Mathematics 195
MATHEMATICAL METHODS FOR OPTIMIZATION
Dynamic Optimization

Spring 2021 version

Lawrence C. Evans
Department of Mathematics
University of California, Berkeley

Contents

PREFACE

Last fall I taught a revised version of Math 170, primarily on finite dimensional optimization. This new spring class Math 195 discusses dynamic optimization, mostly the calculus of variations and optimal control theory. (However, Math 170 is not a prerequisite for Math 195, since we will be developing quite different mathematical tools.)

We continue to be grateful to Kurt and Evelyn Riedel for their very generous contribution to the Berkeley Math Department, in financial support of the redesign and expansion of our undergraduate classes in optimization theory.

The texts *Dynamic Optimization* by Kamien and Schwartz [**K-S**] and *Introduction to Optimal Control Theory* by Macki–Strauss [**M-S**] are good overall references for this class, and I also strongly recommend Levi, *Classical Mechanics with Calculus of Variations and Optimal Control* [**L**]. Part of the content in Chapters 4 and 5 is reworked from my old online lecture notes [**E**].

I have used Inkscape and SageMath for the illustrations. Thanks to David Hoffman for the beautiful pictures of minimal surfaces. I am again very thankful to have had Haotian Gu as my course assistant this term.

INTRODUCTION

Mathematical optimization theory comprises three major subareas:

A. Discrete optimization

B. Finite dimensional optimization

C. Infinite dimensional optimization.

This class covers several topics from infinite dimensional optimization theory, mainly the rigorous mathematical theories for *the calculus of variations* and *optimal control theory*. For these problems the unknowns are *functions*, and our main mathematical tools will be calculus and differential equations techniques. In most of our examples the unknowns will be functions of time, whence the name *dynamic optimization*.

The big math ideas for this class are

(i) First variation, Euler-Lagrange equations

(ii) Hamiltonian dynamics

(iii) Second variation

(iv) Pontryagin maximum principle

(v) Dynamic programming

While reading these notes students should carefully distinguish between the core mathematical theories and their applications. It is essential to understand how to write down for various problems the correct Euler-Lagrange equations or the correct form of the Pontryagin maximum principle. But these in turn may lead to problem-specific difficulties that can be quite hard.

I have written up in detail a lot of tricky mathematics needed for particular problems; students should read the calculations but should not let these particular issues deflect from their understanding of the larger mathematical framework.

Chapter 1

FIRST VARIATION

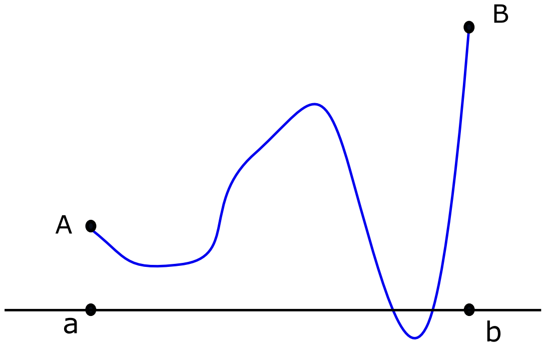
1.1. The calculus of variations

We introduce a class of optimization problems for which the unknown is a function.

DEFINITION. Assume $a < b$ and the points $y^0, y^1 \in \mathbb{R}$ are given. The corresponding set of **admissible functions** is

$$\begin{aligned}\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid\ & y(\cdot) \text{ is continuous and piecewise} \\ & \text{continuously differentiable}, y(a) = y^0, y(b) = y^1\}.\end{aligned}$$

So the graphs of functions $y(\cdot) \in \mathcal{A}$ connect the given endpoints $A = (a, y^0)$ and $B = (b, y^1)$.



Graph of an admissible function

NOTATION. We will often write "$y(\cdot)$" when we wish to emphasize that $y : [a,b] \to \mathbb{R}$ is a function. □

DEFINITION. The **Lagrangian** is a given continuous function

$$L : [a, b] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R},$$

written

$$L = L(x, y, z).$$

DEFINITION. If $y(\cdot) \in \mathcal{A}$ and L is a Lagrangian, we define the corresponding **integral functional**

$$\boxed{I[y(\cdot)] = \int_a^b L(x, y(x), y'(x))\, dx,} \tag{1.1}$$

where

$$y' = \frac{dy}{dx}.$$

Note that *we insert $y(x)$ into the y-variable slot of $L(x, y, z)$, and $y'(x)$ into the z-variable slot of $L(x, y, z)$.*

INTERPRETATION. We can informally think of the number $I[y(\cdot)]$ as being some sort of "energy" associated with the function $y(\cdot)$ (but there can be many other interesting interpretations). □

REMARK. To avoid various technical issues, *we will usually suppress mention of the precise degree of smoothness assumed for various functions that we discuss.* In particular, whenever we write down a derivative (or partial derivative) of some function at some point, the reader should suppose that the function is indeed differentiable there. □

The **basic problem in the calculus of variations** is to study functions $y_0(\cdot) \in \mathcal{A}$ that satisfy

$$\boxed{I[y_0(\cdot)] = \min_{y(\cdot) \in \mathcal{A}} I[y(\cdot)].} \tag{COV}$$

Does such a **minimizer** $y_0(\cdot)$ exist? What are its properties?

Different choices of the Lagrangian L give us different sorts of problems:

EXAMPLE (Shortest path between two points). Consider first the case that

$$L(x, y, z) = (1 + z^2)^{1/2}.$$

Then

$$\begin{aligned} I[y(\cdot)] &= \int_a^b (1 + (y')^2)^{1/2}\, dx \\ &= \text{length of the graph of } y(\cdot). \end{aligned}$$

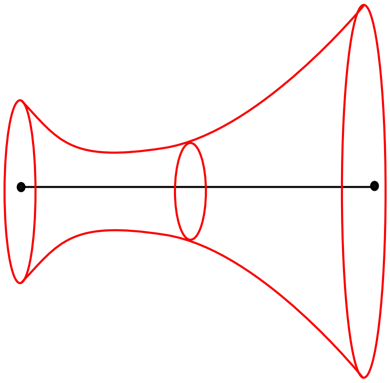
So a minimizer $y_0 \in \mathcal{A}$ will give the shortest path connecting the points $A = (a, y^0)$ and $B = (b, y^1)$, at least among curves that can be written as graphs of functions. The graph of the minimizer $y_0(\cdot)$ is obviously a straight line, but it will be interesting to see later what our general theory says even for this simple problem. □

EXAMPLE (Minimal surfaces of revolution). As a second example, take

$$L(x, y, z) = 2\pi y(1 + z^2)^{1/2}.$$

Then

$$\begin{aligned} I[y(\cdot)] &= 2\pi \int_a^b y\left(1 + (y')^2\right)^{1/2} dx \\ &= \text{area of surface of revolution of the graph.} \end{aligned}$$



A surface of revolution

What curve $y_0(\cdot)$ gives the surface of revolution of least surface area? This is more difficult than the previous example, and we will only later have the tools to handle this. □

1.2. Computing the first variation

1.2.1. Euler-Lagrange equation.

The most important insight of the calculus of variations is the next theorem. It says that *a minimizer $y_0(\cdot) \in \mathcal{A}$ automatically solves a certain ordinary differential equation (ODE).* This equation appears when we compute an appropriate first variation for our minimization problem (COV).

THEOREM 1.2.1. Assume $y_0(\cdot) \in \mathcal{A}$ solves (COV) and $y_0(\cdot)$ is twice continuously differentiable.

Then y_0 solves the nonlinear ODE

$$-\frac{d}{dx}\left(\frac{\partial L}{\partial z}(x, y_0(x), y_0'(x))\right) + \frac{\partial L}{\partial y}(x, y_0(x), y_0'(x)) = 0 \tag{1.2}$$

for $a \leq x \leq b$.

DEFINITIONS. (i) We call

$$\boxed{-\frac{d}{dx}\left(\frac{\partial L}{\partial z}(x, y, y')\right) + \frac{\partial L}{\partial y}(x, y, y') = 0} \tag{E-L}$$

the **Euler-Lagrange equation** corresponding to the Lagrangian L. This is a second-order, and usually nonlinear, ODE for the function $y = y(\cdot)$.

(ii) Solutions $y(\cdot)$ of the Euler-Lagrange equation are called **extremals** (or **critical points** or **stationary points**) of $I[\,\cdot\,]$.

(iii) Problems in mathematics or the sciences that lead to equations of the form (E-L) are called **variational**. □

REMARKS.

(i) Theorem 1.2.1 says that any minimizer y_0 solving (COV) satisfies the Euler-Lagrange differential equation and thus is an extremal. But a given extremal need not be a minimizer.

(ii) Remember that $' = \frac{d}{dx}$. So it is also correct to write (E-L) as

$$-\left(\frac{\partial L}{\partial z}(x, y, y')\right)' + \frac{\partial L}{\partial y}(x, y, y') = 0.$$

(iii) We could apply the chain rule to expand out the first term in (E-L), but it is almost always best not to do so. □

HOW TO WRITE DOWN THE EULER-LAGRANGE EQUATION FOR A SPECIFIC PROBLEM:

Step 1. Given $L = L(x, y, z)$, compute

$$\frac{\partial L}{\partial y}(x, y, z) \quad \text{and} \quad \frac{\partial L}{\partial z}(x, y, z).$$

Step 2. Plug in $y(x)$ for the variable y and $y'(x)$ for z, to obtain

$$\frac{\partial L}{\partial y}(x, y(x), y'(x)) \quad \text{and} \quad \frac{\partial L}{\partial z}(x, y(x), y'(x)).$$

Step 3. Now write (E-L):

$$-\frac{d}{dx}\left(\frac{\partial L}{\partial z}(x, y(x), y'(x))\right) + \frac{\partial L}{\partial y}(x, y(x), y'(x)) = 0.$$

WARNING ABOUT NOTATION. Most books write (E-L) as

$$-\frac{d}{dx}\left(\frac{\partial L}{\partial y'}(x,y,y')\right)+\frac{\partial L}{\partial y}(x,y,y')=0.$$

This is very common, but bad, notation. Note carefully: $L=L(x,y,z)$ is a function of the three real variables x,y,z; it has nothing to do with the derivative y' of some other function $y(\cdot)$. So "$\frac{\partial L}{\partial y'}$" has no meaning. □

The Euler-Lagrange equation is extremely important, since it provides us with a procedure for finding candidates for minimizers $y_0(\cdot)$ of (COV). We do so by trying to solve the (E-L) differential equation.

EXAMPLE. In the first example from page 4, the function y_0 minimizes $I[y]=\int_a^b(1+(y')^2)^{\frac{1}{2}}\,dx$ (the length of graph of y) among functions $y\in\mathcal{A}$.

The Euler-Lagrange equation is an ODE which provides useful information about y_0. For this example we have

$$L=(1+z^2)^{1/2},$$

and therefore

$$\frac{\partial L}{\partial y}=0,\quad \frac{\partial L}{\partial z}=\frac{z}{(1+z^2)^{1/2}}.$$

We insert y' for z and then write down (E-L):

$$\begin{aligned}0&=-\left(\frac{y'}{(1+(y')^2)^{1/2}}\right)'\\&=-y''(1+(y')^2)^{-1/2}-y'\left(-\frac{1}{2}\right)\left(1+(y')^2\right)^{-3/2}2y'y''\\&=-\frac{y''}{(1+(y')^2)^{3/2}}.\end{aligned}$$

Consequently a minimizer y_0 solves $\frac{y_0''}{\left(1+(y_0')^2\right)^{1/2}}=0$, and this implies

$$y_0''=0\qquad(a\le x\le b).$$

Hence the graph of y_0 is indeed a straight line connecting A and B.

GEOMETRIC INTERPRETATION. This conclusion is of course obvious, but our method suggests something interesting, namely that the expression

$$\left(\frac{y'}{(1+(y')^2)^{1/2}}\right)'=\frac{y''}{(1+(y')^2)^{3/2}}$$

may have a geometric meaning. It does. For any twice differentiable curve $y(\cdot)$,

$$\boxed{\kappa = \frac{y''}{(1+(y')^2)^{3/2}}} \tag{1.3}$$

is the **curvature** of the graph of $y(\cdot)$ at the point $(x, y(x))$. *The calculus of variations has automatically produced this important expression for the geometry of planar curves.* And what (E-L) really says is that the graph of our minimizer y_0 has constant curvature $\kappa = 0$. □

EXAMPLE. Compute the Euler-Lagrange equation satisfied by minimizers of

$$I[y(\cdot)] = \int_a^b \frac{(y')^2}{2} - fy\,dx$$

where $f : [a, b] \to \mathbb{R}$ is given.

In this case

$$L(x, y, z) = \frac{z^2}{2} - f(x)y, \quad \frac{\partial L}{\partial y} = -f(x), \quad \frac{\partial L}{\partial z} = z.$$

So (E-L) is the simple linear ODE

$$-y'' = f.$$

□

EXAMPLE. In the second example on page 5, we have

$$L(x, y, z) = 2\pi y(1 + z^2)^{1/2}.$$

Then

$$\frac{\partial L}{\partial y} = 2\pi(1 + z^2)^{1/2}, \quad \frac{\partial L}{\partial z} = \frac{2\pi yz}{(1 + z^2)^{1/2}}.$$

Consequently (E-L) reads

$$0 = -\left(\frac{yy'}{(1 + (y')^2)^{1/2}}\right)' + \left(1 + (y')^2\right)^{1/2}.$$

Which functions y solve this nonlinear ODE? We do not yet have the tools to answer this and so must return to this example later. □

EXAMPLE. Lagrangians of the form

$$L = a(y)z \tag{1.4}$$

are called **null Lagrangians**, meaning that *every function $y : [a, b] \to \mathbb{R}$ automatically solves the associated Euler-Lagrange equation.*

Indeed,

$$-\frac{d}{dx}\left(\frac{\partial L}{\partial z}(y,y')\right)+\frac{\partial L}{\partial y}(y,y')=-\frac{d}{dx}\left(a(y)\right)+a'(y)y'=0$$

for all functions y.

We will learn later that null Lagrangians, especially for more complicated variational problems, can provide useful information. □

1.2.2. Alternative notation.

In the examples above the variable x denotes a spatial position, but for many other applications the independent variable represents time t. In these situations it is appropriate to use different notation.

For such problems we consider Lagrangians

$$L:[0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$$

that depend upon the variables t denoting time, x denoting position, and v denoting velocity. So we will write

$$L=L(t,x,v).$$

The letter T gives a terminal time. We also redefine the admissible class to be

$$\begin{aligned}\mathcal{A}=\{x:[0,T]\to\mathbb{R}\mid x(\cdot)&\text{ is continuous and piecewise}\\ &\text{continuously differentiable}, x(0)=x^0, x(T)=x^1\}\end{aligned}$$

for given points $x^0, x^1\in\mathbb{R}$, and put

$$\boxed{I[x(\cdot)]=\int_0^T L(t,x(t),\dot{x}(t))\,dt.} \tag{1.5}$$

Observe that when the independent variable is t, we usually write

$$\dot{}=\frac{d}{dt}.$$

Employing this new notation, we check that the Euler-Lagrange equation for extremals $x(\cdot)$ now reads

$$\boxed{-\frac{d}{dt}\left(\frac{\partial L}{\partial v}(t,x,\dot{x})\right)+\frac{\partial L}{\partial x}(t,x,\dot{x})=0} \tag{E-L}$$

for $0\le t\le T$. There is no new mathematics here; we are simply changing notation by renaming the variables.

EXAMPLE. Consider the following simple model for the motion of a particle along the real line, moving under the influence of a potential energy. In this interpretation m denotes the mass, $x(t)$ is the position of the particle at time t, and $\dot{x}(t)$ is its velocity.

In addition,

$$\begin{aligned}\frac{m}{2}|\dot{x}(t)|^2 &= \text{kinetic energy at time } t,\\ W(x(t)) &= \text{potential energy at time } t,\end{aligned}$$

where $W : \mathbb{R} \to \mathbb{R}$ is given. The **action** of a path $x : [0,T] \to \mathbb{R}$ is the time integral of the *difference* between the kinetic and potential energies:

$$I[x(\cdot)] = \int_0^T \frac{m}{2}|\dot{x}|^2 - W(x(t))\,dt.$$

What is the corresponding Euler-Lagrange equation?

We have

$$L = \frac{mv^2}{2} - W(x), \quad \frac{\partial L}{\partial x} = -W'(x), \quad \frac{\partial L}{\partial v} = mv,$$

where $' = \frac{d}{dx}$. So (E-L) is

$$-\frac{d}{dt}(m\dot{x}) - W'(x) = 0,$$

which is **Newton's law of motion**:

$$m\ddot{x} = -W'(x).$$

In other words, $ma = f$ for the acceleration $a = \ddot{x}$ and force $f = -W'$. *The calculus of variations provides a systematic derivation for this fundamental law of physics.* □

1.2.3. Derivation.

In this section we prove that minimizers satisfy the Euler-Lagrange equation.

LEMMA 1.2.1. (i) If $f, g : [a,b] \to \mathbb{R}$ are continuously differentiable, we have the **integration by parts** formula

$$\int_a^b f'g\,dx = -\int_a^b fg'\,dx + f(b)g(b) - f(a)g(a). \tag{1.6}$$

(ii) Assume $f : [a,b] \to \mathbb{R}$ is continuous and

$$\int_a^b fw\,dx = 0 \tag{1.7}$$

for all continuously differentiable functions $w : [a,b] \to \mathbb{R}$ such that $w(a) = w(b) = 0$. Then

$$f(x) = 0 \quad \text{for all } a \leq x \leq b.$$

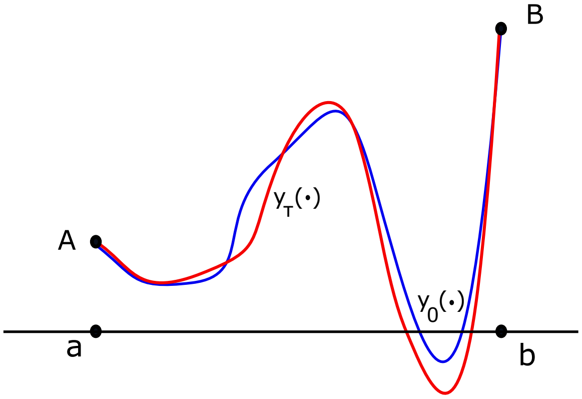
Proof. 1. Integrate $(fg)' = f'g + fg'$ from a to b.

2. A standard approximation argument shows if holds for all continuously differentiable functions w, it is valid also for all merely continuous functions w. Let $\phi : [a,b] \to \mathbb{R}$ be positive for $a < x < b$ and zero at the endpoints a, b. Put $w(x) = \phi(x)f(x)$ above, to find

$$\int_a^b \phi f^2 \, dx = 0.$$

Hence $\phi(x)f^2(x) = 0$ for all $x \in [a,b]$, since the integrand is nonnegative. Then since $\phi(x) > 0$ for all $x \in (a,b)$, we see that $f(x) = 0$ if $x \in (a,b)$. □

Derivation of Euler-Lagrange equation:



Computing the first variation

1. Let $w : [a,b] \to \mathbb{R}$ be continuously differentiable, with $w(a) = w(b) = 0$. Assume $-1 \leq \tau \leq 1$ and define

$$y_\tau(x) = y_0(x) + \tau w(x) \quad (a \leq x \leq b).$$

Note $y_\tau(\cdot) \in \mathcal{A}$, since $y_\tau(a) = y^0, y_\tau(b) = y^1$.

Thus

$$I[y_0(\cdot)] \leq I[y_\tau(\cdot)]$$

since $y_0(\cdot)$ is the minimizer of $I[\,\cdot\,]$. Define

$$i(\tau) = I[y_\tau(\cdot)].$$

Then

$$i(0) \leq i(\tau).$$

So $i(\cdot)$ has a minimum at $\tau = 0$ on the interval $-1 \leq \tau \leq 1$, and therefore

$$\frac{di}{d\tau}(0) = 0.$$

Our task now is to see what information we can extract from this simple formula.

2. We have

$$\begin{aligned} i(\tau) &= I[y_\tau(\cdot)] \\ &= \int_a^b L(x, y_\tau(x), (y_\tau)'(x))\, dx \\ &= \int_a^b L(x, y_0(x) + \tau w(x), y_0'(x) + \tau w'(x))\, dx. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{di}{d\tau}(\tau) &= \int_a^b \frac{\partial}{\partial \tau} L(x, y_0 + \tau w, y_0' + \tau w')\, dx \\ &= \int_a^b \frac{\partial L}{\partial y}(x, y_0 + \tau w, y_0' + \tau w') w + \frac{\partial L}{\partial z}(x, y_0 + \tau w, y_0' + \tau w') w'\, dx, \end{aligned}$$

where we used the chain rule. Next, set $\tau = 0$, to learn that

$$0 = \frac{di}{d\tau}(0) = \int_a^b \frac{\partial L}{\partial y}(x, y_0, y_0') w + \frac{\partial L}{\partial z}(x, y_0, y_0') w'\, dx.$$

We now integrate by parts, to deduce

$$\int_a^b \left[\frac{\partial L}{\partial y}(x, y_0, y_0') - \frac{d}{dx}\left(\frac{\partial L}{\partial z}(x, y_0, y_0') \right) \right] w\, dx = 0.$$

This is valid for all functions w such that $w(a) = w(b) = 0$. According then to the Lemma above, it follows that

$$\frac{\partial L}{\partial y}(x, y_0, y_0') - \frac{d}{dx}\left(\frac{\partial L}{\partial z}(x, y_0, y_0') \right) = 0$$

for all $a \leq x \leq b$. This is (E-L). □

REMARK. The procedure in this proof is called computing the **first variation**. □

1.3. Extensions and generalizations

In this section we discuss various extensions of the basic theory, focussing in particular upon how to use the Euler-Lagrange equation (E-L) to extract useful information. There are several general approaches for this:

(a) deriving exact formulas for extremals;

(b) calculating perturbative corrections from known solutions;

(c) applying rigorous ODE theory;

(d) introducing numerical methods.

In these notes we will stress (a) (although there is certainly no magic way to exactly solve all (E-L) equations) and (c).

1.3.1. Conservation laws.

We introduce first some methods for actually finding solutions of Euler-Lagrange equations in the various special cases. The idea is to try to reduce (E-L) to a (much simpler) first-order equation.

- **SPECIAL CASE 1: $L = L(x, z)$ does not depend on y.**

THEOREM 1.3.1. If L does not depend on y and the function $y(\cdot)$ solves (E-L), then

$$\boxed{\frac{\partial L}{\partial z}(x, y') \quad \text{is constant for } a \leq x \leq b.} \tag{1.8}$$

Proof. Since $\frac{\partial L}{\partial y} = 0$, (E-L) says

$$-\left(\frac{\partial L}{\partial z}(x, y')\right)' = 0;$$

and so $\frac{\partial L}{\partial z}(x, y')$ is a constant. □

REMARK. Why is this result useful? The point is that when

$$\frac{\partial L}{\partial z}(x, y') = C$$

for some constant C, we can perhaps rewrite this to solve for y':

$$y' = f(x, C).$$

Then

$$y(x) = \int_0^x f(t, C)\, dt + D$$

for constants C, D is a formula for general solution of (E-L). We can next try to select C, D so that the boundary conditions $y(a) = y^0, y(b) = y^1$ hold; in which case $y(\cdot) \in \mathcal{A}$. □

EXAMPLE. (a) Write down and then solve (E-L) for

$$I[y(\cdot)] = \int_a^b x^3 (y')^2 \, dx.$$

We have

$$L(x, y, z) = x^3 z^2, \quad \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial z} = 2x^3 z;$$

therefore

$$\frac{\partial L}{\partial z}(x, y'(x)) = 2x^3 y'(x) = C.$$

Hence

$$y'(x) = \frac{C}{2x^3},$$

and so

$$y(x) = \frac{E}{x^2} + F$$

for constants E, F.

(b) Find a minimizer of $I[\,\cdot\,]$ from the admissible class

$$\mathcal{A} = \{y : [1, 2] \to \mathbb{R} \mid y(1) = 3, y(2) = 4\}.$$

We need to select the constants E, F above so that

$$3 = y(1) = E + F, \ 4 = y(2) = \frac{E}{4} + F.$$

Solving, we find that $E = -\frac{4}{3}$, $F = \frac{13}{3}$, and thus

$$y_0(x) = -\frac{4}{3x^2} + \frac{13}{3}.$$

Therefore if (COV) has a solution, it must be this. □

• SPECIAL CASE 2: $L = L(y, z)$ does not depend on x.

THEOREM 1.3.2. If L does not depend on x and the function $y(\cdot)$ solves (E-L), then

$$\boxed{y' \frac{\partial L}{\partial z}(y, y') - L(y, y') \quad \text{is constant for } a \leq x \leq b.} \tag{1.9}$$

REMARK. Conversely, we will see from the proof that if $y' \frac{\partial L}{\partial z}(y, y') - L(y, y')$ is constant, then $y(\cdot)$ solves the Euler-Lagrange equation on any subinterval where $y' \neq 0$. □

Proof.

$$\begin{aligned}\left(L(y,y') - y'\frac{\partial L}{\partial z}(y,y')\right)' &= \frac{\partial L}{\partial y}y' + \frac{\partial L}{\partial z}y'' - y''\frac{\partial L}{\partial z} - y'\left(\frac{\partial L}{\partial z}\right)' \\ &= y'\left(-\left(\frac{\partial L}{\partial z}(y,y')\right)' + \frac{\partial L}{\partial y}(y,y')\right) \\ &= 0,\end{aligned}$$

since the expression in the parentheses is 0 according (E-L). □

Why is this useful? If

$$y'\frac{\partial L}{\partial z}(y,y') - L(y,y') = C,$$

then perhaps we can rewrite this expression into the form

$$y' = g(y,C).$$

This is a nonlinear first-order ODE that is solvable, at least in principle, when $g \neq 0$:

REVIEW: Solving a nonlinear first-order ODE. Let us recall how to solve nonlinear ODE of the form

$$y' = g(y).$$

First, introduce an antiderivative

$$G(y) = \int^{y} \frac{dt}{g(t)},$$

so that $G' = \frac{1}{g}$. Next, try to solve the algebraic expression

$$G(y) = x + D$$

for $y = y(x,D)$, where D is a constant.

We claim that $y(\cdot)$ solves the ODE $y' = g(y)$. To confirm this, notice that $G(y) = x + D$ implies $G'(y)y' = 1$. Hence $y' = g(y)$, since $G' = \frac{1}{g}$. □

EXAMPLE. We are now able to solve the Euler-Lagrange equation from the surface of revolution example on page 8 above. Recall that we have

$$L = y(1+z^2)^{1/2}, \quad \frac{\partial L}{\partial y} = (1+z^2)^{1/2}, \quad \frac{\partial L}{\partial z} = \frac{yz}{(1+z^2)^{1/2}}.$$

Then (E-L) says

$$0 = -\frac{d}{dx}\left(\frac{yy'}{(1+(y')^2)^{1/2}}\right) + (1+(y')^2)^{1/2},$$

and this is a difficult nonlinear second-order ODE.

