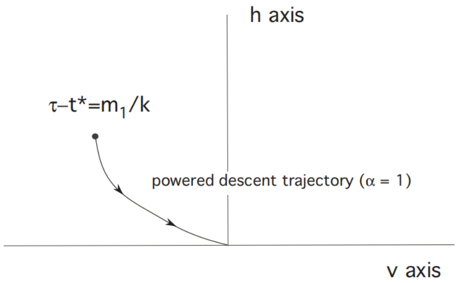
$$\begin{cases} m(t^*) = m_0 \\ v(t^*) = g(\tau - t^*) + \frac{1}{k} \log\left[\frac{m_0 + k(t^* - \tau)}{m_0}\right] \\ h(t^*) = -\frac{g(t^* - \tau)^2}{2} - \frac{m_0}{k^2} \log\left[\frac{m_0 + k(t^* - \tau)}{m_0}\right] + \frac{t^* - \tau}{k} \end{cases} \tag{24}$$

Suppose the total amount of fuel to start with was m_1; so that $m_0 - m_1$ is the weight of the empty spacecraft.

I. After time t^*, as $\alpha \equiv 1$, the fuel is used up at a rate of k. Hence:

$$k\,(\tau - t^*) \leq m_1 \tag{25}$$

and so $0 \leq \tau - t^* \leq \frac{m_1}{k}$. This says that assuming we know the time the spacecraft lands, the time remaining for the spacecraft to turn on its engine for landing shall not exceed $\frac{m_1}{k}$, which is a number we know. Thus, the powered descent trajectory as shown below, would only be partial because of this constraint on fuel:



As long as the status of spacecraft can be at this curve at some time, it is able to proceed a safe landing.

II. Before time t^*, we have $\alpha \equiv 0$. Then the ODE reads:

$$\begin{cases} \dot{h} = v \\ \dot{v} = -g \\ \dot{m} = 0 \end{cases} \tag{26}$$

Thus, we have:

$$\begin{cases} m(t) \equiv m_0 \\ v(t) = -gt + v_0 \\ h(t) = -\frac{1}{2}gt^2 + tv_0 + h_0 \end{cases} \tag{27}$$

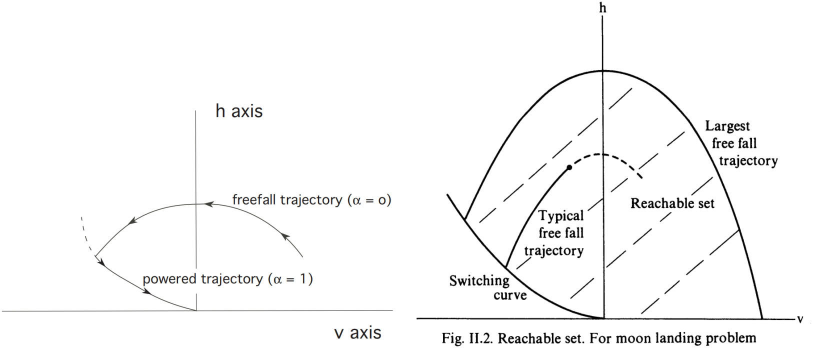
We combine the formulas for $v(t)$ and $h(t)$ to discover:

$$h(t) = h_0 - \frac{1}{2g}\left(v^2(t) - v_0^2\right) \quad (0 \le t \le t^*) \tag{28}$$

We deduce that the freefall trajectory $(v(t), h(t))$ lies on a parabola:

$$h = h_0 - \frac{1}{2g}\left(v^2 - v_0^2\right) \tag{29}$$

If we then move along this parabola (29) until we hit the soft-landing curve from the previous picture, we can then turn on the rocket engine and land safely, as shown in figure (7a).



(a) One successful landing case (b) Reachable set for moon landing

Figure 7: Successful Landing

However, we may miss switching curve, and hence cannot land safely on the moon by switching once, as shown in figure (8). In this case, we may try several switches.

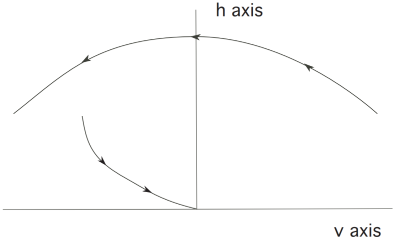


Figure 8: Failed Landing

Remark

The way we approach this model is theoretical. We assume we know the landing time τ of the spacecraft so as to form the powered descent trajectory as well as its relation with the freefall trajectory. In real life we do know this theoretically assumed time τ, but we can capture the height $h(\cdot)$ and velocity $v(\cdot)$ in seconds instead. Therefore, at each time we capture the status $(h(\cdot), v(\cdot))$ of the spacecraft as we are freefalling, we can go back to (24) to check whether there exists a common τ that satisfies the dynamics. If so, or approximately so, we have our current time t as t^*, and the calculated τ, and we need to check whether their difference is less or equal to $\frac{m_1}{k}$. If so, we know we are at the right point to switch and we can land safely without switching anymore. Otherwise, we have to turn to multiple switches or other strategies.