But since L does not depend on x, we can apply Theorem 1.3.2. Now

$$\begin{aligned} y'\frac{\partial L}{\partial z} - L &= y'\frac{yy'}{(1+(y')^2)^{1/2}} - y(1+(y')^2)^{1/2} \\ &= -\frac{y}{(1+(y')^2)^{1/2}}. \end{aligned}$$

Therefore Theorem 1.3.2 tells us that

$$\frac{y}{(1+(y')^2)^{1/2}} = C$$

for some constant C. We solve this expression for

$$y' = \pm\left(\frac{y^2-C^2}{C^2}\right)^{1/2.}$$

We take the positive sign and solve this ODE:

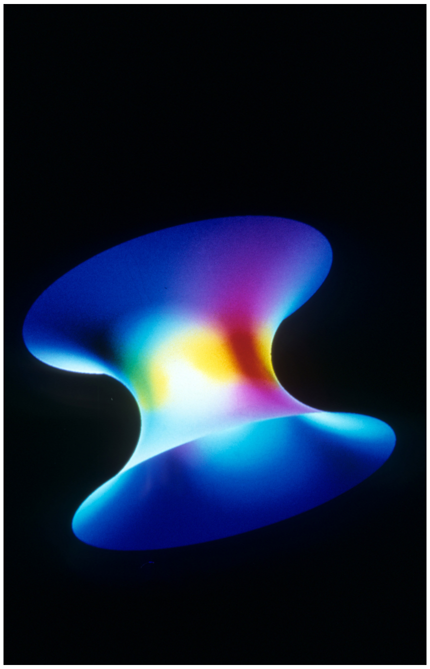
$$\begin{aligned} \frac{dy}{dx} &= \frac{(y^2-C^2)^{1/2}}{C} \\ \frac{dy}{(y^2-C^2)^{1/2}} &= \frac{dx}{C} \\ \int \frac{dy}{(y^2-C^2)^{1/2}} &= \int \frac{dx}{C} \\ \cosh^{-1}\left(\frac{y}{C}\right) &= \frac{x}{C} + D. \end{aligned}$$

(I looked up the expression for the y integral from a table of standard integrals.)

Therefore the curve giving a surface of revolution of least area is

$$y_0(x) = C\cosh\left(\frac{x}{C}+D\right),$$

where we recall that $\cosh(x) = \frac{e^x+e^{-x}}{2}$. The graph for the y-curve is called a **catenary**. The corresponding surface of revolution is a **catenoid**. □



A catenoid

REMARK. To fully resolve our problem we need try to adjust the constants C and D so the solution passes through the given endpoints. This however can be subtle and may not be possible: see Gilbert [**G**]. □

EXAMPLE. (Geometric optics) Suppose that the velocity of light v in some two-dimensional translucent material depends only upon the vertical coordinate y. Then the time for a light ray, moving along the path of the function $y(\cdot)$, to travel between two given points is

$$\int_a^b \frac{ds}{v(y)} = \int_a^b \frac{(1+(y')^2)^{\frac{1}{2}}}{v(y)} dx,$$

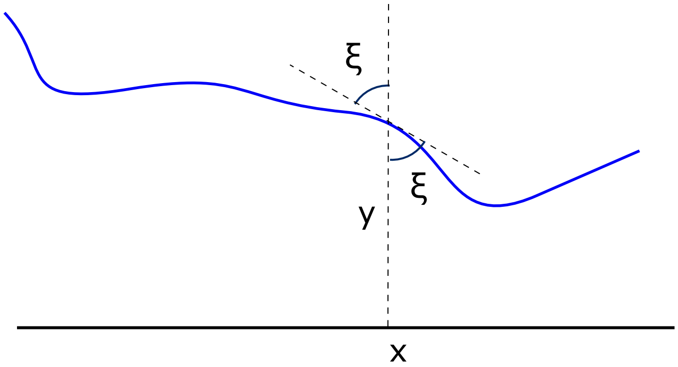
where s denotes arclength along the curve. The Lagrangian

$$L = L(y,z) = \frac{(1+z^2)^{\frac{1}{2}}}{v(y)}$$

does not depend upon x. Consequently if the graph of the function $y(\cdot)$ describes the path along which light travels, we know that

$$L(y,y') - y'\frac{\partial L}{\partial z}(y,y') = \frac{1}{v(y)(1+(y')^2)^{\frac{1}{2}}} = \frac{\sin\xi}{v},$$

is constant, where ξ is the angle of the tangent with the vertical, as drawn.



Angles and derivatives

Thus

$$\frac{\sin \xi}{v(y)} = C, \tag{1.10}$$

for some constant C. This is a continuous version of **Snell's Law** of diffraction (see the Math 170 lecture notes). □

EXAMPLE. Recall for our model for the motion of a particle on the line that

$$I[x(\cdot)] = \int_a^b \frac{m}{2}|\dot{x}|^2 - W(x(t))\,dt$$

with

$$L = \frac{mv^2}{2} - W(x).$$

We compute

$$\begin{aligned}\dot{x}\frac{\partial L}{\partial v} - L &= \dot{x}(m\dot{x}) - \left(\frac{m(\dot{x})^2}{2} - W(x)\right)\\ &= \frac{m(\dot{x})^2}{2} + W(x).\end{aligned}$$

Since L does not depend upon t, Theorem 1.3.2 implies that the above expression is constant. So

$$\begin{aligned}\text{total energy} &= \text{kinetic energy} + \text{potential energy}\\ &= \frac{m(\dot{x})^2}{2} + W(x)\end{aligned}$$

is constant for all times $a \leq t \leq b$. *The calculus of variations therefore predicts the physical law of conservation of total energy.* □

1.3.2. Transversality conditions.

a. Free endpoint problems. Our definition of the admissible class $\mathcal{A}$ on page 3 forces the prescribed boundary conditions that $y(a) = y^0, y(b) = y^1$. But what if we change the admissible class so as to require only, say, that $y(a) = y^0$? That is, suppose we *redefine* the class of admissible functions, now to be

$$\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid y(\cdot) \text{ is continuous and piecewise continuously differentiable}, y(a) = y^0\}, \tag{1.11}$$

and so require nothing about the values at $x = b$ for functions $y(\cdot) \in \mathcal{A}$.

We as usual define

$$I[y(\cdot)] = \int_a^b L(x, y, y')\,dx$$

for functions $y(\cdot) \in \mathcal{A}$; and seek to understand the behavior of a minimizer $y_0(\cdot) \in \mathcal{A}$ of this **free endpoint problem**.

THEOREM 1.3.3. Let the admissible class be given by (1.11). Assume $y_0(\cdot) \in \mathcal{A}$ solves (COV) and is twice continuously differentiable.

(i) Then y_0 solves the Euler-Lagrange equation

$$-\left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)' + \frac{\partial L}{\partial y}(x, y_0, y_0') = 0 \qquad (a \le x \le b) \tag{1.12}$$

(ii) Furthermore,

$$\boxed{\frac{\partial L}{\partial z}(b, y_0(b), y_0'(b)) = 0.} \tag{1.13}$$

INTERPRETATION. So the Euler-Lagrange equation is as before, whereas the new formula (1.13) appears at the free endpoint $x = b$.

This so-called **transversality condition** (or **natural boundary condition**) is implicit in the variational formulation and, as we will see, appears automatically when we compute the first variation. □

Proof. 1. We appropriately modify our earlier derivation of (E-L). So let $w : [a,b] \to \mathbb{R}$, with $w(a) = 0$. Assume $-1 \le \tau \le 1$ and define

$$y_\tau(x) = y_0(x) + \tau w(x) \quad (a \le x \le b).$$

Observe that $y_\tau(\cdot) \in \mathcal{A}$, since $y_\tau(a) = y^0$. Then $I[y_0(\cdot)] \le I[y_\tau(\cdot)]$. Define $i(\tau) = I[y_\tau(\cdot)]$, and, as before, observe that $\frac{di}{d\tau}(0) = 0$.

As in the earlier proof, we have

$$0 = \frac{di}{d\tau}(0) = \int_a^b \frac{\partial L}{\partial y}(x, y_0, y_0')w + \frac{\partial L}{\partial z}(x, y_0, y_0')w'\, dx.$$

Integrate by parts in the second term, remembering that $w(a) = 0$, but that $w(b)$ need not necessarily vanish:

$$\begin{aligned} \int_a^b \left[\frac{\partial L}{\partial y}(x, y_0, y_0') - \left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)'\right] w\, dx \\ + \frac{\partial L}{\partial z}(b, y_0(b), y_0'(b))w(b) = 0. \end{aligned} \tag{1.14}$$

2. If we now assume also that $w(b) = 0$, then (1.14) gives

$$\int_a^b \left[\frac{\partial L}{\partial y}(x, y_0, y_0') - \left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)'\right] w\, dx = 0.$$

That this integral identity holds for all variations w satisfying $w(a) = w(b) = 0$ implies the (E-L) equation (1.12).

Now, drop the assumption that $w(b) = 0$ and return to (1.14). Since we now know that (1.12) holds, we deduce from (1.14) that

$$\frac{\partial L}{\partial z}(b, y_0(b), y_0'(b))w(b) = 0.$$

This is valid for all choices of $w(b)$ and consequently the natural boundary condition (1.13) holds. □

EXAMPLE. The minimizer $y_0(\cdot)$ of

$$I[y(\cdot)] = \int_0^1 \frac{(y')^2}{2} - fy\, dx,$$

subject to $y(0) = 0$, satisfies

$$-y_0'' = f \quad (0 \leq x \leq 1), \quad y_0(0) = 0,\ y_0'(1) = 0.$$

□

b. Free endtime problems. For many interesting time-dependent problems, we are asked to minimize an integral functional that depends both upon a curve $x : [0, T] \to \mathbb{R}$ *and a variable end time* $T > 0$ at which some specified terminal condition holds.

To be more precise, let us switch to the notation introduced in Section 1.2.2 for time-dependent problems. We select points $x^0, x^1 \in \mathbb{R}$ and introduce the new admissible class

$$\mathcal{A} = \{(T, x(\cdot)) \mid T > 0, x : [0, T] \to \mathbb{R},\ x(0) = x^0, x(T) = x^1\}, \tag{1.15}$$

over which we propose to minimize

$$I[T, x(\cdot)] = \int_0^T L(t, x(t), \dot{x}(t))\, dt.$$

Note carefully: we are now selecting both the terminal time $T > 0$ and the curve $x : [0, T] \to \mathbb{R}$.

THEOREM 1.3.4. Let the admissible class be given by (1.15) and assume $(T_0, x_0(\cdot)) \in \mathcal{A}$ minimizes $I[\,\cdot\,]$.

(i) Then x_0 solves the Euler-Lagrange equation

$$-\frac{d}{dt}\left(\frac{\partial L}{\partial v}(t, x_0, \dot{x}_0)\right) + \frac{\partial L}{\partial x}(t, x_0, \dot{x}_0) = 0 \qquad (0 \le t \le T_0). \tag{1.16}$$

(ii) Furthermore, we have

$$\boxed{\frac{\partial L}{\partial v}(T_0, x_0(T_0), \dot{x}_0(T_0))\dot{x}_0(T_0) - L(T_0, x_0(T_0), \dot{x}_0(T_0)) = 0.} \tag{1.17}$$

INTERPRETATION. The Euler-Lagrange equation (1.16) is same as before; whereas the new **free endtime transversality condition** (1.17) appears at the optimal endtime T_0. This is new information that we will discover by computing a variation in the arrival time. □

REMARK. If $L = L(x, v)$ does not depend on t, then according to Theorem 1.3.2, we furthermore see that

$$\boxed{\frac{\partial L}{\partial v}(x_0, \dot{x}_0)\dot{x}_0 - L(x_0, \dot{x}_0) = 0} \qquad (0 \le t \le T_0). \tag{1.18}$$

□

Proof. 1. That the usual Euler-Lagrange equation (1.16) holds follows as in the proof of Theorem 1.3.3.

2. To prove (1.17), we introduce a new sort of variation by scaling our minimizer $x_0(\cdot)$ in the time variable. For this, select $-1 < \sigma < 1$ and define

$$T_\sigma = \tfrac{T_0}{1+\sigma}, \quad x_\sigma(t) = x_0(t(1+\sigma)) \qquad (0 \le t \le T_\sigma).$$

Then $(T_\sigma, x_\sigma(\cdot)) \in \mathcal{A}$ and thus

$$j(\sigma) = \int_0^{T_\sigma} L(t, x_\sigma(t), \dot{x}_\sigma(t))\, dt$$

has a minimum at $\sigma = 0$. Therefore

$$\frac{dj}{d\sigma}(0) = 0.$$

3. Now

$$\frac{dT_\sigma}{d\sigma} = -\frac{T_0}{(1+\sigma)^2}$$

and

$$\frac{\partial x_\sigma}{\partial \sigma} = \dot{x}_0(t(1+\sigma))t.$$

Therefore

$$\begin{aligned} 0 = \frac{dj}{d\sigma}(0) = -T_0 L(T_0, x_0(T_0), \dot{x}_0(T_0)) \\ + \int_0^{T_0} \frac{\partial L}{\partial x}(t, x_0, \dot{x}_0) w + \frac{\partial L}{\partial v}(t, x_0, \dot{x}_0)\dot{w}\, dt, \end{aligned}$$

for

$$w(t) = \dot{x}_0(t)t.$$

We integrate by parts in the integral and recall (1.16), to deduce that

$$0 = -T_0 L(T_0, x_0(T_0), \dot{x}_0(T_0)) + \frac{\partial L}{\partial v}(T_0, x_0(T_0), \dot{x}_0(T_0))\dot{x}_0(T_0)T_0.$$

This gives (1.17), since $T_0 > 0$. □

The interesting book by Kamien and Schwartz [**K-S**] has further information about transversality conditions in various more complicated situations.

1.3.3. Integral constraints.

Constraints involving integrals appear often in the calculus of variations. For a model such problem, assume in addition to the Lagrangian $L = L(x, y, z)$ we are given also a function $G : [a, b] \times \mathbb{R} \to \mathbb{R}$, $G = G(x, y)$. Define then the integral functional

$$\boxed{J[y(\cdot)] = \int_a^b G(x, y)\, dx.}$$

We use $J[\,\cdot\,]$ to define a new admissible class:

$$\mathcal{A} = \{y : [a, b] \to \mathbb{R} \mid y(a) = y^0,\ y(b) = y^1, J[y(\cdot)] = 0\}.$$

So we are now requiring the additional **integral constraint** that $J[y(\cdot)] = 0$. We continue as usual to write

$$I[y(\cdot)] = \int_a^b L(x, y, y')\, dx.$$

THEOREM 1.3.5. Assume that $y_0(\cdot) \in \mathcal{A}$ is a minimizer of $I[\,\cdot\,]$ over $\mathcal{A}$. Suppose also that

$$\frac{\partial G}{\partial y}(x, y_0) \text{ is not identically zero for all } a \le x \le b. \tag{1.19}$$

Then there exists $\lambda_0 \in \mathbb{R}$ such that

$$\boxed{-\left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)' + \frac{\partial L}{\partial y}(x, y_0, y_0') + \lambda_0 \frac{\partial G}{\partial y}(x, y_0) = 0} \tag{1.20}$$

for $a \le x \le b$.

INTERPRETATION. We understand λ_0 as the **Lagrange multiplier** corresponding to the integral equality constraint that $J[x(\cdot)] = 0$. The hypothesis (1.19) is a **constraint qualification** condition, ensuring the existence of the Lagrange multiplier. See the Math 170 notes for lots more about constraint qualification conditions in finite dimensional optimization. □

Proof. 1. Select $w : [a, b] \to \mathbb{R}$ with $w(a) = w(b) = 0$. We want to design a variation involving w, but setting $y_0 + \tau w$ for small τ will not work, since this function will probably not belong to $\mathcal{A}$. We must build some sort of correction, to restore the integral constraint.

Now the condition (1.19) implies that we can find a smooth function $v : [a, b] \to \mathbb{R}$, such that $v(a) = v(b) = 0$ and

$$\int_a^b \frac{\partial G}{\partial y}(x, y_0) v \, dx \neq 0. \tag{1.21}$$

Define

$$\Phi(\tau, \sigma) = \int_a^b G(x, y_0 + \tau w + \sigma v) \, dx.$$

Then

$$\Phi(0, 0) = \int_a^b G(x, y_0) \, dx = J[y_0(\cdot)] = 0$$

and

$$\frac{\partial \Phi}{\partial \sigma}(0, 0) = \int_a^b \frac{\partial G}{\partial y}(x, y_0) v \, dx \neq 0.$$

Therefore the Implicit Function Theorem (see the Appendix) tells us that for some small $\tau_0 >$ there exists a function

$$\phi : [-\tau_0, \tau_0] \to \mathbb{R}$$

such that $\phi(0) = 0$ and

$$\Phi(\tau, \phi(\tau)) = \Phi(0, 0) = 0 \qquad (-\tau_0 \le \tau \le \tau_0).$$

Let us differentiate this expression in τ, to learn that

$$\frac{\partial \Phi}{\partial \tau}(0,0) + \frac{\partial \Phi}{\partial \sigma}(0,0)\phi'(0) = 0,$$

where $\phi' = \frac{d\phi}{d\tau}$. Since

$$\frac{\partial \Phi}{\partial \tau}(0,0) = \int_a^b \frac{\partial G}{\partial y}(x, y_0) w \, dx,$$

it follows that

$$\int_a^b \frac{\partial G}{\partial y}(x, y_0) w \, dx + \phi'(0) \int_a^b \frac{\partial G}{\partial y}(x, y_0) v \, dx = 0. \tag{1.22}$$

2. Now define

$$y_\tau(x) = y_0(x) + \tau w(x) + \phi(\tau) v(x) \qquad (-\tau_0 \leq \tau \leq \tau_0).$$

Then $y_\tau(a) = y^0$, $y_\tau(b) = y^1$, and

$$J[y_\tau(\cdot)] = \int_a^b G(x, y_0 + \tau w + \phi(\tau) v) \, dx = \Phi(\tau, \phi(\tau)) = 0;$$

therefore $y_\tau(\cdot) \in \mathcal{A}$. Hence $i(\tau) = I[y_\tau(\cdot)]$ has a minimum at $\tau = 0$, and consequently

$$\frac{di}{d\tau}(0) = 0.$$

We will extract the Lagrange multiplier from this simple looking equality.

3. We compute

$$\begin{aligned}\frac{di}{d\tau}(\tau) = &\int_a^b \frac{\partial L}{\partial y}(x, y_0 + \tau w + \phi(\tau) v, y_0' + \tau w' + \phi(\tau) v')(w + \phi'(\tau) v) \\ &+ \int_a^b \frac{\partial L}{\partial z}(x, y_0 + \tau w + \phi(\tau) v, y_0' + \tau w' + \phi(\tau) v')(w' + \phi'(\tau) v') \, dx.\end{aligned}$$

Now put $\tau = 0$ and then integrate by parts:

$$\begin{aligned} 0 = \frac{di}{d\tau}(0) &= \int_a^b \frac{\partial L}{\partial y}(w + \phi'(0) v) + \frac{\partial L}{\partial z}(w' + \phi'(0) v') \, dx \\ &= \int_a^b \left[\frac{\partial L}{\partial y} - \left(\frac{\partial L}{\partial z} \right)' \right] (w + \phi'(0) v) \, dx, \end{aligned} \tag{1.23}$$

where L is evaluated at (x, y_0, y_0').

4. We next define the Lagrange multiplier to be

$$\lambda_0 = -\frac{\int_a^b \left[\frac{\partial L}{\partial y} - \left(\frac{\partial L}{\partial z} \right)' \right] v \, dx}{\int_a^b \frac{\partial G}{\partial y} v \, dx},$$

in which L is evaluated at (x, y_0, y_0') and G is evaluated at (x, y_0). Then (1.22) implies

$$\phi'(0)\int_a^b \left[\frac{\partial L}{\partial y} - \left(\frac{\partial L}{\partial z}\right)'\right] v\,dx = -\phi'(0)\lambda_0 \int_a^b \frac{\partial G}{\partial y} v\,dx = \lambda_0 \int_a^b \frac{\partial G}{\partial y}(x, y_0) w\,dx.$$

We utilize this calculation in (1.23), to find that

$$\int_a^b \left[\frac{\partial L}{\partial y} - \left(\frac{\partial L}{\partial z}\right)' + \lambda_0 \frac{\partial G}{\partial y}\right] w\,dx = 0.$$

This identity is valid for all functions w as above, and therefore the ODE (1.20) holds.

□

EXAMPLE. (Isoperimetric problem) We wish to find a curve $y(\cdot) \in \mathcal{A}$ to minimize the length

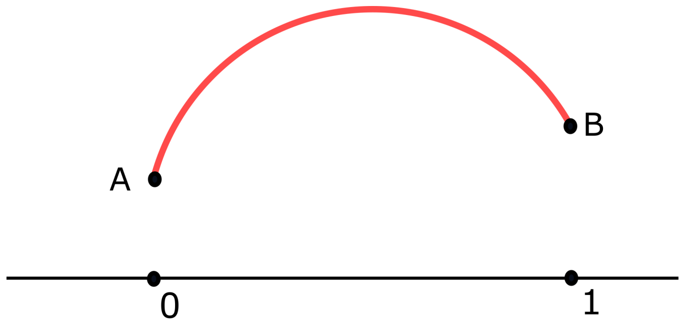
$$I[y(\cdot)] = \int_0^1 (1 + (y')^2)^{1/2}\,dx$$

among curves connecting the given endpoints A, B and having with a given area a under the the graph:

$$J[y(\cdot)] = \int_0^1 y\,dx = a.$$

The Euler-Lagrange equation reads

$$\left(\frac{y'}{(1 + (y')^2)^{1/2}}\right)' = \lambda.$$



Recall from (1.3) that this says *the curvature κ is constant.* Therefore (as we will prove later, on page 48) the graph of $y(\cdot)$ is an arc of a circle connecting the given endpoints. □

Generalization. We can extend the foregoing to handle more complicated integral constraints having the form

$$\boxed{J[y(\cdot)] = \int_a^b G(x, y(x), y'(x))\, dx}$$

where $G : [a,b] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $G = G(x,y,z)$.

THEOREM 1.3.6. Assume that $y_0 \in \mathcal{A}$ is a minimizer of $I[\,\cdot\,]$ over $\mathcal{A}$. Suppose also that

$$(1.24) \qquad -\left(\frac{\partial G}{\partial z}(x, y_0, y_0')\right)' + \frac{\partial G}{\partial y}(x, y_0, y_0')$$

is not identically zero on the interval $[a,b]$.

Then there exists $\lambda_0 \in \mathbb{R}$ such that

$$(1.25) \qquad -\left(\frac{\partial L}{\partial z}(x, y_0, y_0') + \lambda_0 \frac{\partial G}{\partial z}(x, y_0, y_0')\right)' + \left(\frac{\partial L}{\partial y}(x, y_0, y_0') + \lambda_0 \frac{\partial G}{\partial y}(x, y_0, y_0')\right) = 0$$

for $a \le x \le b$.

We omit the proof, which is similar to that for the previous theorem.

REMARKS. (i) The ODE (1.25) is of course the Euler-Lagrange equation for the new Lagrangian $K = L + \lambda_o G$.

(ii) It is not hard to see that the same Euler-Lagrange equation (1.25) holds if we change the constraint to read $J[y(\cdot)] = C$ for any constant C. □

EXAMPLE. (Hanging chain) A chain of constant mass density and length l hangs between the points $A^{\pm} = (\pm a, 0)$. What is its shape?

The problem is to minimize the gravitational potential energy

$$I[y(\cdot)] = \int_{-a}^{a} y(1 + (y')^2)^{\frac{1}{2}}\, dx,$$

subject to $y(\pm a) = 0$ and the length constraint

$$J[y(\cdot)] = \int_{-a}^{a} (1 + (y')^2)^{\frac{1}{2}}\, dx = l.$$

Here $L = y(1+z^2)^{\frac{1}{2}}$, $G = (1+z^2)^{\frac{1}{2}}$. Since $K = L + \lambda G$ does not depend on x, we see from the Euler-Lagrange equation (1.25) that

$$-y'\left(\frac{\partial L}{\partial z} + \lambda \frac{\partial G}{\partial z}\right)(y, y') + (L + \lambda G)(y, y') = \frac{y + \lambda}{(1 + (y')^2)^{\frac{1}{2}}}$$

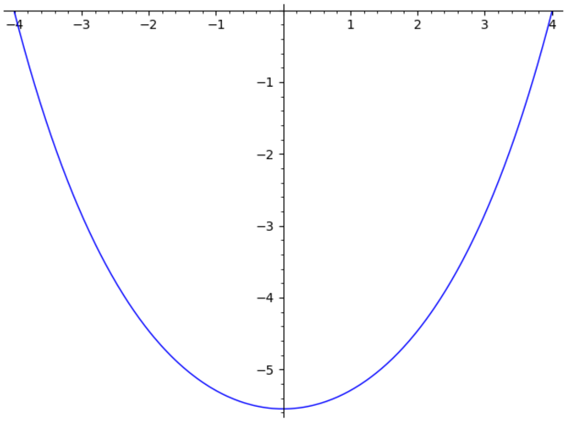
is constant. Thus $\bar{y} = y + \lambda$ satisfies

$$\frac{\bar{y}}{(1+(\bar{y}')^2)^{\frac{1}{2}}} = C$$

for some constant C, and this is an ODE we have solved earlier, on page 16. We thereby obtain the symmetric catenary $\bar{y}(x) = C\cosh\left(\frac{x}{C}\right)$; and consequently

$$y_0(x) = C\cosh\left(\frac{x}{C}\right) - \lambda.$$

We now adjust C so that $\int_{-a}^{a}(1+(y')^2)^{\frac{1}{2}}\,dx = l$, and then select λ so that $y(\pm a) = 0$.



Catenary

Now if $l < 2a$, the admissible class is empty and we will not be able to select C as above. If $l = 2a$, the admissible class consists only of one configuration, for which chain is stretched horizontally between its left and right endpoints. The constraint qualification condition (1.24) then fails. □

1.3.4. Systems.

We next turn attention to calculus of variations problems for functions $\mathbf{y} : [a,b] \to \mathbb{R}^n$. The new difficulties are mostly notational, as the basic ideas are the same as above.

NOTATION. (i) We write

$$\mathbf{y}(x) = \begin{bmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{bmatrix}, \quad \mathbf{y}'(x) = \begin{bmatrix} y_1'(x) \\ \vdots \\ y_n'(x) \end{bmatrix}.$$

(ii) The admissible class is

$$\mathcal{A} = \{\mathbf{y} : [a,b] \to \mathbb{R}^n \mid y(\cdot) \text{ is continuous and piecewise} \\ \text{continuously differentiable}, \mathbf{y}(a) = y^0, \mathbf{y}(b) = y^1\},$$

where $y^0, y^1 \in \mathbb{R}^n$ are given.

(iii) We are given a Lagrangian function $L : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$L = L(x,y,z) = L(x, y_1, \dots, y_n, z_1, \dots, z_n),$$

with

$$\nabla_y L = \begin{bmatrix} \frac{\partial L}{\partial y_1} \\ \vdots \\ \frac{\partial L}{\partial y_n} \end{bmatrix}, \quad \nabla_z L = \begin{bmatrix} \frac{\partial L}{\partial z_1} \\ \vdots \\ \frac{\partial L}{\partial z_n} \end{bmatrix}.$$

(iv) We write

$$I[\mathbf{y}(\cdot)] = \int_a^b L(x, \mathbf{y}(x), \mathbf{y}'(x))\, dx.$$

Our problem now is to study functions $\mathbf{y}_0(\cdot) \in \mathcal{A}$ that satisfy

$$\text{(COV)} \qquad \boxed{I[\mathbf{y}_0(\cdot)] = \min_{\mathbf{y}(\cdot) \in \mathcal{A}} I[\mathbf{y}(\cdot)].}$$

THEOREM 1.3.7. Suppose $\mathbf{y}_0(\cdot) \in \mathcal{A}$ solves (COV) and is twice continuously differentiable. Then $\mathbf{y}_0(\cdot)$ solves the Euler-Lagrange system of ODE

$$\text{(E-L)} \qquad \boxed{-\frac{d}{dx}\left(\frac{\partial L}{\partial z_k}(x, \mathbf{y}, \mathbf{y}')\right) + \frac{\partial L}{\partial y_k}(x, \mathbf{y}, \mathbf{y}') = 0} \qquad (k = 1, \dots, n).$$

REMARKS. (i) In vector notation (E-L) reads

$$-\left(\nabla_z L(x, \mathbf{y}, \mathbf{y}')\right)' + \nabla_y L(x, \mathbf{y}, \mathbf{y}') = 0.$$

These comprise n coupled second-order ODE for the n unknown functions $y_1(\cdot), \dots, y_n(\cdot)$ that are the components of $\mathbf{y}(\cdot)$

(ii) If time t is the independent variable and $L = L(t,x,v)$, the Euler-Lagrange system of ODE is written

$$\text{(E-L)} \qquad -\frac{d}{dt}\left(\frac{\partial L}{\partial v_k}(t, \mathbf{x}, \dot{\mathbf{x}})\right) + \frac{\partial L}{\partial x_k}(t, \mathbf{x}, \dot{\mathbf{x}}) = 0$$

for $k = 1, \dots, n$. In vector form, this is

$$(1.26) \qquad -\frac{d}{dt}\left(\nabla_v L(t, \mathbf{x}, \dot{\mathbf{x}})\right) + \nabla_x L(t, \mathbf{x}, \dot{\mathbf{x}}) = 0.$$

□

Proof. 1. We extend our previous derivation of the Euler-Lagrange equation to this vector case. Select $\mathbf{w} : [a,b] \to \mathbb{R}$, written

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix},$$

such that $\mathbf{w}(a) = \mathbf{w}(b) = 0$. Then define

$$\mathbf{y}_\tau(x) = \mathbf{y}_0(x) + \tau \mathbf{w}(x) \qquad (a \le x \le b)$$

for $-1 \le \tau \le 1$. We have $\mathbf{y}_\tau(\cdot) \in \mathcal{A}$, and consequently

$$I[\mathbf{y}_0(\cdot)] \le I[\mathbf{y}_\tau(\cdot)].$$

Define $i(\tau) = I[\mathbf{y}_\tau(\cdot)]$, so that $i(\cdot)$ has a minimum at $\tau = 0$. Therefore

$$\frac{di}{d\tau}(0) = 0.$$

2. Since

$$i(\tau) = \int_a^b L(x, \mathbf{y}_0(x) + \tau \mathbf{w}(x), \mathbf{y}_0'(x) + \tau \mathbf{w}'(x))\, dx,$$

we can apply the chain rule to compute

$$\begin{aligned} \frac{di}{d\tau}(\tau) = \int_a^b \sum_{l=1}^n \frac{\partial L}{\partial y_l}(x, \mathbf{y}_0 + \tau \mathbf{w}, \mathbf{y}_0' + \tau \mathbf{w}') w_l \\ + \sum_{l=1}^n \frac{\partial L}{\partial z_l}(x, \mathbf{y}_0 + \tau \mathbf{w}, \mathbf{y}_0' + \tau \mathbf{w}') w_l' \, dx. \end{aligned}$$

Thus

$$0 = \frac{di}{d\tau}(0) = \int_a^b \sum_{l=1}^n \frac{\partial L}{\partial y_l}(x, \mathbf{y}_0, \mathbf{y}_0') w_l + \sum_{l=1}^n \frac{\partial L}{\partial z_l}(x, \mathbf{y}_0, \mathbf{y}_0') w_l' \, dx.$$

Now fix some index $k \in \{1, \dots, n\}$ and put

$$\mathbf{w} = [0 \ \dots \ 0 \, w \, 0 \ \dots \ 0]^T,$$

where the real-valued function $w : [a,b] \to \mathbb{R}$ appears in the k-th slot. Then we have

$$\int_U \frac{\partial L}{\partial y_k}(x, \mathbf{y}_0, \mathbf{y}_0') w + \frac{\partial L}{\partial z_k}(x, \mathbf{y}_0, \mathbf{y}_0') w' \, dx = 0.$$

Upon integrating by parts, we deduce as usual that the k-th equation of the stated Euler-Lagrange system (E-L) holds. □

EXAMPLE. (Motion of particle in space) For this example the independent variable is t (for time) and if $\mathbf{x} : [a,b] \to \mathbb{R}^n$, we regard $\mathbf{x}(t)$ as the position at time t of a particle with mass m moving in $\mathbb{R}^n$. The **action** of any such path is

$$I[\mathbf{x}(\cdot)] = \int_a^b \frac{m|\dot{\mathbf{x}}|^2}{2} - W(\mathbf{x})\, dt.$$

So

$$L = \frac{m|v|^2}{2} - W(x), \quad \nabla_v L = mv, \quad \nabla_x L = -\nabla W(x).$$

The Euler-Lagrange system of equations give Newton's law

$$m\ddot{\mathbf{x}} = -\nabla W(\mathbf{x})$$

for the motion of a particle in space governed by the potential energy W.

The path of the particle is thus an extremal of the action. It is sometimes said that the path of the particle satisfies the *principle of least action*. But this terminology is misleading: the path is an extremal of the action, but is not necessarily a minimizer. □

THEOREM 1.3.8. Suppose $\mathbf{y}(\cdot)$ is an extremal.

(i) If $L = L(x,z)$ does not depend on y, then

$$\nabla_z L(x, \mathbf{y}') \text{ is constant for } a \leq x \leq b.$$

More generally, if L does not depend upon y_k for some $k \in \{1, \dots, n\}$, then

$$\frac{\partial L}{\partial z_k}(x, \mathbf{y}, \mathbf{y}') \text{ is constant for } a \leq x \leq b. \tag{1.27}$$

(ii) If $L = L(y,z)$ does not depend on x, then

$$\mathbf{y}' \cdot \nabla_z L(\mathbf{y}, \mathbf{y}') - L(\mathbf{y}, \mathbf{y}') \text{ is constant for } a \leq x \leq b. \tag{1.28}$$

The proof of (1.27) is simple, and the proof of (1.28) is similar to that for our earlier Theorem 1.3.2.

EXAMPLE. For the motion of a particle in space, the Lagrangian does not depend upon t, and therefore the total energy

$$\dot{\mathbf{x}} \cdot \nabla L(\mathbf{x}, \dot{\mathbf{x}}) - L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{m|\dot{\mathbf{x}}|^2}{2} + W(\mathbf{x})$$

is conserved. □

1.3.5. Routh's method. We explain next a technique that can sometimes be invoked to reduce the number of unknowns in an Euler-Lagrange system of ODE.

The simplest case is $m = 2$, for which the unknown is $\mathbf{y} = [y_1\, y_2]^T$. The basic idea is that if the Lagrangian

$$L = L(x, y, z) = L(x, y_1, z_1, z_2)$$

does not depend upon y_2, we can then convert the full (E-L) system

$$\begin{cases} -\left(\frac{\partial L}{\partial z_1}(x, \mathbf{y}, \mathbf{y}')\right)' + \frac{\partial L}{\partial y_1}(x, \mathbf{y}, \mathbf{y}') = 0 \\ -\left(\frac{\partial L}{\partial z_2}(x, \mathbf{y}, \mathbf{y}')\right)' = 0 \end{cases} \tag{1.29}$$

into a single ODE for the single unknown y_1.

To do this, first observe that the second Euler-Lagrange equation in (1.29) implies

$$\frac{\partial L}{\partial z_2}(x, y_1, y_1', y_2') = C \tag{1.30}$$

for some constant C. We assume next that we can rewrite the algebraic identity

$$\frac{\partial L}{\partial z_2}(x, y_1, z_1, z_2) = C$$

to solve for z_2:

$$z_2 = \phi(x, y_1, z_1, C).$$

Thus

$$y_2' = \phi(x, y_1, y_1', C). \tag{1.31}$$

DEFINITION. Routh's function is

$$\boxed{R(x, y_1, z_1) = L(x, y_1, z_1, \phi(x, y_1, z_1, C)) - C\phi(x, y_1, z_1, C).} \tag{1.32}$$

THEOREM 1.3.9. Assume that $\mathbf{y} = [y_1\, y_2]^T$ solves the (E-L) system (1.29) and that the conservation law (1.30) holds.

Then y_1 solves the single (E-L) equation determined by Routh's function:

$$\boxed{-\left(\frac{\partial R}{\partial z_1}(x, y_1, y_1')\right)' + \frac{\partial R}{\partial y_1}(x, y_1, y_1') = 0.} \tag{1.33}$$

REMARK. And so if we can solve the ODE (1.33) for the unknown function y_1, we can then recover y_2 by integrating (1.31). □

Proof. We calculate

$$\frac{\partial R}{\partial x} = \frac{\partial L}{\partial x} + \left(\frac{\partial L}{\partial z_2} - C\right)\frac{\partial \phi}{\partial x}$$

and

$$\frac{\partial R}{\partial z_1} = \frac{\partial L}{\partial z_1} + \left(\frac{\partial L}{\partial z_2} - C\right)\frac{\partial \phi}{\partial z_1}.$$

Hence (1.31) and (1.30) imply

$$\frac{\partial R}{\partial x}(x, y_1, y_1') = \frac{\partial L}{\partial x}(x, y_1, y_1', y_2').$$

and

$$\frac{\partial R}{\partial z_1}(x, y_1, y_1') = \frac{\partial L}{\partial z_1}(x, y_1, y_1', y_2').$$

Then the first equation in (1.29) lets us compute that

$$\begin{aligned} -\left(\frac{\partial R}{\partial z_1}(x, y_1, y_1')\right)' &+ \frac{\partial R}{\partial y_1}(x, y_1, y_1') \\ &= -\left(\frac{\partial L}{\partial z_1}(x, \mathbf{y}, \mathbf{y}')\right)' + \frac{\partial L}{\partial y_1}(x, \mathbf{y}, \mathbf{y}') = 0. \end{aligned}$$

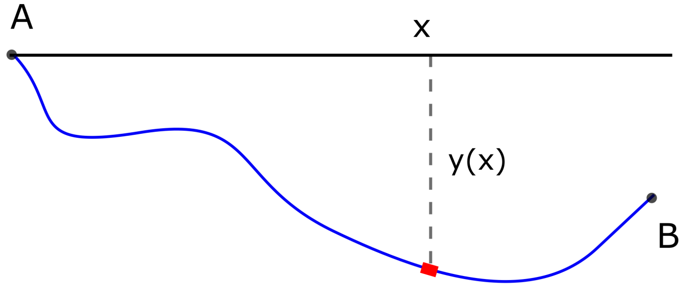
□

1.4. Applications

Following are some more substantial, and more interesting, applications of our theory.

1.4.1. Brachistochrone.

Given two points A, B as drawn, we can interpret the graph of a function $y(\cdot)$ joining these points as a wire path along which a bead of unit mass slides without friction under the influence of gravity. How do we design the slide so as to minimize the time it takes for the bead to slide from A to B?



For simplicity, we assume that $A = (0,0)$ and that $y(x) \leq 0$ for all $0 \leq x \leq b$. As the particle slides its total energy (= kinetic energy + potential energy) is constant. Therefore

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

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

$$v = (-2gy)^{\frac{1}{2}}.$$

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

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

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

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

Now

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

and consequently

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

Since L does not depend on x, it follows from Theorem 1.3.2 that

$$y'\frac{\partial L}{\partial z}(y, y') - L(y, y')$$

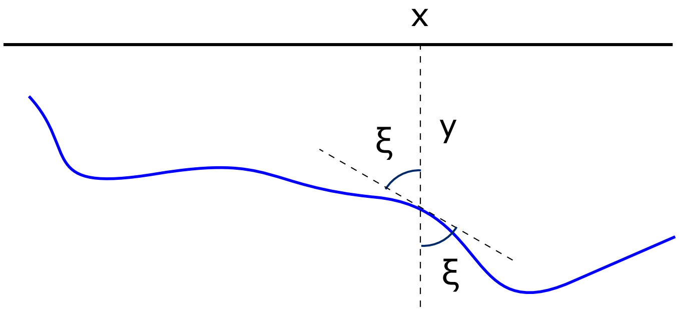
is constant. Therefore

$$y(1 + (y')^2) = C \tag{1.34}$$

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

GEOMETRIC INTERPRETATION. It is possible to directly integrate the ODE (1.34) (see Kot [**K**]), but the following geometric insights are more interesting. We first check if the graph of $y(\cdot)$ is the blue curve drawn below, the angle ξ satisfies

$$\sin \xi = \frac{1}{(1 + (y')^2)^{\frac{1}{2}}}.$$

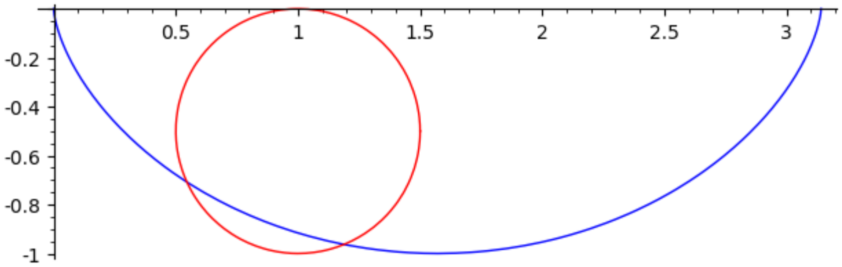


Angles and derivatives

Hence the ODE (1.34) says geometrically that

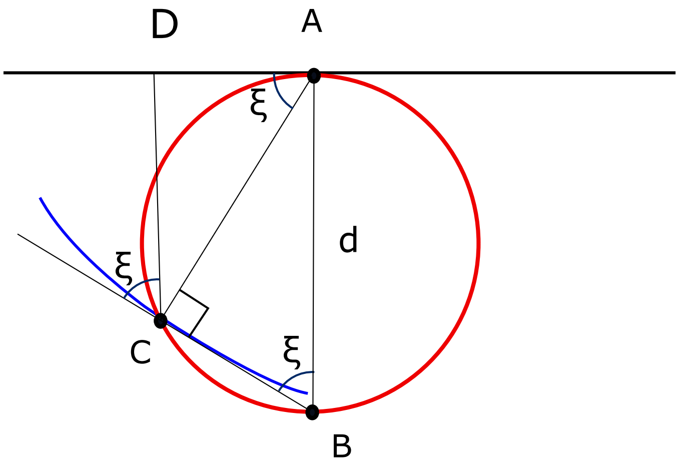
$$\frac{\sin \xi}{(-y)^{\frac{1}{2}}} \quad \text{is constant;} \tag{1.35}$$

and, according to the Remark on page 14, this in turn implies $y(\cdot)$ solves the full Euler-Lagrange equation. (Compare all this with the geometric optics example on page 17.)



A cycloid

Now (1.35) turns out to imply that the brachistochrone path is along a **cycloid**, the curve traced by a point on the rim of a circle as it rolls horizontally. Levi [**L**, pages 190–192] and Melzak [**M**, page 96] provide the following elegant geometric proof. 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 contract of the circle with the line, then the vector AC is perpendicular to the velocity vector $\mathbf{v}$.



Geometry of brachistochrone

Thus **v**, which is tangent to the blue cycloid curve, is parallel to CB, B denoting the point directly opposite from A on the circle. Consequently

$$|AB| = d.$$

Elementary geometry shows for the angles ξ as drawn that $|AC| = d \sin \xi$ and

$$-y = |DC| = |AC| \sin \xi = d \sin^2 \xi.$$

This gives (1.35). □

REMARK. See also Levi [**L**, page 55] for an argument showing that the cycloid is a **tautochrone**. This means that if we release from rest two beads from different locations along the cycloid wire curve, they arrive at the lowest point at the same time. □

1.4.2. Terrestrial brachistochrone.

A modern variant of the classical brachistochrone problem asks us the find a curve $r = r(\theta)$ in polar coordinates, describing a tunnel through the earth within which a (frictionless) vehicle falls and thereby travels in the least time between points on the surface.

As derived in Smith [**S**, pages 131-133] and Kot [**K**, pages 6-7], the transit time along a given curve is

$$I[r(\cdot)] = \left(\frac{R}{g}\right)^{\frac{1}{2}} \int_{\theta_a}^{\theta_b} \left(\frac{(r')^2 + r^2}{R^2 - r^2}\right)^{\frac{1}{2}} d\theta,$$

where R is the radius of the earth and g is the gravitational acceleration at the surface.

Since the Lagrangian

$$L(r,s) = \left(\frac{s^2+r^2}{R^2-r^2}\right)^{\frac{1}{2}}$$

does not depend on θ, we know that for any minimizing curve the expression

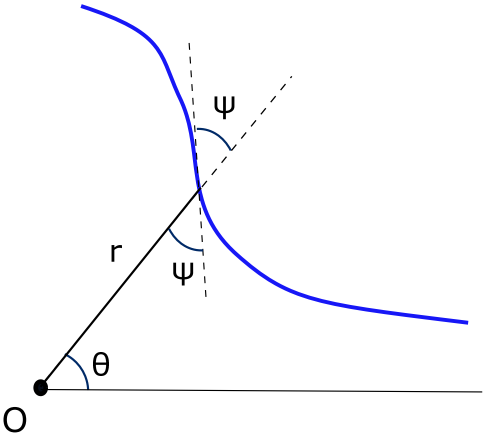
$$r'\frac{\partial L}{\partial s}(r,r') - L(r,r')$$

is constant. We compute this and simplify, to find for some constant C that

$$\left(\frac{r'}{r}\right)^2 + 1 = C\frac{r^2}{R^2-r^2} \tag{1.36}$$

GEOMETRIC INTERPRETATION. If ψ denotes, as drawn, the angle between the path traced by $r = r(\theta)$ and the radial vector r, we have

$$r' = r\cot\psi.$$

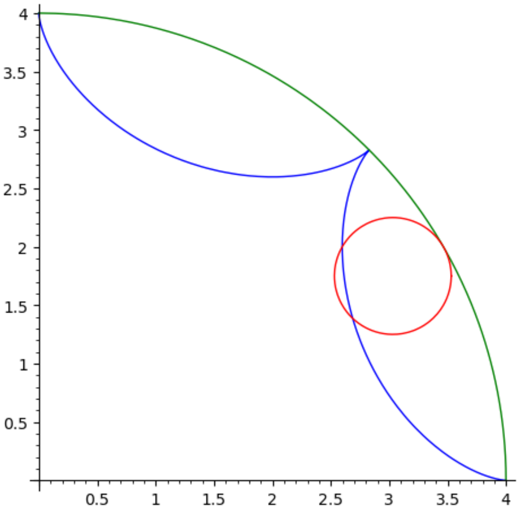


Angles and derivatives in polar coordinate

Hence the ODE (1.36) says geometrically that

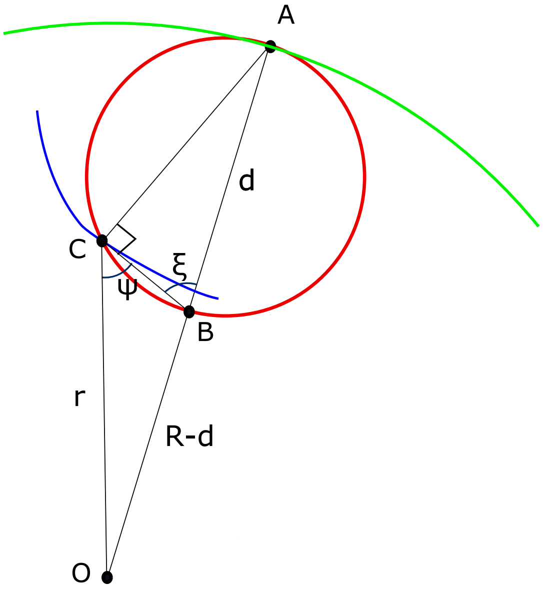
$$\sin\psi\left(\frac{R^2}{r^2}-1\right)^{-\frac{1}{2}} \text{ is constant.} \tag{1.37}$$

We now modify the geometric reasoning from the conventional brachistochrone. A circle of radius d rolls along the inside of a circle with center O and radius $R > d$, and a point C on the smaller circle sweeps out a **hypocycloid**.



Hypocycloid

If A is the instantaneous point of contract of the smaller circle with the larger, then the vector AC is perpendicular to the velocity vector $\mathbf{v}$.



Geometry of terrestrial brachistochrone

Therefore the point B as drawn is directly opposite from A, with $|AB| = d$. We apply the Law of Cosines to the triangle ACO, to deduce that

$$r^2 = d^2 \sin^2 \xi + R^2 - 2dR \sin \xi \cos(\tfrac{\pi}{2} - \xi).$$

Thus

$$r^2 = \sin^2 \xi (d^2 - 2Rd) + R^2. \tag{1.38}$$

The Law of Sines applied to the triangle BCO tells us as well that

$$\frac{\sin \psi}{R - d} = \frac{\sin(\pi - \xi)}{r} = \frac{\sin \xi}{r}.$$

Use this identity in (1.38), to learn that

$$\frac{R^2}{r^2} - 1 = \sin^2 \psi \, \frac{2dR - d^2}{(R - d)^2}.$$

Therefore the hypocycloid traced out by the point C satisfies the geometric form (1.37) of the terrestrial brachistochrone equation. □

1.4.3. Lagrangian and Hamiltonian dynamics.

In this section we explain how to rewrite certain Euler-Lagrange equations into a new formulation. For this, we assume the independent variable is time t, and the Lagrangian is $L = L(x, v)$. In this notation Euler-Lagrange equations are

$$-\frac{d}{dt}\left(\nabla_v L(\mathbf{x}, \dot{\mathbf{x}})\right) + \nabla_x L(\mathbf{x}, \dot{\mathbf{x}}) = 0, \tag{E-L}$$

where

$$\nabla_x L = \begin{bmatrix} \frac{\partial L}{\partial x_1} \\ \vdots \\ \frac{\partial L}{\partial x_n} \end{bmatrix}, \quad \nabla_v L = \begin{bmatrix} \frac{\partial L}{\partial v_1} \\ \vdots \\ \frac{\partial L}{\partial v_n} \end{bmatrix}.$$

Hamilton's equations. We propose to convert (E-L) into a system of $2n$ first-order ODE, having the elegant form

$$\boxed{\begin{cases} \dot{\mathbf{x}} = \nabla_p H(\mathbf{x}, \mathbf{p}) \\ \dot{\mathbf{p}} = -\nabla_x H(\mathbf{x}, \mathbf{p}). \end{cases}} \tag{H}$$

These equations involve a new function $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $H = H(x, p)$, called the **Hamiltonian**, that we will define below. We will write

$$\nabla_x H = \begin{bmatrix} \frac{\partial H}{\partial x_1} \\ \vdots \\ \frac{\partial H}{\partial x_n} \end{bmatrix}, \quad \nabla_p H = \begin{bmatrix} \frac{\partial H}{\partial p_1} \\ \vdots \\ \frac{\partial H}{\partial p_n} \end{bmatrix}.$$

The unknowns in (H) are the two functions $\mathbf{x}, \mathbf{p} : [0, \infty) \to \mathbb{R}^n$, where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}.$$

Assume hereafter that $\mathbf{x}(\cdot)$ solves (E-L) for all times $t \geq 0$.

DEFINITION. Set

$$\boxed{\mathbf{p}(t) = \nabla_v L(\mathbf{x}(t), \dot{\mathbf{x}}(t))} \qquad (t \geq 0).$$

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

The Hamiltonian. We will need the following hypothesis:

$$(1.39) \qquad \begin{cases} \text{Assume for all } x, p \in \mathbb{R}^n \text{ that the equation} \\ \qquad\qquad p = \nabla_v L(x, v) \\ \text{can be uniquely solved for } v \text{ as a function of } x \text{ and } p\text{:} \\ \qquad\qquad v = \boldsymbol{\phi}(x, p). \end{cases}$$

We consequently have the identity

$$(1.40) \qquad \nabla_v L(x, \boldsymbol{\phi}(x, p)) = p \quad (x, p \in \mathbb{R}^n).$$

DEFINITION. The **Hamiltonian** H corresponding to the Lagrangian L is

$$(1.41) \qquad \boxed{H(x, p) = p \cdot \boldsymbol{\phi}(x, p) - L(x, \boldsymbol{\phi}(x, p))} \quad (x, p \in \mathbb{R}^n).$$

THEOREM 1.4.1.

(i) The functions $\mathbf{x}, \mathbf{p} : [0, \infty) \to \mathbb{R}^n$ solve **Hamilton's equations** (H).