Application 3: Fuller's Problem, Chattering Controls

We consider the following dynamics:

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \alpha, \quad -1 \leq \alpha \leq 1 \tag{30}$$

with $\mathbf{x}(0) = \mathbf{x}^0 = \begin{bmatrix} y^0 \\ v^0 \end{bmatrix}$, where y^0 is the initial position and v^0 is the initial velocity.

Our payoff functional is:

$$P_x[\alpha(\cdot)] = -\frac{1}{2} \int_0^T (x_1)^2 \, dt \tag{31}$$

The value function is:

$$v(x) = \sup_{\alpha(\cdot) \in \mathcal{A}} P_x[\alpha(\cdot)]$$

The corresponding Hamilton-Jacobi-Bellman (HJB) equation is:

$$\max_{a \in A} \{\mathbf{f} \cdot \nabla v + r\} = 0$$

where

$$A = [-1, 1], \quad \mathbf{f} = \begin{bmatrix} x_2 \\ a \end{bmatrix}, \quad r = -\frac{1}{2}x_1^2$$

Therefore,

$$\max_{|a| \leq 1} \left\{ x_2 \frac{\partial v}{\partial x_1} + a \frac{\partial v}{\partial x_2} - \frac{1}{2}x_1^2 \right\} = 0$$

and consequently, the Hamilton-Jacobi-Bellman equation is:

$$\begin{cases} x_2 \frac{\partial v}{\partial x_1} + \left| \frac{\partial v}{\partial x_2} \right| = \frac{1}{2}x_1^2 \quad \text{in } \mathbb{R}^2 \setminus \{0\} \\ v(0) = 0 \end{cases} \tag{32}$$

Solving the HJB equation

Using the definition of the value function, we can check that it satisfies the two symmetry conditions:

$$v(-x) = v(x) \quad \left(x \in \mathbb{R}^2\right) \tag{I}$$

and

$$v\left(\lambda^2 x_1, \lambda x_2\right) = \lambda^5 v\left(x_1, x_2\right) \quad \left(x \in \mathbb{R}^2, \lambda > 0\right) \tag{II}$$

To verify (II), suppose that $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is an optimal trajectory starting at the point $x^0 = \begin{bmatrix} x_1^0 \\ x_2^0 \end{bmatrix}$, corresponding to an optimal control α.

Then we start instead at the perturbed point $x^0_\lambda = \begin{bmatrix} \lambda^2 x^0_1 \\ \lambda x^0_2 \end{bmatrix}$ for some $\lambda > 0$, by (30), optimal trajectories and control are

$$\mathbf{x}_\lambda(t) = \begin{bmatrix} \lambda^2 x_1\left(\frac{t}{\lambda}\right) \\ \lambda x_2\left(\frac{t}{\lambda}\right) \end{bmatrix}, \quad \alpha_\lambda(t) = \alpha\left(\frac{t}{\lambda}\right)$$

Consequently,

$$P\left[\alpha_\lambda(\cdot)\right] = -\frac{1}{2}\int_0^{T_\lambda} \left(x^1_\lambda\right)^2 dt = -\frac{\lambda^4}{2}\int_0^{T_\lambda} (x_1)^2\left(\frac{t}{\lambda}\right) dt = \lambda^5 P[\alpha(\cdot)]$$

The scaling identity (II) holds.

$\square$

We assume that the optimal switching should occur on the boundary Γ between two regions of the form

$$\begin{aligned} I &:= \{(x_1, x_2)\, |x_1 > -\beta x_2 |\, x_2\, |\} \\ II &:= \{(x_1, x_2)\, |x_1 < -\beta x_2 |\, x_2\, |\} \end{aligned}$$

for some as yet unknown constant $\beta > 0$.

Motivated by the structure of (32), we look a function v for which the scaling symmetry (II) holds, satisfying

$$\frac{\partial v}{\partial x_2} < 0 \text{ in Region I}, \quad \frac{\partial v}{\partial x_2} = 0 \text{ on } \Gamma, \quad \frac{\partial v}{\partial x_2} > 0 \text{ in Region II}$$

and in particular, v solves the linear PDE

$$x_2 \frac{\partial v}{\partial x_1} - \frac{\partial v}{\partial x_2} = \frac{1}{2}(x_1)^2 \quad \text{in Region I.} \tag{33}$$

In view (II), let us start by looking for a particular solution of (33) having the polynomial form

$$v = Ax_2^5 + Bx_1x_2^3 + Cx_1^2x_2.$$

If we plug this guess into (33) and match coefficients, we discover that

$$v = -\frac{1}{15}x_2^5 - \frac{1}{3}x_1x_2^3 - \frac{1}{2}x_1^2x_2$$

is a solution.