(ii) Furthermore,

$$(1.42) \qquad H(\mathbf{x}, \mathbf{p}) \text{ is constant on } [0, \infty).$$

Proof. 1. We compute using (1.41) and (1.40) that

$$(1.43) \qquad \begin{aligned} \nabla_p H(x, p) &= \boldsymbol{\phi}(x, p) + (\nabla_p \boldsymbol{\phi})^T (x, p)(p - \nabla_v L(x, \boldsymbol{\phi}(x, p))) \\ &= \boldsymbol{\phi}(x, p) \end{aligned}$$

and

$$(1.44) \qquad \begin{aligned} \nabla_x H(x, p) &= (\nabla_x \boldsymbol{\phi})^T (x, p)(p - \nabla_v L(x, \boldsymbol{\phi}(x, p))) - \nabla_x L(x, \boldsymbol{\phi}(x, p)) \\ &= -\nabla_x L(x, \boldsymbol{\phi}(x, p)). \end{aligned}$$

Put $x = \mathbf{x}$ and $p = \mathbf{p}$ into these formulas, and note that (1.39) implies

$$\dot{\mathbf{x}} = \boldsymbol{\phi}(\mathbf{x}, \mathbf{p}).$$

From (1.43), it follows that $\nabla_p H(\mathbf{x}, \mathbf{p}) = \dot{\mathbf{x}}$. And (1.44) gives

$$\begin{aligned} \nabla_x H(\mathbf{x}, \mathbf{p}) &= -\nabla_x L(\mathbf{x}, \dot{\mathbf{x}}) \\ &= -\frac{d}{dt}\left(\nabla_v L(\mathbf{x}, \dot{\mathbf{x}})\right) \quad \text{according to (E-L)} \\ &= -\dot{\mathbf{p}}. \end{aligned}$$

2. To see that $H(\mathbf{x}, \mathbf{p})$ is constant in time, compute

$$\begin{aligned} \frac{d}{dt} H(\mathbf{x}, \mathbf{p}) &= \nabla_x H(\mathbf{x}, \mathbf{p}) \cdot \dot{\mathbf{x}} + \nabla_p H(\mathbf{x}, \mathbf{p}) \cdot \dot{\mathbf{p}} \\ &= \nabla_x H(\mathbf{x}, \mathbf{p}) \cdot \nabla_p H(\mathbf{x}, \mathbf{p}) - \nabla_p H(\mathbf{x}, \mathbf{p}) \cdot \nabla_x H(\mathbf{x}, \mathbf{p}) \\ &= 0. \end{aligned}$$

□

REMARK. (Lagrangians, Hamiltonians and convex duality) If we assume for each $x \in \mathbb{R}^n$ that

$$v \mapsto L(x, v) \quad \text{is uniformly convex}$$

and that the superlinear growth condition

$$\lim_{|v| \to \infty} \frac{L(x, v)}{|v|} = \infty$$

holds, then the Hamiltonian is the **dual convex function**

$$\boxed{H(x, p) = \max_{v \in \mathbb{R}^n} \{x \cdot v - L(x, v)\},} \tag{1.45}$$

the maximum occurring for $v = \boldsymbol{\phi}(x, p)$. (See the Math 170 notes for more about convex duality.) □

EXAMPLE. (Motion within a magnetic field) Let $\mathbf{B} : \mathbb{R}^3 \to \mathbb{R}^3$ denote a time-independent magnetic field. A charged particle within this magnetic field moves according to the **Lorentz equation**

$$m\ddot{\mathbf{x}} = q(\dot{\mathbf{x}} \times \mathbf{B}(\mathbf{x})), \tag{1.46}$$

in which m is the mass of the particle and q is its charge.

We now show that this equation follows from Hamilton's equations (H) for

$$H = \frac{1}{2m}|p - q\mathbf{A}(x)|^2, \tag{1.47}$$

where the **magnetic potential** field $\mathbf{A}$ satisfies

$$\nabla \times \mathbf{A} = \mathbf{B}.$$

We compute

$$\nabla_p H = \frac{p - q\mathbf{A}(x)}{m}, \quad \nabla_x H = -\frac{q(\nabla \mathbf{A}(x))^T(p - q\mathbf{A}(x))}{m},$$

since $\nabla\left(|\mathbf{A}|^2\right) = 2(\nabla\mathbf{A})^T\mathbf{A}$. So Hamilton's equations read

$$\begin{cases} \dot{\mathbf{x}} = \frac{\mathbf{p} - q\mathbf{A}(\mathbf{x})}{m} \\ \dot{\mathbf{p}} = \frac{q(\nabla\mathbf{A}(\mathbf{x}))^T(\mathbf{p} - q\mathbf{A}(\mathbf{x}))}{m}. \end{cases}$$

We now show that these imply the Lorentz equation (1.46), by computing

$$\begin{aligned} m\ddot{\mathbf{x}} &= \dot{\mathbf{p}} - q\nabla\mathbf{A}\dot{\mathbf{x}} \\ &= \frac{q(\nabla\mathbf{A})^T(\mathbf{p} - q\mathbf{A})}{m} - q\nabla\mathbf{A}(\mathbf{x})\dot{\mathbf{x}} \\ &= q((\nabla\mathbf{A})^T - \nabla\mathbf{A})\dot{\mathbf{x}} \\ &= q(\dot{\mathbf{x}} \times (\nabla \times \mathbf{A})) \\ &= q(\dot{\mathbf{x}} \times \mathbf{B}). \end{aligned}$$

In this calculation we employed the vector calculus rule

$$(\nabla\mathbf{g} - (\nabla\mathbf{g})^T)y = (\nabla \times \mathbf{g}) \times y$$

for $y \in \mathbb{R}^3$ and $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3$.

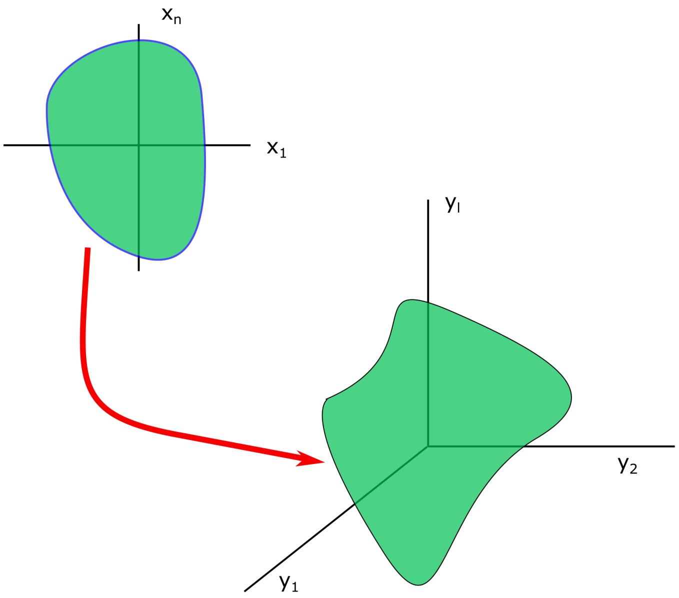
Taylor [**T**] is a good text for more on physical applications of variational principles and Hamilton's equations. □

1.4.4. Geodesics.

Let $U \subseteq \mathbb{R}^n$ be an open region. Assume that we are given a function $\mathbf{y} : U \to \mathbb{R}^l$, which we write as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_l \end{bmatrix}.$$

We call $\mathbf{y}$ a **coordinate patch**.



DEFINITION. The **metric tensor** G is the $n \times n$ symmetric matrix function whose entries are

$$g_{ij} = \frac{\partial \mathbf{y}}{\partial x_i} \cdot \frac{\partial \mathbf{y}}{\partial x_j} \qquad (i, j = 1, \dots, n).$$

We assume G is everywhere positive definite: $G \succ 0$.

NOTATION. The matrix G is therefore invertible. We will write

$$g^{ij}$$

for the (i, j)-th entry of the inverse matrix G^{-1}.

DEFINITION. The corresponding **Christoffel symbols** are

$$\boxed{\Gamma_{ij}^m = \frac{1}{2} \sum_{k=1}^n g^{mk} \left(\frac{\partial g_{ik}}{\partial x_j} + \frac{\partial g_{jk}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_k} \right).} \tag{1.48}$$

DEFINITION. The **energy** of a curve $\mathbf{x} : [0, T] \to U$ is

$$E[\mathbf{x}(\cdot)] = \frac{1}{2} \int_0^T \sum_{i,j=1}^n g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j \, dt.$$

THEOREM 1.4.2.

(i) The Euler-Lagrange equations for the energy $E[\,\cdot\,]$ are

$$\boxed{\ddot{x}_m + \sum_{i,j=1}^{n} \Gamma_{ij}^{m}(\mathbf{x})\dot{x}_i\dot{x}_j = 0} \qquad (m = 1, \dots, n). \tag{1.49}$$

(ii) If $\mathbf{x}$ solves (1.49), then

$$\sum_{i,j=1}^{n} g_{ij}(\mathbf{x})\dot{x}_i\dot{x}_j \ \text{ is constant.} \tag{1.50}$$

DEFINITION. A curve $\mathbf{x}(\cdot)$ solving the system of ODE (1.49) is called a **geodesic**. We will see later that (1.50) says *geodesics have constant speed.*

□

Proof. 1. The Lagrangian is

$$L = L(x, v) = \frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(x)v_i v_j,$$

with

$$\frac{\partial L}{\partial v_k} = \sum_{i=1}^{n} g_{ik}v_i, \qquad \frac{\partial L}{\partial x_k} = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial g_{ij}}{\partial x_k} v_i v_j.$$

We insert these into the Euler-Lagrange equation $-\frac{d}{dt}\left(\frac{\partial L}{\partial v_k}\right) + \frac{\partial L}{\partial x_k} = 0$, to find

$$\frac{d}{dt}\left(\sum_{i=1}^{n} g_{ik}\dot{x}_i\right) - \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial g_{ij}}{\partial x_k}\dot{x}_i\dot{x}_j = 0.$$

Therefore

$$\begin{aligned} 0 &= \sum_{i=1}^{n} g_{ik}\ddot{x}_i + \sum_{i,j=1}^{n}\left(\frac{\partial g_{ik}}{\partial x_j} - \frac{1}{2}\frac{\partial g_{ij}}{\partial x_k}\right)\dot{x}_i\dot{x}_j \\ &= \sum_{i=1}^{n} g_{ik}\ddot{x}_i + \frac{1}{2}\sum_{i,j=1}^{n}\left(\frac{\partial g_{ik}}{\partial x_j} + \frac{\partial g_{jk}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_k}\right)\dot{x}_i\dot{x}_j. \end{aligned}$$

Multiply by g^{mk}, sum on m, and recall $\sum_{k=1}^{n} g^{mk}g_{ki} = \delta_{mi}$, to deduce

$$\ddot{x}_m + \frac{1}{2}\sum_{i,j,k=1}^{n} g^{mk}\left(\frac{\partial g_{ik}}{\partial x_j} + \frac{\partial g_{jk}}{\partial x_i} - \sum_{i,j=1}^{n}\frac{\partial g_{ij}}{\partial x_k}\right)\dot{x}_i\dot{x}_j = 0.$$

We recall the definition (1.48) of the Christoffel symbols to complete the derivation of (1.49).

2. Since the Lagrangian L does not depend upon the independent variable t, Theorem 1.3.8 tells us that the expression

$$\dot{\mathbf{x}} \cdot \nabla_z L(\mathbf{x}, \dot{\mathbf{x}}) - L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j$$

is constant for times $0 \leq t \leq T$. □

Length and energy. We discuss next how minimizing the energy is equivalent to minimizing length. We henceforth take $T = 1$ in the definiton of the energy.

DEFINITIONS.

(i) The **length** of a curve $\mathbf{x} : [0, 1] \to U$ is

$$L[\mathbf{x}(\cdot)] = \int_0^1 \Big(\sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j \Big)^{\frac{1}{2}} \, dt.$$

(This is the Euclidean length of the image of $\mathbf{x}(\cdot)$ under the coordinate chart $\mathbf{y}(\cdot)$.)

(ii) The **distance** between two points $A, B \in R^n$ in the metric determined by G is

$$\operatorname{dist}(A, B) = \min \left\{ L[\mathbf{x}] \mid \mathbf{x} : [0, 1] \to \mathbb{R}^n, \mathbf{x}(0) = A, \mathbf{x}(1) = B \right\}.$$

□

It turns out that minimizing energy is equivalent to minimizing length:

THEOREM 1.4.3. A curve that minimizes the energy among paths joining A and B for $0 \leq t \leq 1$ if and only if it also minimizes the length.

Furthermore

$$\min \left\{ E[\mathbf{x}] \mid \mathbf{x} : [0, 1] \to \mathbb{R}^n, \mathbf{x}(0) = A, \mathbf{x}(1) = B \right\} = \frac{\operatorname{dist}(A, B)^2}{2}. \tag{1.51}$$

Proof. Let us assume that a curve $\mathbf{x} = [x_1 \cdots x_n]^T$ joining A and B gives the distance:

$$\operatorname{dist}(A, B) = \int_0^1 \Big(\sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j \Big)^{\frac{1}{2}} \, dt.$$

We can if necessary reparameterize $\mathbf{x}(\cdot)$ to have constant speed, so that

$$\Big(\sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j \Big)^{\frac{1}{2}} = \operatorname{dist}(A, B) \quad (0 \leq t \leq 1). \tag{1.52}$$

Next, assume that a curve $\mathbf{y}(\cdot)$ minimizes the energy among paths connecting A to B:

$$E[\mathbf{y}] = \min\left\{E[\mathbf{w}] \mid \mathbf{w} : [0,1] \to \mathbb{R}^n, \mathbf{w}(0) = A, \mathbf{w}(1) = B\right\}.$$

According to Theorem 1.4.2,

$$\frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(\mathbf{y})\dot{y}_i\dot{y}_j = E \tag{1.53}$$

is constant in time. Then (1.52) implies

$$E = E[\mathbf{y}] \leq E[\mathbf{x}] = \frac{1}{2}\int_0^1 \sum_{i,j=1}^{n} g_{ij}(\mathbf{x})\dot{x}_i\dot{x}_j\,dt = \frac{\operatorname{dist}(A,B)^2}{2}.$$

But also

$$\begin{aligned}\operatorname{dist}(A,B) = L[\mathbf{x}] \leq L[\mathbf{y}] &= \int_0^1 \Big(\sum_{i,j=1}^{n} g_{ij}(\mathbf{y})\dot{y}_i\dot{y}_j\Big)^{\frac{1}{2}}\,dt \\ &\leq \left(\int_0^1 \sum_{i,j=1}^{n} g_{ij}(\mathbf{y})\dot{y}_i\dot{y}_j\,dt\right)^{\frac{1}{2}} = (2E)^{\frac{1}{2}},\end{aligned}$$

according to (1.53). Hence $E = \frac{\operatorname{dist}(A,B)^2}{2}$; and therefore $L[\mathbf{x}] = L[\mathbf{y}]$, $E[\mathbf{y}] = E[\mathbf{x}]$. □

EXAMPLE (**Hyperbolic metric**). A famous model in geometry is the **Poincaré hyperbolic plane**, for which $n = 2$ and

$$g_{ij} = \frac{1}{(x_2)^2}\delta_{ij} \qquad (i \leq i, j \leq 2)$$

in the region $\mathbb{H} = \{-\infty < x_1 < \infty,\ 0 < x_2 < \infty\}$.

We calculate using the definition (1.48) that

$$\Gamma_{12}^1 = \Gamma_{21}^1 = \Gamma_{22}^2 = -\frac{1}{x_2}, \quad \Gamma_{11}^2 = \frac{1}{x_2},$$

and the remaining Christoffel symbols vanish. Employing these formulas in (1.49) yields this system of geodesic ODE for the hyperbolic plane:

$$\begin{cases}\ddot{x}_1 = \frac{2\dot{x}_1\dot{x}_2}{x_2} \\ \ddot{x}_2 = \frac{(\dot{x}_2)^2 - (\dot{x}_1)^2}{x_2}.\end{cases} \tag{1.54}$$

We can extract geometric information from these equations:

THEOREM 1.4.4. The path of any trajectory solving (1.54) is either along a vertical line or else along a circle centered on the x_1-axis.

Proof. 1. Consider a solution of the ODE system (1.54) with $\dot{x}_1 \neq 0$. Define

$$a = x_1 + \frac{x_2 \dot{x}_2}{\dot{x}_1}. \tag{1.55}$$

Then

$$\begin{aligned}
\dot{a} &= \dot{x}_1 + \frac{\dot{x}_2\dot{x}_2 + x_2\ddot{x}_2}{\dot{x}_1} - \frac{x_2\dot{x}_2\ddot{x}_1}{(\dot{x}_1)^2} \\
&= \dot{x}_1 + \frac{(\dot{x}_2)^2}{\dot{x}_1} + \frac{x_2}{\dot{x}_1}\left(\frac{(\dot{x}_2)^2 - (\dot{x}_1)^2}{x_2}\right) - \frac{x_2\dot{x}_2}{(\dot{x}_1)^2}\left(\frac{2\dot{x}_1\dot{x}_2}{x_2}\right) \\
&= 0;
\end{aligned}$$

consequently a is constant.

2. We claim that the motion of the point $\mathbf{x} = [x_1\, x_2]^T$ lies within a circle with center $(a, 0)$. To confirm this, let us use (1.55) to calculate that

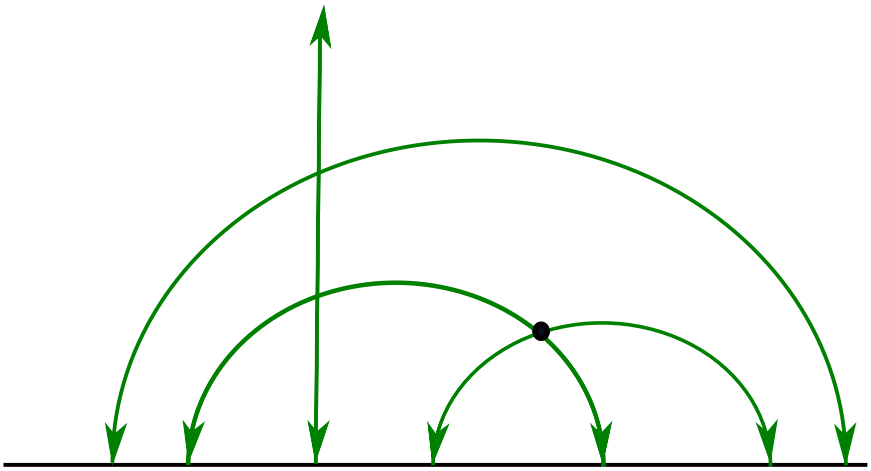
$$\begin{aligned}
\frac{d}{ds}\left\{\frac{(x_1 - a)^2 + (x_2)^2}{2}\right\} &= (x_1 - a)\dot{x}_1 + x_2\dot{x}_2 \\
&= \left(-\frac{x_2\dot{x}_2}{\dot{x}_1}\right)\dot{x}_1 + x_2\dot{x}_2 \\
&= 0.
\end{aligned}$$

Therefore

$$(x_1 - a)^2 + x_2^2 = r^2$$

for some appropriate radius $r > 0$. □

GEOMETRIC INTERPRETATION. Thus the geodesics in the hyperbolic half plane are either vertical, or else approach the x_1-axis as $s \to \pm\infty$. The half circles they traverse have infinite length. qed



Geodesics in the hyperbolic plane

□

1.4.5. Maxwell's fisheye.

This example concerns curves $\mathbf{x}(\cdot)$ taking values in $\mathbb{R}^3$ that are extremals for the Lagrangian

$$L(x,v) = \frac{|v|}{1+|x|^2} \qquad (x, v \in \mathbb{R}^3). \tag{1.56}$$

Then

$$\nabla_v L = \frac{v}{|v|(1+|x|^2)}, \quad \nabla_x L = -\frac{2|v|x}{(1+|x|^2)^2}.$$

The (E-L) system is therefore

$$-\frac{d}{dt}\left(\frac{\dot{\mathbf{x}}}{|\dot{\mathbf{x}}|(1+|\mathbf{x}|^2)}\right) - \frac{2|\dot{\mathbf{x}}|\mathbf{x}}{(1+|\mathbf{x}|^2)^2} = 0. \tag{1.57}$$

where $\dot{} = \frac{d}{dt}$. We reparameterize in terms of the arclength $s = s(t)$, which satisfies

$$\frac{ds}{dt} = |\dot{\mathbf{x}}|.$$

Then (1.57) becomes

$$\left(\frac{\mathbf{x}'}{1+|\mathbf{x}|^2}\right)' = -\frac{2\mathbf{x}}{(1+|\mathbf{x}|^2)^2}. \tag{1.58}$$

where $' = \frac{d}{ds}$.

We will now investigate properties of solutions of the system (1.58), with $|\mathbf{x}'| \equiv 1$.

THEOREM 1.4.5. Any trajectory solving (1.58) is either along a line through the origin, or else along a circle in a plane passing through the origin.

GEOMETRIC INTERPRETATION. The proof below shows how we can sometimes deduce interesting geometric information about solutions of a system of ODE.

In particular, the Lagrangian (1.56) has the remarkable property that all solutions of the corresponding Euler-Lagrange system of ODE move in circles (or along straight lines). And even though L is radial in x, the centers of these circles need not be the origin.

□

Proof. 1. We compute using the cross product and the ODE (1.58) that

$$\left(\frac{\mathbf{x}'}{1+|\mathbf{x}|^2} \times \mathbf{x}\right)' = \left(\frac{\mathbf{x}'}{1+|\mathbf{x}|^2}\right)' \times \mathbf{x} + \left(\frac{\mathbf{x}'}{1+|\mathbf{x}|^2} \times \mathbf{x}'\right)$$

$$= -\frac{2\mathbf{x}}{(1+|\mathbf{x}|^2)^2} \times \mathbf{x} = 0.$$

Hence

$$\frac{\mathbf{x}'}{1+|\mathbf{x}|^2} \times \mathbf{x} = b \qquad \text{for some vector } b \in \mathbb{R}^3.$$

If $b \neq 0$, then since $\mathbf{x} \cdot b = 0$, the trajectory lies in the plane through the origin perpendicular to b. If $b = 0$, then $\mathbf{x}$ and $\mathbf{x}'$ are everywhere parallel and so the motion is along a line.

2. We henceforth assume $b \neq 0$. Upon rotating coordinates if necessary, we may assume that b is parallel to $[0\,0\,1]^T$. Thus we covert to the two-dimensional case that $\mathbf{x} = [x_1\, x_2]^T$ solves the (now two-dimensional) system of ODE (1.58). Carrying out the differentiation in the term on the left hand side of (1.58), we can now write (E-L) as

$$\mathbf{x}'' = \frac{2}{1+|\mathbf{x}|^2}((\mathbf{x}\cdot\mathbf{x}')\mathbf{x}' - \mathbf{x}). \tag{1.59}$$

Let $\mathbf{t} = \mathbf{x}'$ be the unit tangent vector to the curve in the plane. Then

$$\mathbf{t}' = \kappa\mathbf{n} \tag{1.60}$$

where $\mathbf{n}$ is the unit normal and $\kappa \geq 0$ is the curvature. So $\{\mathbf{t}, \mathbf{n}\}$ is an orthonormal frame moving along the curve. If we differentiate the expression $\mathbf{t}\cdot\mathbf{n} = 0$, we see that

$$\mathbf{n}' = -\kappa\mathbf{t}. \tag{1.61}$$

Since $\mathbf{x} = (\mathbf{x}\cdot\mathbf{t})\mathbf{t} + (\mathbf{x}\cdot\mathbf{n})\mathbf{n}$, $\mathbf{x}' = \mathbf{t}$ and $\mathbf{x}'' = \mathbf{t}' = \kappa\mathbf{n}$, we deduce from (1.59) that

$$\kappa = -2\frac{\mathbf{x}\cdot\mathbf{n}}{1+|\mathbf{x}|^2}. \tag{1.62}$$

3. We will now show that κ is constant. Let us calculate using (1.61) that

$$\begin{aligned}
\kappa' &= -2\left(\frac{\mathbf{x}'\cdot\mathbf{n} + \mathbf{x}\cdot\mathbf{n}'}{1+|\mathbf{x}|^2} - \frac{(\mathbf{x}\cdot\mathbf{n})2\mathbf{x}\cdot\mathbf{x}'}{(1+|\mathbf{x}|^2)^2}\right)\\
&= -2\left(\frac{\mathbf{t}\cdot\mathbf{n} - \kappa\mathbf{x}\cdot\mathbf{t}}{1+|\mathbf{x}|^2} - \frac{(\mathbf{x}\cdot\mathbf{n})2\mathbf{x}\cdot\mathbf{t}}{(1+|\mathbf{x}|^2)^2}\right)\\
&= \frac{2(\mathbf{x}\cdot\mathbf{t})}{(1+|\mathbf{x}|^2)^2}\Big(\kappa(1+|\mathbf{x}|^2) + (\mathbf{x}\cdot\mathbf{n})2\Big)\\
&= 0,
\end{aligned}$$

the last equality following from (1.62).

4. Finally we show that the trajectory lies on a circle if $\kappa > 0$. Define

$$\mathbf{c} = \mathbf{x} + \frac{1}{\kappa}\mathbf{n}.$$

Then

$$\mathbf{c}' = \mathbf{x}' + \tfrac{1}{\kappa}\mathbf{n}' = \mathbf{t} - \tfrac{1}{\kappa}\kappa\mathbf{t} = 0,$$

and hence $\mathbf{c} \equiv c$ for some point $c \in \mathbb{R}^2$. Furthermore,

$$(|\mathbf{x} - c|^2)' = 2(\mathbf{x} - c) \cdot \mathbf{x}' = -\frac{2}{\kappa}\mathbf{n} \cdot \mathbf{t} = 0.$$

Consequently the trajectory moves along the circle of radius κ^{-1} and center c. □

PHYSICAL INTERPRETATION. In the 19th century, there was interest in the optical properties of the eyes of fishes, and in particular the question as to how such eyes, which are often quite flat, focus images. Maxwell had found the geometric properties of extremals for the Langrangian (1.56), and there was some thought that these may explain the optics of fish eyes.

In the 20th century R. Luneburg generalized Maxwell's ideas to design lenses with various interesting properties, made from transparent materials with radially varying refractive index. (The refractive index of an optical medium is $n = \frac{c}{v}$, where c is the speed of light in a vacuum and v is the speed within the medium.) □

Chapter 2

SECOND VARIATION

2.1. Computing the second variation

We return to our standard calculus of variations problem of characterizing minimizers $y_0(\cdot)$ of the functional

$$I[y(\cdot)] = \int_a^b L(x, y, y')\,dx$$

over the admissible class

$$\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid y(a) = y^0, y(b) = y^1\}.$$

We have so far examined in great detail the Euler-Lagrange equation, the derivation of which corresponds to taking the first variation. This chapter turns attention to the second variation.

2.1.1. Integral and pointwise versions.

THEOREM 2.1.1. Suppose $y_0(\cdot) \in \mathcal{A}$ is a minimizer.

(i) Then

$$\begin{aligned}(2.1)\qquad \int_a^b \frac{\partial^2 L}{\partial z^2}(x, y_0, y_0')(w')^2 \\ + 2\frac{\partial^2 L}{\partial y \partial z}(x, y_0, y_0')ww' + \frac{\partial^2 L}{\partial y^2}(x, y_0, y_0')w^2\,dx \geq 0\end{aligned}$$

for all $w : [a,b] \to \mathbb{R}$ with $w(a) = w(b) = 0$.

(ii) Furthermore,

$$\boxed{\frac{\partial^2 L}{\partial z^2}(x, y_0(x), y_0'(x)) \geq 0} \qquad (a \leq x \leq b). \tag{2.2}$$

REMARKS.

(i) We call the left hand side of (2.1) the **second variation** of $I[\,\cdot\,]$ about $y_0(\cdot)$, evaluated at $w(\cdot)$.

(ii) If the mapping

$$z \mapsto L(x, y_0(x), z) \quad \text{is convex,} \tag{2.3}$$

then (2.2) holds. This observation strongly suggests that the convexity of the Lagrangian L in the variable z will be a useful hypothesis if we try to find minimizers: see Section 2.4.2. □

Proof. 1. We extend our earlier first variation proof of the Euler-Lagrange equation. Select $w : [a, b] \to \mathbb{R}$, with $w(a) = w(b) = 0$, and define $y_\tau(x) = y_0(x) + \tau w(x)$ for $-1 \leq \tau \leq 1$. Then $y_\tau(\cdot) \in \mathcal{A}$, and so $\tau \mapsto i(\tau)$ has a minimum at $\tau = 0$ on the interval $-1 \leq \tau \leq 1$. Therefore

$$\frac{d^2 i}{d\tau^2}(0) \geq 0.$$

2. Differentiating twice with respect to τ, we find

$$\frac{d^2 i}{d\tau^2}(\tau) = \int_a^b \frac{\partial^2 L}{\partial y^2}(x, y_0 + \tau w, y_0' + \tau w')w^2$$
$$+ 2\frac{\partial^2 L}{\partial y \partial z}(x, y_0 + \tau w, y_0' + \tau w')ww' + \frac{\partial^2 L}{\partial z^2}(x, y_0 + \tau w, y_0' + \tau w')(w')^2\, dx.$$

Put $\tau = 0$ and recall $\frac{d^2 i}{d\tau^2}(0) \geq 0$, to prove (2.1).

3. We need to design appropriate functions w to extract pointwise information from (2.1). For this, define $\phi : \mathbb{R} \to \mathbb{R}$ by setting

$$\phi(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 2 - x & \text{if } 1 \leq x \leq 2 \end{cases}$$

on the interval $[0, 2]$ and then extend ϕ to be 2-periodic on all of $\mathbb{R}$. Thus ϕ is a "sawtooth function" with corners at the integers and $\phi' = \pm 1$ elsewhere.

Next let $\zeta : [a.b] \to \mathbb{R}$ be any continuously differentiable function with $\zeta(a) = \zeta(b) = 0$. Then for each $\varepsilon > 0$, define

$$w_\varepsilon(x) = \varepsilon \phi(\tfrac{x}{\varepsilon})\zeta(x).$$

Then for all but finitely many points, w_ε is differentiable:

$$w'_\varepsilon(x) = \phi'(\tfrac{x}{\varepsilon})\zeta(x) + \varepsilon\phi(\tfrac{x}{\varepsilon})\zeta'(x).$$

Note that the second term on the right is less than or equal to $A\varepsilon$ for some appropriate constant A.

We now plug in w_ε in place of w in (2.1). Then upon making some simple estimates, we learn that

$$\int_a^b \frac{\partial^2 L}{\partial z^2}(x, y_0, y_0')(\phi'(\tfrac{x}{\varepsilon}))^2\zeta^2\,dx + D\varepsilon \geq 0,$$

for some constant D. But

$$\left(\phi'(\tfrac{x}{\varepsilon})\right)^2 = 1$$

except at finitely many points (which have no effect on the integral). Therefore upon sending $\varepsilon \to 0$, we deduce that

$$\int_a^b \frac{\partial^2 L}{\partial z^2}(x, y_0, y_0')\zeta^2\,dx \geq 0$$

for all functions ζ as above. This implies (2.2). □

2.1.2. Weierstrass condition.

In this section we strengthen (2.2):

THEOREM 2.1.2. Suppose $y_0(\cdot) \in \mathcal{A}$ is a minimizer that is continuously differentiable.

Then for all $z \in \mathbb{R}$ and all points $a \leq x \leq b$,

$$\boxed{L(x, y_0(x), z) \geq L(x, y_0(x), y_0'(x)) + \frac{\partial L}{\partial z}(x, y_0(x), y_0'(x))(z - y_0'(x)).} \tag{2.4}$$

This is the **Weierstrass condition** for a minimizer.

GEOMETRIC INTERPRETATION. This inequality says that for fixed x and $y_0(x)$, the graph of the function

$$z \mapsto L(x, y_0(x), z)$$

lies above the tangent line at the point $z = y_0'(x)$. This of course is clear if $z \mapsto L(x, y_0(x), z)$ is convex.

Notice that (2.4) does not follow from any sort of local second variation argument, since z need not be close to $y_0'(x)$. Rather, the Weierstrass condition is a consequence of the global minimality of $y_0(\cdot)$. □

Proof. To simplify the exposition, we will assume for simplicity that $L = L(x,z)$ does not depend upon y.

1. Select any $z \in \mathbb{R}$ and $0 < \delta < 1$. Choose any point $a < x_0 < b$ and select $\varepsilon > 0$ so small that the interval $[x_0 - \delta\varepsilon, x_0 + \varepsilon]$ lies within $[a,b]$. Now define

$$y(x) = \begin{cases} y_0(x) & (a \le x \le x_0 - \delta\varepsilon) \\ l_1(x) & (x_0 - \delta\varepsilon \le x \le x_0) \\ l_2(x) & (x_0 \le x \le x_0 + \varepsilon) \\ y_0(x) & (x_0 + \varepsilon \le x \le b), \end{cases}$$

where l_1, l_2 are linear functions selected so that $y(\cdot)$ is continuous and

$$\begin{cases} l_1'(x) = z & (x_0 - \delta\varepsilon \le x \le x_0) \\ l_2'(x) = \frac{y_0(x_0+\varepsilon) - y_0(x_0-\delta\varepsilon)}{\varepsilon} - \delta z & (x_0 \le x \le x_0 + \varepsilon) \end{cases}$$

Then $y(\cdot) \in \mathcal{A}$ and thus

$$I[y_0(\cdot)] \le I[y(\cdot)]$$

Since $y(\cdot) = y_0(\cdot)$ outside the interval $[x_0 - \delta\varepsilon, x_0 + \varepsilon]$, it follows that

$$\begin{aligned} \int_{x_0-\delta\varepsilon}^{x_0+\varepsilon} L(x, y_0')\,dx &\le \int_{x_0-\delta\varepsilon}^{x_0+\varepsilon} L(x, y')\,dx \\ &= \int_{x_0-\delta\varepsilon}^{x_0} L(x,z)\,dx + \int_{x_0}^{x_0+\varepsilon} L\left(x, \tfrac{y_0(x_0+\varepsilon)-y_0(x_0-\delta\varepsilon)}{\varepsilon} - \delta z\right) dx. \end{aligned}$$

Then

$$\begin{aligned} \frac{1}{\delta\varepsilon}\int_{x_0-\delta\varepsilon}^{x_0} L(x, y_0')\,dx &\le \frac{1}{\delta\varepsilon}\int_{x_0-\delta\varepsilon}^{x_0} L(x,z)\,dx \\ &\quad + \frac{1}{\delta\varepsilon}\int_{x_0}^{x_0+\varepsilon} L\left(x, \tfrac{y_0(x_0+\varepsilon)-y_0(x_0-\delta\varepsilon)}{\varepsilon} - \delta z\right) - L(x, y_0')\,dx. \end{aligned} \tag{2.5}$$

2. We must examine carefully the integral on the right. Now

$$L\left(x, \tfrac{y_0(x_0+\varepsilon)-y_0(x_0-\delta\varepsilon)}{\varepsilon} - \delta z\right) - L(x, y_0') = \tfrac{\partial L}{\partial z}(x, y_0')a_\varepsilon + r_\varepsilon, \tag{2.6}$$

for

$$a_\varepsilon = \frac{y_0(x_0+\varepsilon) - y_0(x_0-\delta\varepsilon)}{\varepsilon} - \delta z - y_0'(x),$$

where the remainder term r_ε satisfies the estimate

$$|r_\varepsilon| \le A a_\varepsilon^2. \tag{2.7}$$

According to L'Hospital's Rule,

$$\lim_{\varepsilon \to 0} a_\varepsilon = (1+\delta)y_0'(x_0) - \delta z - y_0'(x). \tag{2.8}$$

Then (2.7) implies

$$\limsup_{\varepsilon\to 0} |r_\varepsilon| \le B(\delta^2 + (y_0'(x_0) - y_0'(x))^2). \tag{2.9}$$

for a suitable constant B.

3. Now send $\varepsilon \to 0$ in (2.5), using (2.6),(2.8) and (2.9):

$$L(x_0, y_0') \le L(x_0, z) + \tfrac{\partial L}{\partial z}(x_0, y_0')(y_0'(x_0) - z) + B\delta.$$

Finally, let $\delta \to 0$, to deduce

$$L(x_0, y_0') + \tfrac{\partial L}{\partial z}(x_0, y_0')(z - y_0'(x_0)) \le L(x_0, z);$$

this is (2.4) for $L = L(x, z)$. □

2.2. Positive second variation

Suppose now that $y(\cdot) \in \mathcal{A}$ is an extremal, and consequently solves the Euler-Lagrange equation

$$-\left(\frac{\partial L}{\partial z}(x, y, y')\right)' + \frac{\partial L}{\partial y}(x, y, y') = 0 \qquad (a \le x \le b). \tag{E-L}$$

We devote the rest of this chapter to the question:

When is $y(\cdot)$ in fact a local minimizer of $I[\,\cdot\,]$?

2.2.1. Riccati equation.

NOTATION. Let us write

$$\boxed{A = \frac{\partial^2 L}{\partial z^2}(x, y, y'),\ B = \frac{\partial^2 L}{\partial y \partial z}(x, y, y'),\ C = \frac{\partial^2 L}{\partial y^2}(x, y, y').} \tag{2.10}$$

Therefore the second variation of $I[\,\cdot\,]$ about $y(\cdot)$ is

$$\int_a^b A(w')^2 + 2Bww' + Cw^2\,dx \tag{2.11}$$

for $w : [a, b] \to \mathbb{R}$ with $w(a) = w(b) = 0$. □

We assume hereafter that

$$A > 0 \quad \text{on } [a, b]. \tag{2.12}$$

As a first approach towards understanding local minimizers, we introduce a new nonlinear equation:

DEFINITION. The **Riccati equation** associated with (E-L) and the extremal $y(\cdot)$ is the first-order ODE

$$\text{(R)} \qquad \boxed{q' = -\frac{(q-B)^2}{A} + C} \qquad (a \le x \le b),$$

for the functions A, B, C defined by (2.10). □

A remarkable observation is that if (R) has a solution, then the second variation is positive:

THEOREM 2.2.1. Assume that there exists a solution $q(\cdot)$ of the Riccati equation (R).

Then the second variation of $I[\,\cdot\,]$ around $y(\cdot)$ is positive:

$$\int_a^b A(w')^2 + 2Bww' + Cw^2\,dx > 0 \tag{2.13}$$

for all nonzero $w : [a,b] \to \mathbb{R}$ with $w(a) = w(b) = 0$.

Proof. We calculate that

$$\begin{aligned}\int_a^b A(w')^2 &+ 2Bww' + Cw^2\,dx \\ &= \int_a^b A(w')^2 + 2Bww' + \left(q' + \tfrac{(q-B)^2}{A}\right) w^2\,dx \\ &= \int_a^b A(w')^2 - 2(q-B)ww' + \tfrac{(q-B)^2}{A} w^2\,dx \\ &= \int_a^b A\left(w' - \tfrac{q-B}{A} w\right)^2 dt \ge 0.\end{aligned}$$

Furthermore, if $w' - \frac{q-B}{A}w = 0$ on the interval $[a,b]$ and $w(a) = 0$, then uniqueness of the solution of this ODE imples $w = 0$. □

2.2.2. Conjugate points.

We show next that we can construct a solution to (R) if we can find a positive solution of a related *linear* ODE.

DEFINITIONS. (i) The **linearization of the Euler-Lagrange equation** (E-L) about the extremal $y(\cdot)$ is the linear operator

$$\boxed{Ju = -(Au' + Bu)' + Bu' + Cu,} \tag{2.14}$$

defined for twice continuously differentiable functions $u : [a,b] \to \mathbb{R}$.

(ii) We call the ODE

$$\boxed{Ju = 0} \qquad (a \leq x \leq b)$$

Jacobi's equation. □

We can construct solutions of the Riccati equation from *positive* solutions of Jacobi's equation:

THEOREM 2.2.2. Suppose that

$$\begin{cases} Ju = 0 & (a \leq x \leq b) \\ u > 0 & (a \leq x \leq b). \end{cases} \tag{2.15}$$

Then

$$\boxed{q = A\frac{u'}{u} + B} \tag{2.16}$$

solves the Riccati equation (R).

Proof. We calculate

$$\begin{aligned} 0 &= -(Au' + Bu)' + Bu' + Cu \\ &= -(qu)' + Bu' + Cu \\ &= -q'u + (B - q)u' + Cu \\ &= -q'u - \frac{(q-B)^2}{A}u + Cu. \end{aligned}$$

Divide by $u > 0$ to derive (R). □

In light of the previous theorem, we need to understand when we can find a positive solution of the Jacobi equation on a given interval $[a, b]$. This depends upon whether or not the interval contains conjugate points:

DEFINITION. Select $a \in \mathbb{R}$. We say that a point $c > a$ is a **conjugate point** with respect to a if there exists a function $u : [a, c] \to \mathbb{R}$ such that

$$\begin{cases} Ju = 0 \qquad (a \leq x \leq c) \\ u(a) = u(c) = 0 \\ u \neq 0. \end{cases} \tag{2.17}$$

□

THEOREM 2.2.3. If there are no conjugate points within the interval $(a, b]$, then there exists a solution $q(\cdot)$ of the Riccati equation (R).

Consequently, the second variation of $I[\,\cdot\,]$ about $y(\cdot)$ is positive.

Proof. 1. It can be shown that if there are no conjugate points with respect to a within $(a,b]$, then in fact for some nearby point $d<a$ there are no conjugate points with respect to d within $(d,b]$. (The proof of this relies upon some mathematical ideas a bit beyond the scope of these notes.)

2. Consider then the initial-value problem

$$\begin{cases} Ju=0 & (d\leq x\leq b) \\ u(d)=0, & u'(d)=1. \end{cases} \tag{2.18}$$

Standard ODE theory tells us that this problem has a unique solution. We observe furthermore that

$$u(x)>0 \qquad (d<x\leq d+\varepsilon)$$

for some sufficiently small $\varepsilon>0$, since $u'(d)=1$.

We claim that in fact

$$u(x)>0 \qquad (d<x\leq b). \tag{2.19}$$

To see this, assume instead that $d<c\leq b$ is a point where $u(c)=0$. Then c would be a conjugate point with respect to d, and this is not possible.

Since by continuity u is strictly positive on the interval $[a,b]$, it follows from Theorem 2.2.2 that $q=A\frac{u'}{u}+B$ solves (R). Theorem 2.2.1 now implies that the second variation is positive. □

2.3. Strong local minimizers

We now build upon the ideas just developed. We next show that the absence of conjugate points implies not only that the second variation is positive, but also that $y(\cdot)$ is a local minimizer of $I[\,\cdot\,]$, in fact a *strong* local minimizer.

DEFINITION. We say $y:[a,b]\to\mathbb{R}$ is a **strong local minimizer** of $I[\,\cdot\,]$ if there exists $\delta>0$ such that

$$\max_{[a,b]}|y-\bar{y}|\leq\delta \tag{2.20}$$

implies

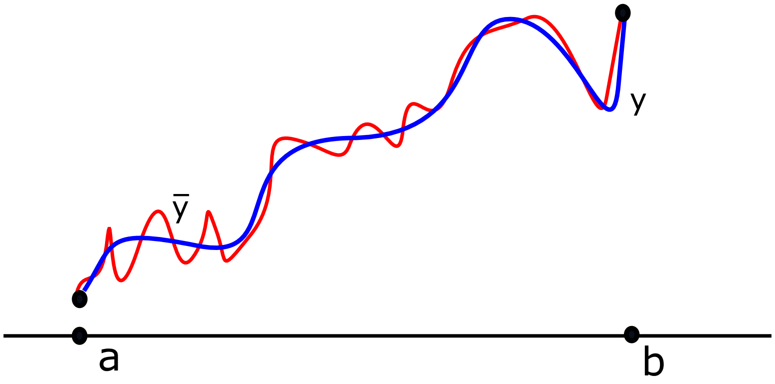
$$I[y(\cdot)]\leq I[\bar{y}(\cdot)] \tag{2.21}$$

for all $\bar{y}:[a,b]\to\mathbb{R}$ satisfying

$$\bar{y}(a)=y(a),\ \ \bar{y}(b)=y(b).$$

REMARK. We call such a local minimizer *strong* since we require only that (2.20) be valid, and not necessarily also that

$$\max_{[a,b]}|y'-\bar{y}'|\leq\delta. \tag{2.22}$$



y and $\bar{y}$ are close, but y' and $\bar{y}'$ are not

In particular (2.20) does *not* imply that $L(x, \bar{y}, \bar{y}')$ is close to $L(x, y, y')$ pointwise. □

2.3.1. Fields of extremals.

We begin with the assertion that $y(\cdot)$ is a strong local minimizer, provided we can embed it within a field of extremals. This means that we can find a family of other solutions of (E-L), depending upon a parameter c, that surround $y(\cdot)$:

DEFINITION. We say that $w : [a, b] \times [-1, 1] \to \mathbb{R}$, $w = w(x, c)$, is a **field of extremals** about $y(\cdot)$ provided:

(i) For each $-1 \leq c \leq 1$,

$$-\left(\frac{\partial L}{\partial z}(x, w, w')\right)' + \frac{\partial L}{\partial y}(x, w, w') = 0 \qquad (a \leq x \leq b); \tag{2.23}$$

(ii)

$$y(x) = w(x, 0) \qquad (a \leq x \leq b); \tag{2.24}$$

and

(iii)

$$\frac{\partial w}{\partial c}(x, 0) > 0 \qquad (a \leq x \leq b). \tag{2.25}$$

NOTATION. In (2.23) and below we write

$$w'(x, c) = \frac{\partial w}{\partial x}(x, c).$$

□

REMARK. The importance of condition (2.25) is that it ensures that the graphs of the functions

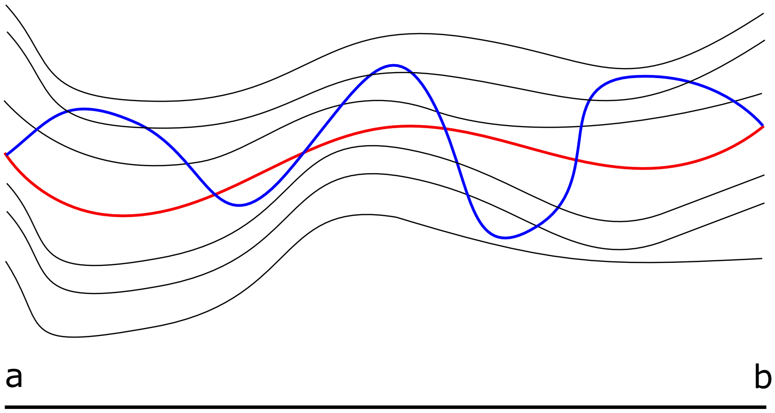
$$\{w(\cdot,c) \mid |c| \le \varepsilon\}$$

do not intersect on the interval $[a,b]$, if $\varepsilon > 0$ is sufficiently small. Therefore each point close enough to the graph of $y(\cdot)$ lies on the graph of precisely one of the functions $\{w(\cdot,c) \mid |c| \le \varepsilon\}$. □

THEOREM 2.3.1. Suppose $y(\cdot)$ is an extremal that lies within a field of extremals w, as above. Assume also that

$$z \mapsto L(x,y,z) \text{ is convex} \quad (a \le x \le b, y \in \mathbb{R}). \tag{2.26}$$

Then $y(\cdot)$ is a strong local minimizer of $I[\,\cdot\,]$.



The graph of y is red, the graph of $\bar{y}$ is blue and the graphs of the other extremals are black

Proof. 1. Let $\bar{y} : [a,b] \to \mathbb{R}$ satisfy $\bar{y}(a) = y(a)$, $\bar{y}(b) = y(b)$ and

$$\max_{[a,b]} |y - \bar{y}| \le \delta$$

for some small number $\delta > 0$.

As noted in the Remark above, the Implicit Function Theorem implies in view of (2.25) that we can uniquely write

$$\bar{y}(x) = w(x, c(x)) \qquad (a \le x \le b) \tag{2.27}$$

for a function $c : [a,b] \to (-1,1)$ provided $\delta > 0$ is small enough. Furthermore, (2.24) forces

$$c(a) = c(b) = 0. \tag{2.28}$$

2. According to (2.27) we have

$$\bar{y}'(x) = w'(x, c(x)) + \frac{\partial w}{\partial c}(x, c(x))c'(x).$$

Therefore the convexity condition (2.26) implies

$$\begin{aligned} L(x, \bar{y}, \bar{y}') &\geq L(x, \bar{y}, w')\big|_{c=c(x)} + \frac{\partial L}{\partial z}(x, \bar{y}, w')\frac{\partial w}{\partial c}c'\big|_{c=c(x)} \\ &= L(x, y, y') + E, \end{aligned} \tag{2.29}$$

where

$$E = L(x, w, w')\big|_{c=c(x)} - L(x, y, y') + \frac{\partial L}{\partial z}(x, w, w')\frac{\partial w}{\partial c}c'\big|_{c=c(x)}.$$

3. We now show that E is the derivative of a function F that vanishes at $x = a, b$. To see this, we compute using (2.23) that

$$\begin{aligned} E &= \int_0^{c(x)} \frac{\partial}{\partial c}L(x, w(x, c), w'(x, c))\, dc + \frac{\partial L}{\partial z}(x, w, w')\frac{\partial w}{\partial c}c'\big|_{c=c(x)} \\ &= \int_0^{c(x)} \frac{\partial L}{\partial y}\frac{\partial w}{\partial c} + \frac{\partial L}{\partial z}\frac{\partial w'}{\partial c}\, dc + \frac{\partial L}{\partial z}(x, w, w')\frac{\partial w}{\partial c}c'\big|_{c=c(x)} \\ &= \int_0^{c(x)} \left(\frac{\partial L}{\partial z}\right)'\frac{\partial w}{\partial c} + \frac{\partial L}{\partial z}\frac{\partial w'}{\partial c}\, dc + \frac{\partial L}{\partial z}(x, w, w')\frac{\partial w}{\partial c}c'\big|_{c=c(x)} \\ &= F' \end{aligned}$$

for

$$F(x) = \int_0^{c(x)} \frac{\partial L}{\partial z}(x, w, w')\frac{\partial w}{\partial c}\, dc.$$

We used here the calculus formula

$$\frac{d}{dx}\left(\int_a^{g(x)} f(x, t)\, dt\right) = \int_a^{g(x)} \frac{\partial f}{\partial x}(x, t)\, dt + f(x, g(x))g'(x).$$

Observe also that

$$F(b) = F(a) = 0, \tag{2.30}$$

since $c(a) = c(b) = 0$, according to (2.28).

4. It now follows from (2.29) and (2.30) that

$$\int_a^b L(x, \bar{y}, \bar{y}')\, dx \geq \int_a^b L(x, y, y') + F'\, dx = \int_a^b L(x, y, y')\, dx.$$

Therefore

$$I[y(\cdot)] \leq I[\bar{y}(\cdot)].$$

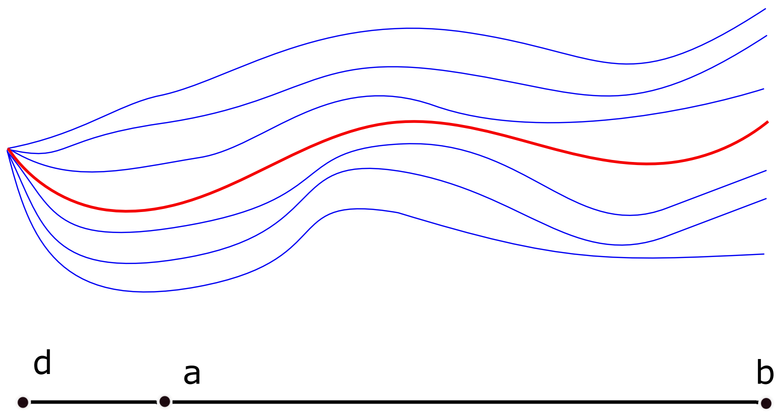
□

REMARK. This direct proof, which uses ideas from Ball-Murat [**B-M**], is more straightforward than conventional arguments involving Hilbert's invariant integral (as explained, for instance, in Kot [**K**]).

Clarke and Zeidan present in [**C-Z**] a very interesting alternative approach, using the Riccati equation (R) to build an appropriate calibration function. □

2.3.2. More on conjugate points.

We now need to understand when we can embed a given extremal within a field of extremals. A clue comes from our earlier discussion of the Riccati equation and conjugate points.



The graph of y is red and the graphs of the other extremals are blue

THEOREM 2.3.2. Let $y : [a, b] \to \mathbb{R}$ be an extremal, and suppose that there are no conjugate points with respect to a within the interval $[a, b]$.

(i) Then we can embed $y(\cdot)$ within a field w of extremals.

(ii) Consequently if $z \mapsto L(x, y, z)$ is convex, then $y(\cdot)$ is a strong local minimizer of $I[\,\cdot\,]$.

Proof. 1. First, as noted in the proof of Theorem 2.2.3, for some point $d < a$ there are no conjugate points with respect to d within $(d, b]$. By solving the Euler-Lagrange ODE we can extend y to be a solution defined on the larger interval $[d, b]$.

Now solve the initial-value problems

$$
\begin{cases}
-\left(\frac{\partial L}{\partial z}(x, w, w')\right)' + \frac{\partial L}{\partial y}(x, w, w') = 0 \qquad (d \le x \le b) \\
w(d, c) = y(d) \\
w'(d, c) = y'(d) + c
\end{cases}
\tag{2.31}
$$

for the functions $w = w(x, c)$. By uniqueness of the solution to an initial-value problem, we have $y(x) = w(x, 0)$.

2. Define

$$u(x) = \frac{\partial w}{\partial c}(x, 0). \tag{2.32}$$

We now differentiate the ODE in (2.31) with respect to the variable c. This gives the identity

$$0 = -\left(\frac{\partial^2 L}{\partial z \partial y}\frac{\partial w}{\partial c} + \frac{\partial^2 L}{\partial z^2}\frac{\partial w'}{\partial c}\right)' + \frac{\partial^2 L}{\partial y^2}\frac{\partial w}{\partial c} + \frac{\partial^2 L}{\partial y \partial z}\frac{\partial w'}{\partial c}.$$

We now put $c = 0$ and recall the definitions (2.10) and (2.32):

$$0 = -(Bu + Au')' + Cu + Bu' = Ju.$$

Upon differentiating as well the initial conditions in (2.31), we see that u solves the initial-value problem

$$\begin{cases} Ju = 0 & (d \leq x \leq b) \\ u(d) = 0, \quad u'(d) = 1. \end{cases} \tag{2.33}$$

3. As shown in the proof of Theorem 2.2.3, the hypothesis of no conjugate points implies that $u = \frac{\partial w}{\partial c}(\cdot, 0)$ is strictly positive on the interval $[a, b]$. Therefore $w : [a, b] \times [-1, 1] \to \mathbb{R}$ is a field of extremals.