Now the general solution of the linear, homogeneous PDE

$$x_2 \frac{\partial w}{\partial x_1} - \frac{\partial w}{\partial x_2} = 0$$

has the form

$$w = f\left(x_1 + \frac{1}{2}x_2^2\right)$$

Hence the general solution of (33) is

$$v = -\frac{1}{15}x_2^5 - \frac{1}{3}x_1 x_2^3 - \frac{1}{2}x_1^2 x_2 + f\left(x_1 + \frac{1}{2}x_2^2\right)$$

In order to satisfy the scaling condition (II), we take f to be homogeneous and so have

$$v = -\frac{1}{15}x_2^5 - \frac{1}{3}x_1 x_2^3 - \frac{1}{2}x_1^2 x_2 - \gamma\left(x_1 + \frac{1}{2}x_2^2\right)^{\frac{5}{2}}$$

for another as yet unknown constant $\gamma > 0$.

We want next to adjust the constants β, γ so that $\frac{\partial v}{\partial x_2} = 0$ on Γ. Now

$$\frac{\partial v}{\partial x_2} = -\frac{1}{3}x_2^4 - x_1 x_2^2 - \frac{1}{2}x_1^2 - \frac{5\gamma}{2}\left(x_1 + \frac{1}{2}x_2^2\right)^{\frac{3}{2}} x_2$$

Therefore on $\Gamma_+ := \left\{x_1 = -\beta x_2^2, x_2 > 0\right\}$, we have

$$\frac{\partial v}{\partial x_2} = x_2^4 \left[-\frac{1}{3} + \beta - \frac{1}{2}\beta^2 - \frac{5\gamma}{2}\left(-\beta + \frac{1}{2}\right)^{\frac{3}{2}}\right]$$

and on $\Gamma_- := \left\{x_1 = \beta x_2^2, x_2 < 0\right\}$, we have

$$\frac{\partial v}{\partial x_2} = x_2^4 \left[-\frac{1}{3} - \beta - \frac{1}{2}\beta^2 + \frac{5\gamma}{2}\left(\beta + \frac{1}{2}\right)^{\frac{3}{2}}\right]$$

Consequently, we need to select β, γ so that

$$\left\{\begin{array}{l} -\frac{1}{3} + \beta - \frac{1}{2}\beta^2 - \frac{5\gamma}{2}\left(-\beta + \frac{1}{2}\right)^{\frac{3}{2}} = 0 \\ -\frac{1}{3} - \beta - \frac{1}{2}\beta^2 + \frac{5\gamma}{2}\left(\beta + \frac{1}{2}\right)^{\frac{3}{2}} = 0 \end{array}\right. \tag{34}$$

Solving each equation for γ, we see (34) implies

$$\phi(\beta) := \left(\beta + \frac{1}{2}\right)^{\frac{3}{2}} \left(-\frac{1}{3} + \beta - \frac{1}{2}\beta^2\right) - \left(-\beta + \frac{1}{2}\right)^{\frac{3}{2}} \left(\frac{1}{3} + \beta + \frac{1}{2}\beta^2\right)$$
$$= 0$$

Since $\phi(0) < 0, \phi\left(\frac{1}{2}\right) > 0$, and $\phi(\theta)$ is continuous, then there exists $0 < \beta < \frac{1}{2}$ such that $\phi(\beta) = 0$. We can then find $\gamma > 0$ so that β, γ solve (34). A further calculation confirms that for these choices, $\frac{\partial v}{\partial x_2} < 0$ within Region I.

We then use (I) to extend our definition of v to $\mathbb{R}^2 \setminus \{0\}$, making $\frac{\partial v}{\partial x_2} > 0$ within Region II. This completes our construction of the value function satisfying Hamilton-Jacobi-Bellman equation, in this model.

$\square$

Form of ODE Solutions

We now figure out the optimal trajectory of the Fuller's Problem. When $\alpha = \pm 1$, indeed,

$$\frac{d}{dt}\left(x_1 \mp \frac{1}{2}(x_2)^2\right) = \dot{x}_1 \mp \alpha \cdot x_2 = x_2 \mp (\pm 1) \cdot x_2 = 0$$

Then a solution of (30) moves along a parabola of the form

$$x_1 = \frac{x_2^2}{2} + C \quad (\text{ drawn in green })$$

when $\alpha = 1$, and along a parabola of the form

$$x_1 = -\frac{x_2^2}{2} + C \quad (\text{ drawn in red })$$

when $\alpha = -1$. Furthermore, the optimal control switches from 1 to -1 (or vice versa) at the (blue) switching curve Γ given by the formula $x_1 = -\beta |x_2| x_2$ for $0 < \beta < \frac{1}{2}$.