Now recall Theorem 2.3.1 to deduce that $y(\cdot)$ is a strong local minimizer, provided we suppose also that $z \mapsto L$ is convex. □

REMARK. We are able to transform (2.31) into (2.33) by differentiating in the parameter c, since the Jacobi differential operator

$$Ju = -(Au' + Bu)' + Bu' + Cu$$

is the linearization of the Euler-Lagrange equation about $y(\cdot)$. □

2.4. Existence of minimizers

2.4.1. Examples.

We next build upon our insights from second variation calculations to discuss the existence of minimizers.

We have thus far simply assumed for various variational problems that minimizers exist, but in fact this issue can be quite complicated. Even quite simple looking problems may have no solution:

EXAMPLE. For instance, the functional

$$I[y(\cdot)] = \int_0^1 y^2 - (y')^2 \, dx$$

does not have a minimizer over the admissible set

$$\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid y(0) = 0, y(1) = 0\}.$$

This is clear from the second variation condition (2.2), which definitely fails for all functions $y_0(\cdot) \in \mathcal{A}$ since

$$L = y^2 - z^2, \quad \frac{\partial^2 L}{\partial z^2} < 0.$$

In fact, we have

$$\inf_{\mathcal{A}} I[y(\cdot)] = -\infty.$$

To prove this, define for each integer k

$$y_k(x) = \sin(k\pi x) \in \mathcal{A}$$

and observe that $I[y_k(\cdot)] \to -\infty$ as $k \to \infty$. □

EXAMPLE. As an even easier example of nonexistence of minimizers consider

$$I[y(\cdot)] = \int_0^1 y \, dx.$$

It is not hard to see that

$$\inf_{y(\cdot) \in \mathcal{A}} I[y(\cdot)] = -\infty \tag{2.34}$$

for the admissible set $\mathcal{A}$ as above. □

EXAMPLE. This example illustrates a different mechanism for the failure of minimizers to exist. We investigate

$$I[y(\cdot)] = \int_0^1 x^2 (y')^2 \, dx, \tag{2.35}$$

over the admissible set $\mathcal{A} = \{y : [0,1] \to \mathbb{R} \mid y(0) = 1, y(1) = 0\}$.

Define

$$y_k(x) = \begin{cases} 0 & (\frac{1}{k} \le x \le 1) \\ 1 - kx & (0 \le x \le \frac{1}{k}). \end{cases}$$

Then

$$I[y_k(\cdot)] = \int_0^{\frac{1}{k}} x^2 k^2 \, dx = \frac{1}{3k} \to 0 \quad \text{as } k \to \infty,$$

and consequently

$$\inf_{y(\cdot)\in\mathcal{A}} I[y(\cdot)] = 0.$$

But the minimum is not attained, since $\int_0^1 x^2(y')^2\,dx > 0$ for all $y(\cdot)\in\mathcal{A}$. □

EXAMPLE. We next show that the problem of minimizing

$$I[y(\cdot)] = -\int_0^1 xyy'\,dx \tag{2.36}$$

over $\mathcal{A} = \{y : [0,1] \to \mathbb{R} \mid y(0) = 0, y(1) = b\}$ has a solution only if $b = 0$.

To see this, we first integrate by parts:

$$I[y(\cdot)] = -\frac{1}{2}\int_0^1 x(y^2)'\,dx = -\frac{b^2}{2} + \frac{1}{2}\int_0^1 y^2\,dx.$$

For any $\varepsilon > 0$ we can find a function $y(\cdot)\in\mathcal{A}$ such that $\int_0^1 y^2\,dx < \varepsilon$. Hence

$$\inf_{y(\cdot)\in\mathcal{A}} I[y(\cdot)] = -\frac{b^2}{2}.$$

But the minimum is not attained, because $\int_0^1 y^2\,dx > 0$ for all $y\in\mathcal{A}$ (unless $b = 0$). □

2.4.2. Convexity and minimizers.

The examples in the previous section illustrate that in general variational problems need not have solutions. And even when we can show the existence of minimizers, this theory depends upon the advanced mathematical topics of measure theory and functional analysis, which are beyond the scope of these notes. Instead, in this section we provide a brief discussion about the key assumption for the existence theory for minimizers and mention some typical theorems.

The primary requirement, strongly motivated by our earlier second variation necessary conditions (2.2) and (2.4), is that

$$\text{the mapping } z \mapsto L(x,y,z) \text{ is convex} \tag{2.37}$$

for each (x,y).

This is the fundamental hypothesis for most existence theorems for minimizers in the calculus of variations. Indeed, a basic existence theorem states that if L satisfies (2.37) and also certain growth conditions, then there exists at least one minimizer $y_0(\cdot)$ within some appropriate class of admissible functions $\mathcal{A}$.

REMARK. Modern existence theory in fact separates the questions of (i) the existence of minimizers and (ii) their regularity (or smoothness) properties. The basic strategy is expand the admissible class $\mathcal{A}$, to include various functions that are less smooth than continuous and piecewise continuously differentiable (as we have supposed throughout these notes). The precise definition for this expanded admissible class is rather subtle, but the idea is that the more functions we accept as belonging to $\mathcal{A}$, the easier it will be to find a minimizer.

And once we have resolved the problem (i) of existence of a minimizer, we can then ask how smooth it is. A typical theorem concerning (ii) requires that L be smooth and satisfies the strengthened convexity condition that

$$\frac{\partial^2 L}{\partial z^2}(x, y, z) \geq \gamma > 0 \tag{2.38}$$

for some constant γ. The regularity assertion is then that a minimizer $y_0(\cdot)$ will be a smooth function of the variable x. □

Chapter 3

MULTIVARIABLE VARIATIONAL PROBLEMS

3.1. Multivariable calculus of variations

In this chapter we extend the theory to variational problems for functions of more than one variable. There will be few new mathematical ideas, and so the only difficulties will be notational.

NOTATION. In the following U will always denote an open subset of $\mathbb{R}^n$, with smooth boundary ∂U. Its closure is $\bar{U} = U \cup \partial U$. □

DEFINITION. Assume $g : \partial U \to \mathbb{R}$ is a given function. The corresponding set of *admissible functions* is

$$\mathcal{A} = \{u : \bar{U} \to \mathbb{R} \mid u(\cdot) \text{ is continuously differentiable}, u = g \text{ on } \partial U\}.$$

NOTATION. We will often write "$u(\cdot)$" to emphasize that $u : \bar{U} \to \mathbb{R}$ is a function. The gradient of a function $u \in \mathcal{A}$ is

$$\nabla u = \begin{bmatrix} \frac{\partial u}{\partial x_1} \\ \vdots \\ \frac{\partial u}{\partial x_n} \end{bmatrix}.$$

□

Assume that in addition we are given a function $L : U \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$,

$$L = L(x, y, z),$$

called the **Lagrangian**. So L is a function of the $2n+1$ real variables $(x, y, z) = (x_1, \dots, x_n, y, z_1, \dots, z_n)$.

DEFINITION. If $u(\cdot) \in \mathcal{A}$, we define

$$\boxed{I[u(\cdot)] = \int_U L(x, u(x), \nabla u(x))\, dx.}$$

Note that we insert the number $u(x)$ into the y-variable slot of $L(x, y, z)$, and the vector $\nabla u(x)$ into the z-variable slot of $L(x, y, z)$.

The **basic problem of the multivariable calculus of variations** is to characterize functions $u_0(\cdot) \in \mathcal{A}$ that satisfy

$$\text{(COV)} \qquad \boxed{I[u_0(\cdot)] = \min_{u(\cdot) \in \mathcal{A}} I[u(\cdot)].}$$

EXAMPLE (Dirichlet energy). As a first example, take

$$L(x, y, z) = \frac{|z|^2}{2} = \frac{1}{2} \sum_{k=1}^{n} z_k^2.$$

Then

$$I[u(\cdot)] = \frac{1}{2} \int_U |\nabla u(x)|^2\, dx,$$

an expression called the *Dirichlet energy* of u. What function $u_0(\cdot)$ minimizes this energy among all other candidates satisfying the given boundary conditions that $u = g$ on ∂U? □

EXAMPLE (Minimal surfaces). For a geometric example, we consider now

$$L(x, y, z) = (1 + |z|^2)^{1/2}.$$

Then

$$\begin{aligned} I[u(\cdot)] &= \int_U \left(1 + |\nabla u(x)|^2\right)^{1/2} dx \\ &= \text{surface area of the graph of } u. \end{aligned}$$

Among all candidates that equal g on the boundary of U, which function $u_0(\cdot)$ gives the surface having the least area? □

3.2. First variation

3.2.1. Euler-Lagrange equation.

We show next that *a minimizer $u_0(\cdot) \in \mathcal{A}$ of our multivariable variational problem automatically solves a certain partial differential equation (PDE).*

THEOREM 3.2.1. Assume $u_0(\cdot) \in \mathcal{A}$ is twice continuously differentiable and solves (COV).

Then u_0 satisfies the nonlinear PDE

$$-\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial z_k}(x, u_0(x), \nabla u_0(x))\right) + \frac{\partial L}{\partial y}(x, u_0(x), \nabla u_0(x)) = 0 \tag{3.1}$$

within the region U.

REMARKS.

(i) The **Euler-Lagrange equation** corresponding to L is the PDE

$$\boxed{-\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial z_k}(x, u, \nabla u)\right) + \frac{\partial L}{\partial y}(x, u, \nabla u) = 0.} \tag{E-L}$$

Solutions u of the Euler-Lagrange equation are **extremals** (or **critical points** or **stationary points**) of $I[\,\cdot\,]$. Consequently, a minimizer $u_0(\cdot)$ is an extremal, subject to the **boundary conditions** that $u = g$ on ∂U.

(ii) Using vector calculus notation, we can also write (E-L) as

$$-\operatorname{div}\left(\nabla_z L(x, u, \nabla u)\right) + \frac{\partial L}{\partial y}(x, u, \nabla u) = 0.$$

Here "div" stands for divergence. □

WARNING ABOUT NOTATION. Most books write the (E-L) PDE as

$$-\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial u_{x_k}}(x, u, \nabla u)\right) + \frac{\partial L}{\partial u}(x, u, \nabla u) = 0$$

or, worse, as

$$-\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial(\frac{\partial u}{\partial x_k})}(x, u, \nabla u)\right) + \frac{\partial L}{\partial u}(x, u, \nabla u) = 0.$$

This is spectacularly bad notation: the Lagrangian $L = L(x, y, z)$ is a function of the $2n+1$ real variables $(x, y, z) = (x_1, \dots, x_n, y, z_1, \dots, z_n)$. So the expressions "$\frac{\partial L}{\partial u}$" and "$\frac{\partial L}{\partial u_{x_k}}$" have no meaning. □

EXAMPLE. Let $L(x,y,z)=\frac{|z|^2}{2}$. Then

$$\frac{\partial L}{\partial y}=0,\quad \frac{\partial L}{\partial z_k}=z_k.$$

Consequently the Euler-Lagrange equation is

$$\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left(\frac{\partial u}{\partial x_k}\right)=0.$$

If we define the **Laplacian**

$$\Delta u=\sum_{k=1}^{n}\frac{\partial^2 u}{\partial x_k^2}, \tag{3.2}$$

then the Euler-Lagrange PDE is **Laplace's equation**

$$\Delta u=0 \qquad \text{within } U.$$

A function solving Laplace's equation is called **harmonic**. Thus *a minimizer of Dirichlet's energy functional is a harmonic function.* □

WARNING ABOUT NOTATION. Many physics and engineering texts use the symbol "∇^2" for the Laplacian. In these and the Math 170 notes, we instead employ ∇^2 to denote the matrix of second derivatives: see Section C in the Appendix. □

EXAMPLE. The Euler-Lagrange equation for the functional

$$\int_U \frac{1}{2}|\nabla u|^2-uf\,dx$$

is **Poisson's equation**

$$-\Delta u=f \quad \text{in } U.$$

The Euler-Lagrange equation for the functional

$$\int_U \frac{1}{2}|\nabla u|^2-F(u)\,dx$$

is the **nonlinear Poisson equation**

$$-\Delta u=f(u) \quad \text{in } U,$$

where $f=F'$. □

EXAMPLE. Assume $L=(1+|z|^2)^{1/2}$. Then

$$\frac{\partial L}{\partial y}=0,\quad \frac{\partial L}{\partial z_k}=\frac{z_k}{(1+|z|^2)^{1/2}}.$$

Consequently the Euler-Lagrange equation is

$$\operatorname{div}\left(\frac{\nabla u}{(1+|\nabla u|^2)^{\frac{1}{2}}}\right)=\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left(\frac{\frac{\partial u}{\partial x_k}}{(1+|\nabla u|^2)^{\frac{1}{2}}}\right)=0.$$

This is the **minimal surface equation**.

GEOMETRIC INTERPRETATION. The expression

$$\boxed{H=\operatorname{div}\left(\frac{\nabla u}{(1+|\nabla u|^2)^{\frac{1}{2}}}\right)}$$

has a geometric meaning; it is the **mean curvature** of the graph of u at the point $(x, u(x))$. Therefore *minimal surfaces have zero mean curvature everywhere.* □

EXAMPLE. If $\mathbf{b} : U \to \mathbb{R}^n$ and $f : \mathbb{R} \to \mathbb{R}$, the expression

$$L = z \cdot \mathbf{b}(x)f(y) + F(y)\operatorname{div}\mathbf{b},$$

where $F' = f$, is a null Lagrangian. Recall from page 8 that this means every function automatically solves the associated Euler-Lagrange PDE.

Indeed, for any function $u : U \to \mathbb{R}$ we have

$$-\operatorname{div}(\nabla_z L) + \frac{\partial L}{\partial y} = -\operatorname{div}(\mathbf{b}f(u)) + \nabla u \cdot \mathbf{b}f'(u) + f(u)\operatorname{div}\mathbf{b} = 0.$$

□

EXAMPLE. Let us for this example write points in $\mathbb{R}^{n+1}$ as (x, t) with $x \in \mathbb{R}^n$ denoting position and $t \geq 0$ denoting time. We consider functions $u = u(x, t)$.

The Euler-Lagrange equation for the functional

$$I[u]=\frac{1}{2}\int_0^T\int_{\mathbb{R}^n}\left(\frac{\partial u}{\partial t}\right)^2-|\nabla_x u|^2\,dxdt$$

is the **wave equation**

$$\frac{\partial^2 u}{\partial t^2}-\Delta u=0,$$

where the Laplacian Δ, defined as in (3.2), acts in the x-variables. However, it is easy to see that the functional $I[\,\cdot\,]$ is unbounded from below on any reasonable admissible class of functions. Consequently, *solutions of the wave equation correspond to extremals that are not minimizers.* □

3.2.2. Derivation of Euler-Lagrange PDE.

NOTATION. We write

$$\boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}.$$

for the **outward pointing unit normal** vector field along ∂U. □

LEMMA 3.2.1.

(i) For each $k = 1, \dots, n$ we have the **multivariable integration-by-parts formula**

$$\int_U \frac{\partial f}{\partial x_k} g \, dx = -\int_U f \frac{\partial g}{\partial x_k} \, dx + \int_{\partial U} f g \nu^k \, dS \tag{3.3}$$

for $k = 1, \dots, n$.

(ii) If $f : U \to \mathbb{R}$ is continuous and

$$\int_U f w \, dx = 0 \quad \text{for all continuous } w : \bar{U} \to \mathbb{R}$$

with $w = 0$ on ∂U, then $f \equiv 0$ within U.

Proof. 1.The Divergence Theorem says that

$$\int_U \text{div } \mathbf{h} \, dx = \int_{\partial U} \mathbf{h} \cdot \boldsymbol{\nu} \, dS.$$

Apply this to the vector field $\mathbf{h} = [0, \dots, 0, fg, 0, \dots, 0]^T$ with the nonzero term fg in the k-th slot.

2. Let $\phi : \bar{U} \to \mathbb{R}$ be positive within U and equal to zero on ∂U. Let $w(x) = \phi(x) f(x)$ above, to find

$$\int_U \phi f^2 \, dx = 0.$$

Hence $\phi(x) f^2(x) = 0$ for all $x \in U$ as the integrand is positive. Then since $\phi(x) > 0$, we conclude that $f(x) = 0$ for all $x \in U$. □

Derivation of Euler-Lagrange equation:

1. Let $w : \bar{U} \to \mathbb{R}$ satisfy $w = 0$ on ∂U. Assume $-1 \leq \tau \leq 1$ and define

$$u_\tau(x) = u_0(x) + \tau w(x) \qquad (x \in U).$$

Note $u_\tau(\cdot) \in \mathcal{A}$, since $w = 0$ on ∂U.

Thus

$$I[u_0(\cdot)] \leq I[u_\tau(\cdot)],$$

since u_0 is the minimizer of I. Define $i(\tau) = I[u_\tau(\cdot)]$. Then

$$i(0) \leq i(\tau).$$

So $\tau \mapsto i(\tau)$ has a minimum at $\tau = 0$ on the interval $-1 \leq \tau \leq 1$, and therefore

$$\frac{di}{d\tau}(0) = 0.$$

2. Now

$$i(\tau) = \int_U L(x, u_\tau, \nabla u_\tau)\, dx = \int_U L(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w)\, dx.$$

Therefore

$$\begin{aligned}\frac{di}{d\tau}(\tau) &= \int_U \frac{\partial}{\partial \tau} L(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w)\, dx. \\ &= \int_U \frac{\partial L}{\partial y}(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w) w \\ &\qquad + \sum_{k=1}^n \frac{\partial L}{\partial z_k}(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w) \frac{\partial w}{\partial x_k}\, dx.\end{aligned}$$

Put $\tau = 0$:

$$0 = \frac{di}{d\tau}(0) = \int_U \frac{\partial L}{\partial y}(x, u_0, \nabla u_0) w + \sum_{k=1}^n \frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0) \frac{\partial w}{\partial x_k}\, dx.$$

We now integrate by parts, to deduce that

$$\int_U \left[\frac{\partial L}{\partial y}(x, u_0, \nabla u_0) - \sum_{k=1}^n \frac{\partial}{\partial x_k} \left(\frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0) \right) \right] w\, dx = 0.$$

This is valid for all functions w such that $w = 0$ on ∂U. The lemma before now implies that the (E-L) PDE holds everywhere in U. □

3.3. Second variation

THEOREM 3.3.1. Suppose $u_0(\cdot) \in \mathcal{A}$ is a minimizer and u_0 is continuously differentiable.

(i) Then

$$\begin{aligned}(3.4) \qquad \int_U \sum_{k,l=1}^n & \frac{\partial^2 L}{\partial z_k \partial z_l}(x, u_0, \nabla u_0) \frac{\partial w}{\partial x_k} \frac{\partial w}{\partial x_l} \\ & + 2 \sum_{k=1}^n \frac{\partial^2 L}{\partial z_k \partial y}(x, u_0, \nabla u_0) \frac{\partial w}{\partial x_k} w + \frac{\partial^2 L}{\partial y^2}(x, u_0, \nabla u_0) w^2\, dx \geq 0\end{aligned}$$

for all $w : \bar{U} \to \mathbb{R}$ with $w = 0$ on ∂U.

(ii) Furthermore,

$$\boxed{\nabla_z^2 L(x, u_0(x), \nabla u_0(x)) \succeq 0} \qquad (x \in U). \tag{3.5}$$

DEFINITION. The expression on the left hand side of (3.4) the **second variation** of $I[\,\cdot\,]$ about $u_0(\cdot)$, evaluated at $w(\cdot)$. □

REMARK. The inequality (3.5) means that

$$\sum_{k,l=1}^{n} \frac{\partial^2 L}{\partial z_k \partial z_l}(x, u_0, \nabla u_0) y_k y_l \geq 0 \qquad (y \in \mathbb{R}^n).$$

If the mapping

$$z \mapsto L(x, u_0(x), z) \quad \text{is convex,} \tag{3.6}$$

then (3.5) holds automatically. □

Proof. 1. Select $w : \bar{U} \to \mathbb{R}$, with $w = 0$ on ∂U, and define $u_\tau(x) = u_0(x) + \tau w(x)$ for $-1 \leq \tau \leq 1$. Then $u_\tau(\cdot) \in \mathcal{A}$, and so $\tau \mapsto i(\tau)$ has a minimum at $\tau = 0$ and thus

$$\frac{d^2 i}{d\tau^2}(0) \geq 0.$$

2. Differentiating twice with respect to τ, we find

$$\begin{aligned} \frac{d^2 i}{d\tau^2}(\tau) = \int_U & \frac{\partial^2 L}{\partial y^2}(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w) w^2 \\ & + 2 \sum_{k=1}^{n} \frac{\partial^2 L}{\partial y \partial z_k}(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w) w \frac{\partial w}{\partial x_k} \\ & + \sum_{k,l=1}^{n} \frac{\partial^2 L}{\partial z_k \partial z_l}(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w) \frac{\partial w}{\partial x_k} \frac{\partial w}{\partial x_l} \, dx. \end{aligned}$$

Put $\tau = 0$ and recall $\frac{d^2 i}{d\tau^2}(0) \geq 0$, to prove (3.4).

3. As in the second variation calculation for the scalar case, we define

$$\phi(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 2 - x & \text{if } 1 \leq x \leq 2 \end{cases}$$

and extend ϕ to be 2-periodic on $\mathbb{R}$. Thus ϕ is a sawtooth function and $\phi' = \pm 1$ except at the integers.

Next let $\zeta : U \to \mathbb{R}$ be any continuously differentiable function with $\zeta = 0$ on ∂U. Finally, choose any $y \in \mathbb{R}^n$ and write

$$w_\varepsilon(x) = \varepsilon\phi(\tfrac{x\cdot y}{\varepsilon})\zeta(x).$$

for each $\varepsilon > 0$. Then w_ε is differentiable except along finitely many hyperplanes that intersect U, with

$$\nabla w_\varepsilon(x) = \phi'(\tfrac{x\cdot y}{\varepsilon})\zeta(x)y + \varepsilon\phi(\tfrac{x\cdot y}{\varepsilon})\nabla\zeta(x).$$

Note that the second term on the right is less than or equal to $A\varepsilon$ for some appropriate constant A.

Insert w_ε in place of w in (3.4). It follows that

$$\int_U \sum_{k,l=1}^n \frac{\partial^2 L}{\partial z_k \partial z_l}(x, u_0, \nabla u_0) y_k y_l \zeta^2 \left(\phi'(\tfrac{x\cdot y}{\varepsilon})\right)^2 dx + B\varepsilon \geq 0$$

for some constant B. But

$$\left(\phi'(\tfrac{x\cdot y}{\varepsilon})\right)^2 = 1$$

except along finitely many hyperplanes (which have no effect on the integral). Therefore upon sending $\varepsilon \to 0$, we deduce that

$$\int_U \sum_{k,l=1}^n \frac{\partial^2 L}{\partial z_k \partial z_l}(x, u_0, \nabla u_0) y_k y_l \zeta^2 \, dx \geq 0$$

for all functions ζ as above. This implies (3.5). □

3.4. Extensions and generalizations

3.4.1. Other boundary conditions.

If we change the admissible class of functions and/or if we change the energy functional $I[\,\cdot\,]$, new effects can appear.

For example, suppose we are additionally given a function $B : \partial U \times \mathbb{R} \to \mathbb{R}$, $B = B(x, y)$. We then for this section redefine the energy by adding a boundary integral term:

$$\boxed{I[u(\cdot)] = \int_U L(x, u, \nabla u)\, dx + \int_{\partial U} B(x, u)\, dS.} \tag{3.7}$$

We also redefine

$$\mathcal{A} = \{u : \bar{U} \to \mathbb{R} \mid u \text{ is continuously differentiable}\}, \tag{3.8}$$

so as now to require nothing about the boundary behavior of admissible functions.

THEOREM 3.4.1. Let the energy be given by (3.7) and the admissible class by (3.8). Assume $u_0(\cdot) \in \mathcal{A}$ is a twice continuously differentiable minimizer

(i) Then u_0 solves the usual Euler-Lagrange equation

$$-\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0)\right) + \frac{\partial L}{\partial y}(x, u_0, \nabla u_0) = 0 \tag{3.9}$$

within U.

(ii) Furthermore, we have the boundary condition

$$\boxed{\sum_{k=1}^{n} \frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0)\nu^k + \frac{\partial B}{\partial y}(x, u_0) = 0} \tag{3.10}$$

on ∂U.

In vector notation, (3.10) reads

$$\nabla_z L(x, u_0, \nabla u_0) \cdot \boldsymbol{\nu} + \frac{\partial B}{\partial y}(x, u_0) = 0 \quad \text{on } \partial U. \tag{3.11}$$

INTERPRETATION. The identity (3.10) is the **natural boundary condition** (or **transversality condition**) hidden in our new variational problem. It appears automatically when we compute the first variation. □

Proof. 1. Select any $w : \bar{U} \to \mathbb{R}$, but do not require that w vanishes on the boundary. Define $u_\tau(x) = u_0(x) + \tau w(x)$ $(x \in U)$, and observe $u_\tau(\cdot) \in \mathcal{A}$. Thus $i(\tau) = I[u_\tau(\cdot)]$ has a minimum at $\tau = 0$ and therefore

$$\frac{di}{d\tau}(0) = 0.$$

2. Now

$$\begin{aligned} i(\tau) &= I[u_\tau(\cdot)] \\ &= \int_U L(x, u_0 + \tau w, \nabla u_0 + \tau \nabla w)\, dx + \int_{\partial U} B(x, u_0 + \tau w)\, dS. \end{aligned}$$

Therefore

$$\begin{aligned} 0 = \frac{di}{d\tau}(0) = \int_U \frac{\partial L}{\partial y}(x, u_0, \nabla u_0) w + \sum_{k=1}^{n} \frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0) \frac{\partial w}{\partial x_k}\, dx \\ + \int_{\partial U} \frac{\partial B}{\partial y}(x, u_0) w\, dS. \end{aligned}$$

Integrate by parts, to deduce

$$\int_U \left[\frac{\partial L}{\partial y}(x, u_0(x), \nabla u_0(x)) - \sum_{k=1}^n \frac{\partial}{\partial x_k} \left(\frac{\partial L}{\partial z_k}(x, u_0(x), \nabla u_0(x)) \right) \right] w \, dx$$

$$+ \int_{\partial U} \sum_{k=1}^n \left[\frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0)\nu^k + \frac{\partial B}{\partial y}(x, u_0) \right] w \, dS = 0.$$

This identity is valid for all functions w. If we restrict attention to functions satisfying also $w = 0$ on ∂U, we as usual deduce that the Euler-Lagrange PDE holds within U.

Knowing this, we can then rewrite the foregoing:

$$\int_{\partial U} \sum_{k=1}^n \left[\frac{\partial L}{\partial z_k}(x, u_0, \nabla u_0)\nu^k + \frac{\partial B}{\partial y}(x, u_0) \right] w \, dS = 0.$$

As this identity is valid for all functions w, regardless of their behavior on ∂U, it follows that the boundary condition (3.10) holds everywhere on ∂U. □

EXAMPLE. A minimizer $u_0(\cdot)$ the functional

$$I[u(\cdot)] = \frac{1}{2} \int_U |\nabla u|^2 dx + \int_{\partial U} B(u) \, dS$$

over the admissible class (3.8) solves the nonlinear boundary-value problem

$$\begin{cases} \Delta u_0 = 0 & \text{in } U \\ \frac{\partial u_0}{\partial \nu} + b(u_0) = 0 & \text{on } \partial U, \end{cases}$$

where $b = B'$ and

$$\frac{\partial u}{\partial \nu} = \nabla u \cdot \boldsymbol{\nu}$$

is the **outward normal derivative** of u. □

3.4.2. Integrating factors.

It is often important to determine whether a given PDE (or ODE) problem is **variational**, meaning that it arises as the Euler-Lagrange equation for an appropriate energy functional $I[\,\cdot\,]$.

Consider for instance the linear equation

$$-\Delta u + \boldsymbol{b} \cdot \nabla u = f \quad \text{in } U, \tag{3.12}$$

where $\boldsymbol{b} : U \to \mathbb{R}^n$ is a given vector field, $\boldsymbol{b} = \boldsymbol{b}(x)$. In general (3.12) is not variational.

But for the special case that $\boldsymbol{b} = \nabla\phi$ is the gradient of a potential function $\phi : U \to \mathbb{R}$, let us consider the functional

$$I[u(\cdot)] = \int_U e^{-\phi}\left(\frac{1}{2}|\nabla u|^2 - fu\right)dx. \tag{3.13}$$

The Lagrangian function is

$$L = e^{-\phi(x)}\left(\frac{1}{2}|z|^2 - f(x)y\right).$$

and the corresponding Euler-Lagrange equation for a minimizer u is

$$-\operatorname{div}\left(e^{-\phi}\nabla u\right) - e^{-\phi}f = 0.$$

We simplify and cancel the term $e^{-\phi}$, to rewrite the foregoing as

$$-\Delta u + \nabla\phi\cdot\nabla u = f; \tag{3.14}$$

this is (3.12) when $\boldsymbol{b} = \nabla\phi$.

INTERPRETATION. We can regard $e^{-\phi}$ as an **integrating factor**. This means that if we multiply the PDE (3.14) by this expression, it then becomes variational. □

3.4.3. Integral constraints.

In this section we for simplicity take $L = \frac{1}{2}|z|^2$; so that

$$I[u(\cdot)] = \frac{1}{2}\int_U |\nabla u|^2\,dx.$$

Define also the functional

$$\boxed{J[u(\cdot)] = \int_U G(x, u(x))\,dx,}$$

where $G : U \times \mathbb{R} \to \mathbb{R}$, $G = G(x, y)$. The new admissible class will be

$$\mathcal{A} = \{u : U \to \mathbb{R} \mid u = g \text{ on } \partial U, J[u(\cdot)] = 0\}. \tag{3.15}$$

THEOREM 3.4.2. Assume that $u_0 \in \mathcal{A}$ is a minimizer of $I[\,\cdot\,]$ over $\mathcal{A}$. Suppose also that

$$\frac{\partial G}{\partial y}(x, u_0) \text{ is not identically zero on } U. \tag{3.16}$$

Then there exists $\lambda_0 \in \mathbb{R}$ such that

$$\boxed{-\Delta u_0 + \lambda_0\frac{\partial G}{\partial y}(x, u_0) = 0} \tag{3.17}$$

within U.

We omit the proof, which is similar to that for Theorem 1.3.5. As in that previous theorem, we interpret λ_0 as the Lagrange multiplier for the integral constraint $J[u(\cdot)] = 0$. The requirement (3.16) is a constraint qualification condition.

For an application, see Theorem 3.5.1 below.

3.4.4. Systems.

We can also extend our theory to handle systems. For this, we assume $\mathbf{g} : \partial U \to \mathbb{R}^m$ is given, and redefine the class of admissible functions, now to be

$$\mathcal{A} = \{\mathbf{u} : \bar{U} \to \mathbb{R}^m \mid \mathbf{u} = \mathbf{g} \text{ on } \partial U\}.$$

NOTATION. We write

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}, \quad \nabla \mathbf{u} = \begin{bmatrix} (\nabla u_1)^T \\ \vdots \\ (\nabla u_m)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \dots & \frac{\partial u_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_m}{\partial x_1} & \dots & \frac{\partial u_m}{\partial x_n} \end{bmatrix}.$$

□

Suppose next we have a Lagrangian function $L : U \times \mathbb{R}^m \times \mathbb{M}^{m\times n} \to \mathbb{R}$,

$$L = L(x, y, Z),$$

where $\mathbb{M}^{m\times n}$ denotes the space of real, $m \times n$ matrices.

NOTATION. In this section only, we will write a matrix $Z \in \mathbb{M}^{m\times n}$ as

$$Z = \begin{bmatrix} z_1^1 & \dots & z_n^1 \\ \vdots & \ddots & \vdots \\ z_1^m & \dots & z_n^m \end{bmatrix}.$$

□

DEFINITION. If $\mathbf{u}(\cdot) \in \mathcal{A}$, we define

$$I[\mathbf{u}(\cdot)] = \int_U L(x, \mathbf{u}(x), \nabla \mathbf{u}(x))\, dx.$$

Note that we insert $\mathbf{u}(x)$ into the y-variables of $L(x, y, Z)$, and $\nabla \mathbf{u}(x)$ into the Z-variables.

THEOREM 3.4.3. Assume $\mathbf{u}_0(\cdot) \in \mathcal{A}$ minimizers $I[\,\cdot\,]$ and $\mathbf{u}_0$ is twice continuously differentiable.

Then $\mathbf{u}_0$ solves within U the system of nonlinear PDE

$$\text{(E-L)} \quad -\sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial z_k^l}(x, \mathbf{u}_0(x), \nabla \mathbf{u}_0(x))\right) + \frac{\partial L}{\partial y_l}(x, \mathbf{u}_0(x), \nabla \mathbf{u}_0(x)) = 0$$

for $l = 1, \dots, m$.

Proof. 1. Select $\mathbf{w} : \bar{U} \to \mathbb{R}$ such that $\mathbf{w} = 0$ on ∂U, and then define and define

$$\mathbf{u}_\tau(x) = \mathbf{u}_0(x) + \tau \mathbf{w}(x) \qquad (x \in U).$$

for $-1 \leq \tau \leq 1$ Then $\mathbf{u}_\tau(\cdot) \in \mathcal{A}$, since $\mathbf{w} = 0$ on ∂U. Hence

$$I[\mathbf{u}_0(\cdot)] \leq I[u_\tau(\cdot)].$$

Define $i(\tau) = I[\mathbf{u}_\tau(\cdot)]$. Then $\tau \mapsto i(\tau)$ has a minimum at $\tau = 0$, and therefore

$$\frac{di}{d\tau}(0) = 0.$$

2. We have

$$i(\tau) = I[\mathbf{u}_\tau(\cdot)] = \int_U L(x, \mathbf{u}_0(x) + \tau \mathbf{w}(x), \nabla \mathbf{u}_0(x) + \tau \nabla \mathbf{w}(x))\, dx.$$

Therefore

$$\begin{aligned}\frac{di}{d\tau}(\tau) = \int_U \sum_{l=1}^m \frac{\partial L}{\partial y_l}(x, \mathbf{u}_0 + \tau \mathbf{w}, \nabla \mathbf{u}_0 + \tau \nabla w) w_l \\ + \sum_{l=1}^m \sum_{k=1}^n \frac{\partial L}{\partial z_k^l}(x, \mathbf{u}_0 + \tau \mathbf{w}, \nabla \mathbf{u}_0 + \tau \nabla \mathbf{w}) \frac{\partial w_l}{\partial x_k}\, dx;\end{aligned}$$

and so

$$0 = \frac{di}{d\tau}(0) = \int_U \sum_{l=1}^m \frac{\partial L}{\partial y_l}(x, \mathbf{u}_0, \nabla \mathbf{u}_0) w_l + \sum_{l=1}^m \sum_{k=1}^n \frac{\partial L}{\partial z_k}(x, \mathbf{u}_0, \nabla \mathbf{u}_0) \frac{\partial w_l}{\partial x_k}\, dx.$$

Now fix some index $l \in \{1, \dots, m\}$ and put

$$\mathbf{w} = [0 \, \dots \, 0 \, w \, 0 \, \dots \, 0]^T,$$

where the real-valued function w appears in the l-th slot. Then we have

$$\int_U \frac{\partial L}{\partial y_l}(x, \mathbf{u}_0, \nabla \mathbf{u}_0) w + \sum_{k=1}^n \frac{\partial L}{\partial z_k}(x, \mathbf{u}_0, \nabla \mathbf{u}_0) \frac{\partial w}{\partial x_k}\, dx = 0.$$

Upon integrating by parts, we deduce as usual that the l-th equation of the (E-L) system of PDE holds. □

3.5. Applications

3.5.1. Eigenvalues, eigenfunctions.

Assume for this section that $U \subset \mathbb{R}^n$ is a connected, open set, with smooth boundary ∂U.

THEOREM 3.5.1. (i) There exists a real number $\lambda_1 > 0$ and a smooth function w_1 such that

$$\begin{cases} -\Delta w_1 = \lambda_1 w_1 & \text{in } U \\ w_1 = 0 & \text{on } \partial U \\ w_1 > 0 & \text{in } U, \quad \int_U w_1^2\,dx = 1. \end{cases} \tag{3.18}$$

(ii) Furthermore,

$$\lambda_1 = \min\left\{ \int_U |\nabla u|^2\,dx \mid \int_U u^2\,dx = 1, u = 0 \text{ on } \partial U \right\}. \tag{3.19}$$

DEFINITION. We call λ_1 the **principal eigenvalue** for the Laplacian with zero boundary conditions on ∂U. The function w_1 is a principal **eigenfunction**.

REMARK. We can also write

$$\boxed{\lambda_1 = \min_{u=0 \text{ on } \partial U, u\neq 0} \frac{\int_U |\nabla u|^2\,dx}{\int_U u^2\,dx}.} \tag{3.20}$$

This is **Rayleigh's principle**. □

Proof. If $w_1 \in \mathcal{A}$ is a minimizer for the constrained variational problem (3.19), then

$$\frac{\partial G}{\partial y}(w_1) = 2w_1 \not\equiv 0.$$

Therefore Theorem 3.4.2 tells us that the Euler-Lagrange equation is

$$-\Delta w_1 + \mu w_1 = 0,$$

where μ is the eigenvalue for the constraint. But then

$$\mu = \mu \int_U w_1^2\,dx = -\int_U |\nabla w_1|^2\,dx = -\lambda_1.$$

The existence of a smooth minimizer w_1, with $w_1 > 0$ within U, requires graduate level mathematical ideas beyond the scope of these notes. □

THEOREM 3.5.2. (i) There exist real numbers $0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots$ and smooth function $w_1, w_2, w_3, \ldots$ such that

$$\lambda_k \to \infty$$

and

$$\begin{cases} -\Delta w_k = \lambda_k w_k & \text{in } U \\ w_k = 0 & \text{on } \partial U \\ \int_U w_k^2\,dx = 1. \end{cases} \tag{3.21}$$

(ii) Furthermore,

$$\lambda_k = \min\left\{\int_U |\nabla u|^2\,dx \mid \int_U u^2\,dx = 1, u = 0 \text{ on } \partial U, \right.$$
$$\left.\int_U w_l u\,dx = 0\ (l = 1, \dots, k-1)\right\}. \tag{3.22}$$

DEFINITION. We call λ_k the **k-th eigenvalue** for the Laplacian with zero boundary conditions on ∂U. The function w_k is a corresponding **eigenfunction**. □

Proof. Assume that (3.21) holds for $l = 1, \dots, k-1$. The Euler-Lagrange PDE for the minimization problem (3.22) is

$$-\Delta w_k = \lambda_k w_k + \sum_{l=1}^{k-1} \mu_l w_l, \tag{3.23}$$

where λ_k is the Lagrange multiplier for the constraint $\int_U w^2\,dx = 1$ and μ_l is the Lagrange multiplier for the constraint $\int_U w_l w\,dx = 0$ for $l = 1, \dots, k-1$. Let $1 \le m \le k-1$, multiply (3.23) by w_m, and integrate:

$$-\int_U \Delta w_k w_m\,dx = \int_U (\lambda_k w_k + \sum_{l=1}^{k-1} \mu_l w_l) w_m\,dx = \mu_m.$$

Therefore

$$\mu_m = -\int_U \Delta w_k w_m\,dx = \int_U \nabla w_k \cdot \nabla w_m\,dx$$
$$= -\int_U w_k \Delta w_m\,dx = \lambda_m \int_U w_k w_m\,dx = 0.$$

Thus (3.23) becomes $-\Delta w_k = \lambda_k w_k$. □

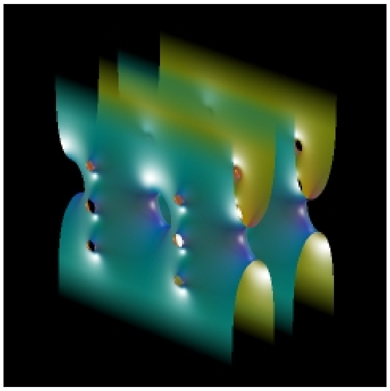
REMARK. (Level curves for eigenfunctions) If U is connected, the first eigenfunction w_1 is positive; but the higher eigenfunctions w_k for $k = 2, \dots$ change sign within U.

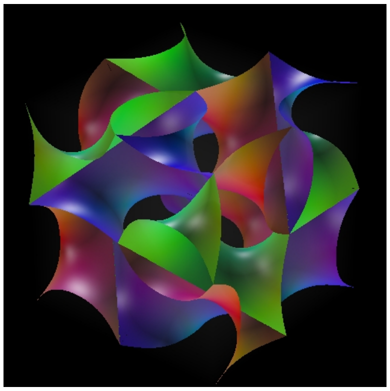
Take a thin metal plate cut into the shape $U \subset \mathbb{R}^2$, sprinkle it with sand, and then connect the plate to an audio speaker. If we we play appropriate high frequencies over the speaker, the plate resonates according to the higher eigenfunctions of the Laplacian (at least approximately) and so the sand collects into the level sets $\{w_k = 0\}$. At higher and higher frequencies, complicated and beautiful structures appear, called **Chladni patterns**: see www.youtube.com/watch?v=wvJAgrUBF4w. □

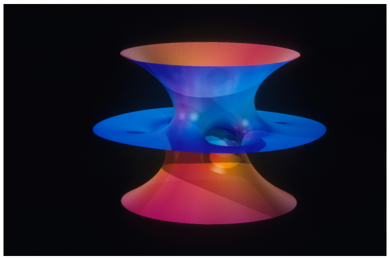
3.5.2. Minimal surfaces.

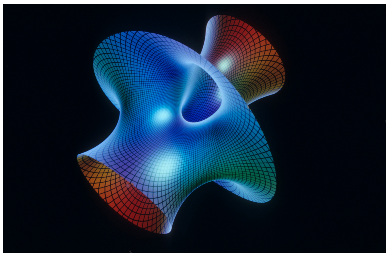
We saw on page 70 that we can apply variational methods to study minimal surfaces that are the graphs of functions $u : \bar{U} \to \mathbb{R}$. The main observation was that for such surfaces the mean curvature H equals 0.

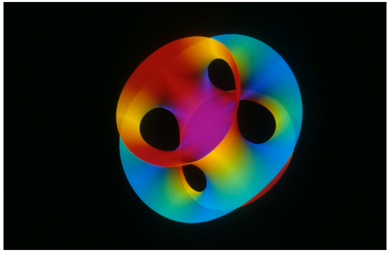
For several centuries there has been intense mathematical investigation of more complicated minimal surfaces that cannot be represented globally as a graph. The study of these is beyond the scope of these notes, but following are some pictures (provided to me by David Hoffman) of some beautiful 2-dimensional minimal surfaces. For all of these $H = \kappa_1 + \kappa_2$ is identically zero, where κ_1, κ_2 are the principal curvatures.





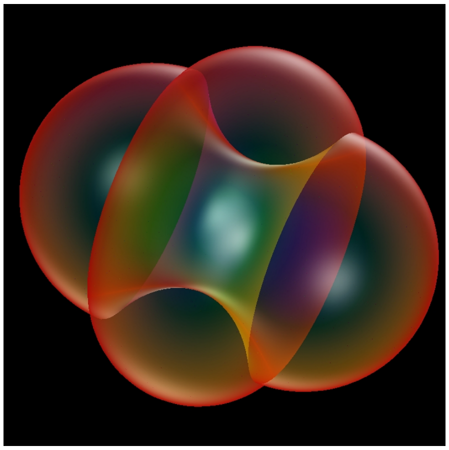






These surfaces are mathematical models for physical *soap films*.

More complicated are mathematical models for *soap bubbles*. These entail minimizing surface area, subject to volume constraints (= volume of the air within the bubbles). Such surfaces have locally *constant mean curvature*, and we can interpret the constant as a Lagrange multiplier for the volume constraint.



Bubbles with constant mean curvature

3.5.3. Harmonic maps.

Let us next study how to minimize the Dirichlet energy among maps from $U \subset \mathbb{R}^n$ into the unit sphere $S^{m-1} \subset \mathbb{R}^m$, satisfying given boundary conditions. We therefore take as the admissible class of mappings

$$\mathcal{A} := \{\mathbf{u} : U \to \mathbb{R}^m \mid \mathbf{u} = \mathbf{g} \text{ on } \partial U, |\mathbf{u}| = 1\}.$$

The corresponding energy is

$$I[\mathbf{u}(\cdot)] = \frac{1}{2}\int_U |\nabla \mathbf{u}|^2\,dx$$

THEOREM 3.5.3. Let $\mathbf{u}_0 \in \mathcal{A}$ satisfy

$$I[\mathbf{u}_0] = \min_{\mathbf{u}\in\mathcal{A}} I[\mathbf{u}].$$

Then

$$\begin{cases} -\Delta \mathbf{u}_0 = |\nabla \mathbf{u}_0|^2 \mathbf{u}_0 & \text{in } U \\ \quad \mathbf{u}_0 = \mathbf{g} & \text{on } \partial U. \end{cases} \tag{3.24}$$

INTERPRETATION. The *function* $\lambda_0 = |\nabla \mathbf{u}_0|^2$ is the Lagrange multiplier corresponding to the *pointwise* constraint $|\mathbf{u}_0| = 1$. □

Proof. 1. Select $\mathbf{w} : U \to \mathbb{R}^m$, with $\mathbf{w} = 0$ on ∂U. Then since $|\mathbf{u}_0| = 1$, it follows that $|\mathbf{u}_0 + \tau\mathbf{w}| \neq 0$ for each sufficiently small τ. Consequently

$$\mathbf{u}(\tau) := \frac{\mathbf{u}_0 + \tau\mathbf{w}}{|\mathbf{u}_0 + \tau\mathbf{w}|} \in \mathcal{A}. \tag{3.25}$$

Thus

$$i(\tau) := I[\mathbf{w}(\tau)]$$

has a minimum at $\tau = 0$, and so, as usual,

$$\frac{di}{d\tau}(0) = 0.$$

2. We have

$$i'(0) = \int_U \nabla \mathbf{u} \cdot \nabla \mathbf{u}'(0)\,dx = 0. \tag{3.26}$$

where $' = \frac{d}{d\tau}$. But we compute directly from (3.25) that

$$\mathbf{u}'(\tau) = \frac{\mathbf{w}}{|\mathbf{u}_0 + \tau\mathbf{w}|} - \frac{[(\mathbf{u}_0 + \tau\mathbf{w})\cdot\mathbf{w}](\mathbf{u}_0 + \tau\mathbf{w})}{|\mathbf{u}_0 + \tau\mathbf{w}|^3}.$$

So $\mathbf{u}'(0) = \mathbf{w} - (\mathbf{u}_0 \cdot \mathbf{w})\mathbf{u}_0$. Put this equality into (3.26):

$$0 = \int_U \nabla \mathbf{u}_0 \cdot \nabla \mathbf{w} - \nabla \mathbf{u}_0 \cdot \nabla((\mathbf{u}_0 \cdot \mathbf{w})\mathbf{u}_0)\,dx. \tag{3.27}$$

We next differentiate the identity $|\mathbf{u}_0|^2 \equiv 1$, to learn that

$$(\nabla \mathbf{u}_0)^T \mathbf{u}_0 = 0.$$

Using this fact, we then verify that

$$\nabla \mathbf{u}_0 \cdot \nabla((\mathbf{u}_0 \cdot \mathbf{w})\mathbf{u}_0) = |\nabla \mathbf{u}_0|^2 (\mathbf{u}_0 \cdot \mathbf{w})$$

in U. Inserting this into (3.27) gives

$$0 = \int_U \nabla \mathbf{u}_0 \cdot \nabla \mathbf{w} - |\nabla \mathbf{u}_0|^2 (\mathbf{u}_0 \cdot \mathbf{w})\, dx.$$

This identity, valid for all functions $\mathbf{w}$ as above, implies the PDE in (3.24). □

3.5.4. Gradient flows.

The Euler-Lagrange equation arising in PDE theory are mostly equilibrium equations that do not entail changes in time. But it is interesting also to consider certain time-dependent equations that can be interpreted as **gradient flows**. Recall that if we are given an "energy" function $\Phi : \mathbb{R}^n \to \mathbb{R}$, the system of ODE

$$\dot{\mathbf{x}} = -\nabla \Phi(\mathbf{x}), \tag{3.28}$$

describes a "downhill" gradient flow

We introduce next a PDE version, corresponding to the energy

$$I[u(\cdot)] = \int_U L(x, u(x), \nabla u(x))\, dx.$$

The corresponding time dependent gradient flow, generalizing (3.28), is the PDE

$$\boxed{\frac{\partial u}{\partial t} = \sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left(\frac{\partial L}{\partial z_k}(x, u, \nabla u)\right) - \frac{\partial L}{\partial y}(x, u, \nabla u).} \tag{3.29}$$

We assume that $u = u(x,t)$ solves this PDE, with the boundary condition

$$u = 0 \quad \text{on } \partial U. \tag{3.30}$$

THEOREM 3.5.4.

(i) The function

$$\phi(t) = I[u(\cdot, t)] \qquad (0 \le t < \infty)$$

is nonincreasing on $[0, \infty)$.

(ii) Assume in addition that $(y, z) \mapsto L(x, y, z)$ is convex. Then ϕ is convex on $[0, \infty)$.

Proof. 1. We differentiate in t, to see that

$$\begin{aligned} \frac{d}{dt} I[u(\cdot, t)] &= \frac{d}{dt} \int_U L(x, u(x,t), \nabla_x u(x,t))\, dx \\ &= \int_U \frac{\partial L}{\partial y}(x, u, \nabla u) \frac{\partial u}{\partial t} + \nabla_z L(x, u, \nabla u) \cdot \nabla_x \left(\frac{\partial u}{\partial t}\right) dx \end{aligned}$$

$$
\begin{aligned}
&= \int_U \left[\frac{\partial L}{\partial y}(x, u, \nabla u) - \operatorname{div}(\nabla_z L(x, u, \nabla u)) \right] \frac{\partial u}{\partial t}\, dx \\
&= -\int_U \left(\frac{\partial u}{\partial t} \right)^2 dx \leq 0.
\end{aligned}
$$

Observe that there is no boundary term when we integrate by parts, since if (3.30) holds for all times t, then $\frac{\partial u}{\partial t} = 0$ on ∂U.

2. Differentiate again in t:

$$
\frac{d^2}{dt^2} I[u(\cdot, t)] = -\frac{d}{dt} \int_U \left(\frac{\partial u}{\partial t} \right)^2 dx = -2 \int_U \frac{\partial u}{\partial t} \frac{\partial^2 u}{\partial t^2}\, dx.
$$

We can also differentiate the PDE (3.29), to find

$$
\begin{aligned}
\frac{\partial^2 u}{\partial t^2} = \sum_{k=1}^n \frac{\partial}{\partial x_k} \left(\frac{\partial^2 L}{\partial z_k \partial y} \frac{\partial u}{\partial t} + \sum_{l=1}^n \frac{\partial^2 L}{\partial z_k \partial z_l} \frac{\partial^2 u}{\partial x_l \partial t} \right) \\
- \frac{\partial^2 L}{\partial y^2} \frac{\partial u}{\partial t} - \sum_{l=1}^n \frac{\partial^2 L}{\partial y \partial z_l} \frac{\partial^2 u}{\partial x_l \partial t}.
\end{aligned}
$$

We insert this into the previous calculation, and integrate by parts:

$$
\begin{aligned}
\frac{d^2}{dt^2} I[u(\cdot, t)] = 2 \int_U \sum_{k,l=1}^n \frac{\partial^2 L}{\partial z_k \partial z_l} \frac{\partial^2 u}{\partial x_k \partial t} \frac{\partial^2 u}{\partial x_l \partial t} \\
+ 2 \sum_{k=1}^n \frac{\partial^2 L}{\partial z_k \partial y} \frac{\partial^2 u}{\partial x_k \partial t} \frac{\partial u}{\partial t} + \frac{\partial^2 L}{\partial y^2} \left(\frac{\partial u}{\partial t} \right)^2 dx \geq 0,
\end{aligned}
$$

the last inequality holding provided L is convex in the variables (y, z). □

Chapter 4

OPTIMAL CONTROL THEORY

4.1. The basic problem

This section provides an informal introduction to optimal control theory, and discusses three model problems. The basic issue is this: we are given some system of interest, whose evolution in time is modeled by a differential equation that depends upon certain parameters. We ask *how to optimally and continually adjust these parameters, so as to maximize a given payoff functional.*

To be more precise, assume that the **state** of our system at time $t \geq 0$ is $\mathbf{x}(t)$, where

$$\mathbf{x} : [0, \infty) \to \mathbb{R}^n$$

solves a system of differential equations having the form

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

$x^0 \in \mathbb{R}^n$ denoting the initial state of the system. Here we are given

$$\mathbf{f} : \mathbb{R}^n \times A \to \mathbb{R}^n,$$

where $A \subseteq \mathbb{R}^m$ is the set of **control (or parameter) values**. The (possibly discontinuous) mapping

$$\alpha : [0, \infty) \to A$$

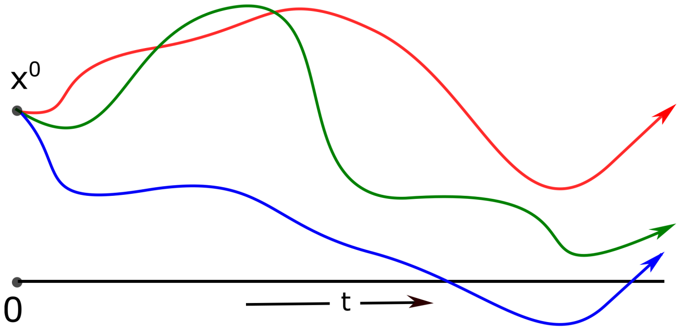
is an **admissible control**.