The picture illustrates the field and a part of a typical optimal trajectory:

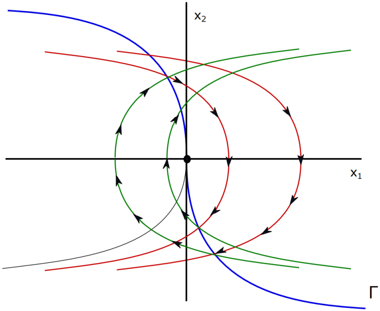


Figure 9: Part of an optimal path for Fuller's problem

A Final Remark on Chattering Control

Theorem.
As $0 < \beta < \frac{1}{2}$, such a trajectory will hit Γ infinitely many times. Consequently, the optimal control $\alpha_0(\cdot)$ will switch between ± 1 infinitely often before driving the state to the origin at a time $T < \infty$. We call $\alpha_0(\cdot)$ a **chattering control**.

Proof.
Assume that A_{i-1}, A_i, A_{i+1} are three successive points of switches:

$$A_{i-1} : (-\beta(x_2^{[i-1]})^2, x_2^{[i-1]})$$

$$A_i : (-\beta(x_2^{[i]})^2, x_2^{[i]})$$

$$A_{i+1} : (-\beta(x_2^{[i+1]})^2, x_2^{[i+1]})$$

Let t_{i-1}, t_i be the time intervals between these switches.

By integrating the ODE dynamics (30) starting from $\mathbf{x^0} = (x_1^0, x_2^0)$, when $\alpha = 1$, we have $x_1^0 = \beta \left(x_2^0\right)^2$, then

$$\begin{cases} x_1(t) = \frac{1}{2}t^2 + x_2^0 t + \beta \left(x_2^0\right)^2 \\ x_2(t) = t + x_2^0 \end{cases} \tag{35}$$

when $\alpha = -1$, we have $x_1^0 = -\beta \left(x_2^0\right)^2$, then

$$\begin{cases} x_1(t) = -\frac{1}{2}t^2 + x_2^0 t - \beta \left(x_2^0\right)^2 \\ x_2(t) = -t + x_2^0 \end{cases} \tag{36}$$

By (35) and (36), we get

$$\begin{cases} x_2^{[i-1]} - t_{i-1} = x_2^{[i]} \\ -\frac{1}{2}\left(t_{i-1}\right)^2 + x_2^{[i-1]} t_{i-1} - \beta \left(x_2^{[i-1]}\right)^2 = \beta \left(x_2^{[i]}\right)^2 \\ x_2^{[i]} + t_i = x_2^{[i+1]} \\ \frac{1}{2}\left(t_i\right)^2 + x_2^{[i]} t_i + \beta \left(x_2^{[i]}\right)^2 = -\beta \left(x_2^{[i+1]}\right)^2 \end{cases} \tag{37}$$

for A_{i-1}, A_i and A_{i+1}.

By eliminating t_{i-1}, t_i, we get:

$$\begin{cases} \left(\beta+\frac{1}{2}\right)\left(x_2^{[i+1]}\right)^2 = -\left(\beta-\frac{1}{2}\right)\left(x_2^{[i-1]}\right)^2 \\ -\left(\beta+\frac{1}{2}\right)\left(x_2^{[i+1]}\right)^2 = \left(\beta-\frac{1}{2}\right)\left(x_2^{[i]}\right)^2 \end{cases} \tag{38}$$

By division, we get:

$$\frac{x_2^{[i+1]}}{x_2^{[i]}} = \frac{x_2^{[i]}}{x_2^{[i-1]}} = -\sqrt{\frac{1-2\beta}{1+2\beta}} \in (-1, 0)$$

as $\beta \in \left(0, \frac{1}{2}\right)$.
Therefore,

$$x_2^{[i]} = \left(-\sqrt{\frac{1-2\beta}{1+2\beta}}\right)^{i-1} x_2^{[1]} \to 0 \text{ as } i \to \infty$$

In other words, the switching points form a geometric progression toward the origin, and there are countably many of them.

On the other hand,

$$\frac{t_i}{t_{i-1}} = -\frac{x_2^{[i+1]} - x_2^{[i]}}{x_2^{[i]} - x_2^{[i-1]}} = \sqrt{\frac{1-2\beta}{1+2\beta}} \in (0,1)$$

Hence

$$\sum_{i=1}^{\infty} t_i < \infty$$

In other words, the time we use for reaching the origin is finite. This completes the proof.

$\square$