We write $\mathcal{A}$ for the collection of all admissible controls, and will always assume that for each $\alpha(\cdot) \in \mathcal{A}$, *the solution* $\mathbf{x}(\cdot)$ *of (ODE) exists and is unique.* We call $\mathbf{x}(\cdot)$ the **response** of the system to the control $\alpha(\cdot) \in \mathcal{A}$.

NOTATION. We write

$$\mathbf{f}(x,a) = \begin{bmatrix} f_1(x,a) \\ \vdots \\ f_n(x,a) \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

□



Three responses to three controls

Given $T > 0$ and functions $r : \mathbb{R}^n \times A \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$, we define for each control $\alpha(\cdot) \in \mathcal{A}$ the **payoff functional**

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

where $\mathbf{x}(\cdot)$ solves (ODE) for the control $\alpha(\cdot)$. We call T the **terminal time**, r is the **running payoff**, and g is the **terminal payoff**.

OPTIMAL CONTROL PROBLEM. Our task is to design a control $\alpha_0(\cdot) \in \mathcal{A}$ that *maximizes* the payoff; thus

$$P[\alpha_0(\cdot)] = \max_{\alpha(\cdot)\in\mathcal{A}} P[\alpha(\cdot)].$$

This is an **optimal control problem**.

REMARK. More precisely, this is a *fixed time, free endpoint* optimal control problem, instances of which appear as the next two examples. Other problems are instead *free time, fixed endpoint*: see the third example following. □

EXAMPLE. (A production and consumption model) Consider an economic activity (such as running a company) that generates an output, some fraction of which we can at each moment reinvest, while consuming the rest. How should we plan our consumption/reinvestment strategy so as to maximize our total consumption over a time period of given length T?

To write down a mathematical model, introduce

$$\begin{aligned} x(t) &= \text{output of company at time } t \\ \alpha(t) &= \text{fraction of output reinvested at time } t. \end{aligned}$$

Since the control $\alpha(\cdot)$ represents a fraction of output, we have $0 \leq \alpha(t) \leq 1$ for times $0 \leq t \leq T$. In other words,

$$\alpha : [0, T] \to A$$

where A denotes the interval $[0, 1] \subset \mathbb{R}$.

Next we model the output of the company as a function of the reinvestment strategy:

$$\text{(ODE)} \qquad \begin{cases} \dot{x}(t) = \gamma\alpha(t)x(t) \quad (0 \leq t \leq T) \\ x(0) = x^0. \end{cases}$$

Here $\gamma > 0$ is a known growth rate. Since $(1 - \alpha(t))x(t)$ is the amount of the output consumed at time t, our total consumption will therefore be

$$\text{(P)} \qquad P[\alpha(\cdot)] = \int_0^T (1 - \alpha(t))x(t)\, dt.$$

We wish to design an optimal reinvestment plan $\alpha_0(\cdot)$ that maximizes $P[\,\cdot\,]$.

This fits into the fixed time, free endpoint control theory formulation from above, with $n = m = 1$ and

$$f(x, a) = kax, \quad r(x, a) = (1 - a)x, \quad g = 0.$$

□

EXAMPLE. (Linear-quadratic regulator) The linear-quadratic regulator is a widely used model, since, as we will later see, it is solvable.

In the simplest case, we take $n = m = 1$ and introduce the system dynamics

$$\text{(ODE)} \qquad \begin{cases} \dot{x}(t) = x(t) + \alpha(t) \quad (0 \leq t \leq T) \\ x(0) = x^0. \end{cases}$$

We also assume $A = \mathbb{R}$; so there is no constraint on the magnitude of control.

We want to *minimize* the quadratic cost functional

$$\int_0^T x^2(t) + \alpha^2(t)\, dt.$$

Since our theory is based upon maximization, we therefore take

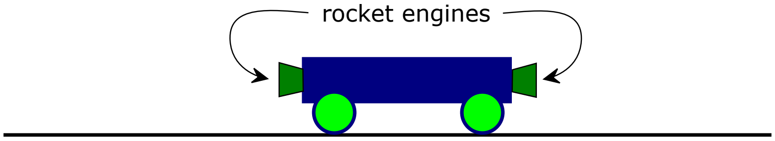
$$\text{(P)} \qquad P[\alpha(\cdot)] = -\int_0^T x^2(t) + \alpha^2(t)\, dt.$$

This falls into the control theory framework, as a fixed time, free endpoint problem, with

$$f(x,a) = x + a, \quad r(x,a) = -(x^2 + a^2), \quad g = 0.$$

□

EXAMPLE. (Rocket railroad car) We study next a railway car that can move along the real line, and whose acceleration can be adjusted by firing rockets at each end of the car. How can we steer the car to the origin in the least amount of time?



We introduce

$y(t)$ = position at time t

$\dot{y}(t)$ = velocity

$\ddot{y}(t)$ = acceleration

$\alpha(t)$ = thrust of rocket engines

T = time the car arrives at the origin, with zero velocity.

We assume concerning the trust that in appropriate physical units we have $-1 \le \alpha(t) \le 1$; consequently $\alpha : [0,T] \to A$ for $A = [-1,1]$. If the car has mass 1, then Newton's Law tells us that

$$\ddot{y}(t) = \alpha(t).$$

We rewrite this problem into the general form discussed before, setting $n = 2, m = 1$. We put

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix}.$$

Then our dynamics are

$$\text{(ODE)} \qquad \dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \alpha(t) \quad (0 \le t \le T)$$

with $\mathbf{x}(0) = x^0 = [y^0\, v^0]^T$, where y^0 is the initial position and v^0 is the initial velocity. Our goal is to steer the railway car to the origin at a time $T > 0$ (so it arrives with zero velocity: $\mathbf{x}(T) = [0\,0]^T$), and to do so in the least time.

We consequently want to maximize

$$\text{(P)} \qquad P[\alpha(\cdot)] = -\int_0^T 1\, dt = -T.$$

This is a free time, fixed endpoint problem, since the time T to reach the origin is not prescribed. □

4.2. Time optimal linear control

When the dynamics are linear in both the state and the control, we can use tools of linear and convex analysis to design explicit optimal controls and/or to deduce detailed information. In this section we illustrate such an approach. (However, our presentation will invoke ideas from measure theory and functional analysis, and will also omit some details.)

Consider therefore the linear control system:

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

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

$$A = [-1, 1]^m \subset \mathbb{R}^m,$$

and consider the class of admissible controls

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

Define

$$\text{(P)} \qquad P[\alpha(\cdot)] = -T$$

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

OPTIMAL TIME LINEAR PROBLEM. We are given the starting point $x^0 \in \mathbb{R}^n$, and want to find an optimal control $\alpha_0(\cdot)$ such that

$$P[\alpha_0(\cdot)] = \max_{\alpha(\cdot)\in\mathcal{A}} P[\alpha(\cdot)].$$

Then

$$T_0 = -\mathcal{P}[\alpha_0(\cdot)] \text{ is the minimum time to steer to the origin.}$$

4.2.1. Linear systems of ODE.

Let us first briefly recall some terminology and basic facts about linear systems of ordinary differential equations.

DEFINITION. Assume $M \in \mathbb{M}^{n\times n}$. Let $\mathbf{X}(\cdot) : [0,\infty) \to \mathbb{M}^{n\times n}$ be the unique solution of the matrix ODE

$$\begin{cases} \dot{\mathbf{X}}(t) = M\mathbf{X}(t) & (t \geq 0) \\ \mathbf{X}(0) = I. \end{cases}$$

We call $\mathbf{X}(\cdot)$ the **fundamental solution**, and sometimes write

$$\mathbf{X}(t) = e^{tM} = \sum_{k=0}^{\infty} \frac{t^k M^k}{k!}.$$

THEOREM 4.2.1. (Solving linear systems)

(i) The unique solution of the homogeneous initial-value problem

$$\begin{cases} \dot{\mathbf{x}} = M\mathbf{x} & (t \geq 0) \\ \mathbf{x}(0) = x^0 \end{cases} \tag{4.1}$$

is

$$\mathbf{x}(t) = \mathbf{X}(t)x^0.$$

(ii) Suppose that $\mathbf{f} : [0,\infty) \to \mathbb{R}^n$. Then the unique solution of the nonhomogeneous initial-value problem

$$\begin{cases} \dot{\mathbf{x}} = M\mathbf{x} + \mathbf{f} & (t \geq 0) \\ \mathbf{x}(0) = x^0. \end{cases} \tag{4.2}$$

is given by the **variation of parameters formula**

$$\mathbf{x}(t) = \mathbf{X}(t)x^0 + \mathbf{X}(t)\int_0^t \mathbf{X}^{-1}(s)\mathbf{f}(s)\,ds.$$

4.2.2. Reachable sets and convexity.

DEFINITION. We define the **reachable set** for time $t > 0$ to be

$$K(t,x^0) = \{x^1 \in \mathbb{R}^n \mid \text{ there exists } \alpha(\cdot) \in \mathcal{A} \text{ such that the corresponding solution of (ODE) satisfies } \mathbf{x}(t) = x^1\}.$$

In other words, $x^1 \in K(t, x^0)$ provided there exists an admissible control that steers the solution of (ODE) from x^0 to x^1 at time t. Using the variation of parameters formula, we see that $x^1 \in K(t, x^0)$ if and only if

$$x^1 = \mathbf{X}(t)x^0 + \mathbf{X}(t)\int_0^t \mathbf{X}^{-1}(s)N\alpha(s)\,ds \tag{4.3}$$

for some control $\alpha(\cdot) \in \mathcal{A}$.

The geometry of the reachable set is important:

THEOREM 4.2.2. For each time $t > 0$, the reachable set $K(t, x^0)$ is convex and closed.

Proof. 1. Let $x^1, x^2 \in K(t, x^0)$. Then there exist controls $\alpha^1(\cdot), \alpha^2(\cdot) \in \mathcal{A}$ such that

$$\begin{aligned} x^1 &= \mathbf{X}(t)x^0 + \mathbf{X}(t)\int_0^t \mathbf{X}^{-1}(s)N\alpha^1(s)\,ds \\ x^2 &= \mathbf{X}(t)x^0 + \mathbf{X}(t)\int_0^t \mathbf{X}^{-1}(s)N\alpha^2(s)\,ds. \end{aligned}$$

Let $0 \leq \theta \leq 1$. Then

$$\theta x^1 + (1-\theta)x^2 = \mathbf{X}(t)x^0 + \mathbf{X}(t)\int_0^t \mathbf{X}^{-1}(s)N(\theta\alpha^1(s) + (1-\theta)\alpha^2(s))\,ds.$$

Since $\theta\alpha^1(\cdot) + (1-\theta)\alpha^2(\cdot) \in \mathcal{A}$, we see that $\theta x^1 + (1-\theta)x^2 \in K(t, x^0)$.

2. We omit the proof that $K(t, x^0)$ is closed, as this requires some knowledge of functional analysis. □

We next exploit the convexity of reachable sets, to deduce nontrivial information about the structure of an optimal control.

THEOREM 4.2.3. (Time optimal linear maximum principle) Assume that $\alpha_0(\cdot)$ is a piecewise continuous optimal control, which steers the system from x^0 to 0 in the least time T_0.

Then there exists a nonzero vector $h \in \mathbb{R}^n$ such that

$$\boxed{h \cdot \mathbf{X}^{-1}(t)N\alpha_0(t) = \max_{a \in A}\{h \cdot \mathbf{X}^{-1}(t)Na\}} \tag{M}$$

for each time $0 \leq t \leq T_0$ that is a point of continuity of $\alpha_0(\cdot)$.

INTERPRETATION. Note that the maximum on the right hand side of (M) is over the finite dimensional set $A = [-1,1]^m$ of control values, and *not* over the infinite dimensional set $\mathcal{A}$ of all admissible controls. And in fact, since the expression $h \cdot \mathbf{X}^{-1}(t)Na$ is linear in a, the maximum will occur among the finitely many corners of the cube A.

The significance is that if we know h, then the maximization principle (M) provides us with a formula for computing $\alpha_0(\cdot)$, or at least for extracting useful information. See the example below for how this works in practice.

We will see later that (M) is a special case of the general Pontryagin Maximum Principle. □

Outline of proof 1. Since T_0 denotes the minimum time it takes to steer to 0, we have

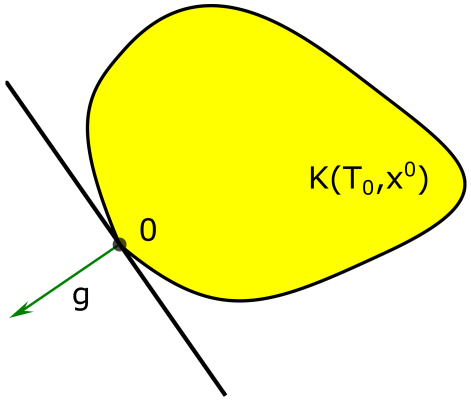
$$0 \notin K(t, x^0) \quad \text{for all times } 0 \leq t < T_0.$$

It follows that

$$0 \in \partial K(T_0, x^0). \tag{4.4}$$

Since $K(T_0, x^0)$ is a nonempty closed convex set, there exists a supporting plane to $K(T_0, x^0)$ at 0: see the Math 170 notes. This means that there exists a nonzero vector $g \in \mathbb{R}^n$, such that

$$g \cdot x \leq 0 \quad \text{for all } x \in K(T_0, x^0). \tag{4.5}$$



2. Given any control $\alpha(\cdot) \in \mathcal{A}$, define $x \in K(T_0, x^0)$ by

$$x = \mathbf{X}(T_0)x^0 + \mathbf{X}(T_0)\int_0^{T_0} \mathbf{X}^{-1}(s)N\alpha(s)\,ds.$$

Note also that

$$0 = \mathbf{X}(T_0)x^0 + \mathbf{X}(T_0)\int_0^{T_0} \mathbf{X}^{-1}(s)N\alpha_0(s)\,ds.$$

Since $g \cdot x \leq 0$, we therefore have

$$\begin{aligned} &g \cdot \left(\mathbf{X}(T_0)x^0 + \mathbf{X}(T_0)\int_0^{T_0} \mathbf{X}^{-1}(s)N\alpha(s)\,ds\right) \\ &\qquad \leq 0 = g \cdot \left(\mathbf{X}(T_0)x^0 + \mathbf{X}(T_0)\int_0^{T_0} \mathbf{X}^{-1}(s)N\alpha_0(s)\,ds\right). \end{aligned}$$

Define

$$h = \mathbf{X}^T(T_0)g;$$

so that $h^T = g^T\mathbf{X}(T_0)$. Then

$$\int_0^{T_0} h^T\mathbf{X}^{-1}(s)N\alpha(s)\,ds \leq \int_0^{T_0} h^T\mathbf{X}^{-1}(s)N\alpha_0(s)\,ds,$$

and therefore

$$\int_0^{T_0} h \cdot \mathbf{X}^{-1}(s)N(\alpha_0(s) - \alpha(s))\,ds \geq 0 \tag{4.6}$$

for all controls $\alpha(\cdot) \in \mathcal{A}$.

3. Now select any time $0 < t < T_0$ and any value $a \in A$. We pick $\delta > 0$ so small that the interval $[t, t+\delta]$ lies in $[0, T_0]$. Define

$$\alpha(s) = \begin{cases} a & \text{if } t \leq s \leq t+\delta \\ \alpha_0(s) & \text{otherwise;} \end{cases}$$

then (4.6) implies

$$\frac{1}{\delta}\int_t^{t+\delta} h \cdot \mathbf{X}^{-1}(s)N(\alpha_0(s) - a)\,ds \geq 0 \tag{4.7}$$

We sent $\delta \to 0$, to deduce that if t is a point of continuity of $\alpha_0(\cdot)$, then

$$h \cdot \mathbf{X}^{-1}(t)N\alpha_0(t) \geq h \cdot \mathbf{X}^{-1}(t)Na \tag{4.8}$$

for all $a \in A$. This implies the maximization assertion (M). □

REMARKS. (i) This outline of the proof needs more details, in particular for the assertion (4.4) that 0 lies on the boundary of the reachable set.

(ii) If an optimal control $\alpha_0(\cdot)$ is measurable, but not necessarily piecewise continuous, the same proof shows that (M) holds for almost every point time t in the interval $[0, T_0]$. □

EXAMPLE. (Rocket railway car) For this problem, introduced on page 92, we have

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \alpha(t)$$

for $n = 2, m = 1$ and

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad A = [-1, 1].$$

According to the maximum principle (M), there exists a *nonzero* vector $h \in \mathbb{R}^2$ such that

$$h \cdot \mathbf{X}^{-1}(t) N \alpha_0(t) = \max_{|a| \leq 1} \left\{ h \cdot \mathbf{X}^{-1}(t) N a \right\} \tag{4.9}$$

for an optimal control $\alpha(\cdot)$. We will now extract from this the useful information that *an optimal control $\alpha_0(\cdot)$ takes on only the values ± 1 and switches between these values most once.*

We must first compute $\mathbf{X}(t) = e^{tM}$. To do so, we observe

$$M^0 = I, \ M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \ M^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = O;$$

and therefore $M^k = O$ for all $k \geq 2$, where O denotes the zero matrix. Consequently,

$$e^{tM} = I + tM = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Then

$$\mathbf{X}^{-1}(t) = \begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix};$$

$$h \cdot \mathbf{X}^{-1}(t) N = [h_1 \, h_2] \begin{bmatrix} -t \\ 1 \end{bmatrix} = -th_1 + h_2.$$

Thus (4.9) asserts

$$(-th_1 + h_2)\alpha_0(t) = \max_{|a| \leq 1} \{(-th_1 + h_2)a\}; \tag{4.10}$$

and this implies that

$$\alpha_0(t) = \operatorname{sgn}(-th_1 + h_2)$$

for the *sign function*

$$\operatorname{sgn} x = \begin{cases} 1 & (x > 0) \\ 0 & (x = 0) \\ -1 & (x < 0). \end{cases}$$

Therefore the optimal control $\alpha_0(\cdot)$ switches at most once; and if $h_1 = 0$, then $\alpha_0(\cdot)$ is constant. (With this information, it is not especially difficult to find optimal controls and trajectories: see page 140.) □

4.3. Pontryagin Maximum Principle

We turn now to general optimal control problems, and learn how to generalize the maximization condition (M) from Theorem 4.2.3.

4.3.1. Fixed time, free endpoint problems.

The dynamics for a fixed time optimal control problem read

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

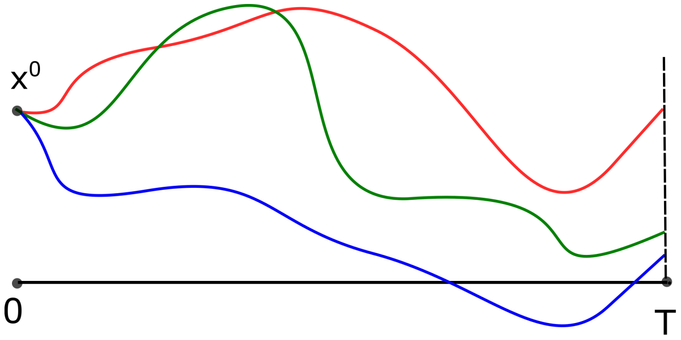
where $T > 0$ and

$$\mathcal{A} = \{\alpha : [0,T] \to A \mid \alpha(\cdot) \text{ is piecewise continuous}\}$$

for a given set $A \subseteq \mathbb{R}^m$. The payoff is

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

Our goal is to characterize an optimal control $\alpha_0(\cdot) \in \mathcal{A}$ that *maximizes* $P[\alpha(\cdot)]$ among all controls $\alpha(\cdot) \in \mathcal{A}$. This is a **fixed time, free endpoint** problem.



Fixed time, free endpoint trajectories

DEFINITION. The **control theory Hamiltonian** is

$$\boxed{H(x,p,a) = \mathbf{f}(x,a) \cdot p + r(x,a)}$$

for $x, p \in \mathbb{R}^n, a \in A$. That is,

$$H(x,p,a) = \sum_{j=1}^{n} f_j(x,a)p_j + r(x,a).$$

NOTATION. (i) We write

$$p = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}, \quad \nabla_x \mathbf{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{n \times n}$$

(ii) Since $\frac{\partial H}{\partial p_i} = f_i$ for $i = 1, \dots, n$, we have

$$\nabla_p H = \mathbf{f}. \tag{4.11}$$

Also, $\frac{\partial H}{\partial x_i} = \sum_{j=1}^n \frac{\partial f_j}{\partial x_i}(x,a)p_j + \frac{\partial r}{\partial x_i}(x,a)$ for $i = 1, \dots, n$. Consequently,

$$\nabla_x H = (\nabla_x \mathbf{f})^T p + \nabla_x r. \tag{4.12}$$

□

Next is our first version of the **Pontryagin Maximum Principle**.

THEOREM 4.3.1. (Fixed time, free endpoint PMP) Suppose $\alpha_0(\cdot)$ is an optimal control for the fixed time, free endpoint problem stated above, and $\mathbf{x}_0(\cdot)$ is the corresponding solution to (ODE).

(i) Then there exists a function $\mathbf{p}_0 : [0,T] \to \mathbb{R}^n$ such that for times $0 \le t \le T$ we have

$$\boxed{\dot{\mathbf{x}}_0(t) = \nabla_p H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t)\right),} \tag{ODE}$$

$$\boxed{\dot{\mathbf{p}}_0(t) = -\nabla_x H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t)\right),} \tag{ADJ}$$

$$\boxed{H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t)\right) = \max_{a \in A} H(\mathbf{x}_0(t), \mathbf{p}_0(t), a),} \tag{M}$$

and

$$\boxed{\mathbf{p}_0(T) = \nabla g\left(\mathbf{x}_0(T)\right).} \tag{T}$$

(ii) In addition,

$$H\left(\mathbf{x}_0, \mathbf{p}_0, \alpha_0\right) \text{ is constant on } [0,T]. \tag{4.13}$$

TERMINOLOGY. (i) We call

$$\mathbf{x}_0(t) = \begin{bmatrix} x_1^0(t) \\ \vdots \\ x_n^0(t) \end{bmatrix}, \quad \mathbf{p}_0(t) = \begin{bmatrix} p_1^0(t) \\ \vdots \\ p_n^0(t) \end{bmatrix}$$

the optimal **state** and **costate** at time t.

(ii) (ADJ) is the **adjoint equation**;

(iii) (M) is the **maximization condition**;

(iv) (T) is the **terminal** (or **transversality**) **condition**. □

REMARKS.

(i) To be more precise, (ODE), (ADJ) and (M) hold at times $0 < t < T$ that are points of continuity of the optimal control $\alpha_0(\cdot)$.

(ii) The most important assertion is (M). In practice, this usually allows us to transform the infinite dimensional problem of finding an optimal control $\alpha_0(\cdot) \in \mathcal{A}$ into a *finite dimensional* problem, at each time t, involving maximization over $A \subseteq \mathbb{R}^m$.

(iii) The costate equation (ADJ) and transversality condition (T) represent a sort of "back propagation" of information from the terminal time T. We can also interpret the costate as a Lagrange multiplier corresponding to the constraint that $\mathbf{x}_0(\cdot)$ solves (ODE).

Note that we specify the *initial* condition $\mathbf{x}_0(0) = x^0$ for (ODE) and the *terminal* condition $\mathbf{p}(T) = \nabla g(\mathbf{x}_0(T))$ for (ADJ). Hence even if $\alpha_0(\cdot)$ is known, solving this coupled system of equations can be tricky.

(iv) Remember from Section 1.4.3 that if $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $H = H(x, p)$, we call

$$\text{(H)} \qquad \begin{cases} \dot{\mathbf{x}} = \nabla_p H(\mathbf{x}, \mathbf{p}) \\ \dot{\mathbf{p}} = -\nabla_x H(\mathbf{x}, \mathbf{p}) \end{cases}$$

a *Hamiltonian system* of ODE. Notice that (ODE), (ADJ) are of this Hamiltonian form, except that now $H = H(x, p, a)$ depends also on the control. Observe furthermore that our assertion (4.13) is similar to (1.42) from Theorem 1.4.1.

□

4.3.2. Other terminal conditions.

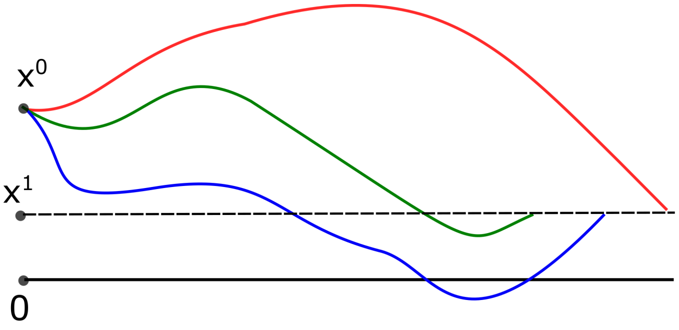
For another important class of optimal control problems, the dynamics are

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \alpha(t)) & (0 \le t \le T) \\ \mathbf{x}(0) = x^0, (T) = x^1, \end{cases} \tag{ODE}$$

where given initial and terminal values x^0 and x^1 are given, but *the terminal time $T > 0$ is not prescribed.* This is a **free time, fixed endpoint problem**, for which our goal is to maximize the payoff

$$P[\alpha(\cdot)] = \int_0^T r\left(\mathbf{x}(t), \alpha(t)\right)\, dt, \tag{P}$$

where $x(\cdot)$ solves (ODE). Our goal is to characterize an optimal control $\alpha_0(\cdot) \in \mathcal{A}$.



Free time, fixed endpoint trajectories

DEFINITION. The **extended Hamiltonian** is

$$\boxed{H(x,p,a,q) = \mathbf{f}(x,a) \cdot p + q r(x,a)} \tag{4.14}$$

for $x, p \in \mathbb{R}^n, a \in A, q \in \mathbb{R}$.

THEOREM 4.3.2. (Free time, fixed endpoint PMP) Suppose $\alpha_0(\cdot)$ is an optimal control for the free time, fixed endpoint control problem and $\mathbf{x}_0(\cdot)$ is the corresponding solution to (ODE), that arrives at the target point at time T_0. Let H be the extended Hamiltonian.

(i) Then there exists a function $\mathbf{p}_0 : [0, T_0] \to \mathbb{R}^n$ and a constant

$$q_0 = 0 \text{ or } 1, \tag{4.15}$$

such that for $0 \le t \le T_0$ we have

$$\dot{\mathbf{x}}_0(t) = \nabla_p H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t), q_0\right), \tag{ODE}$$

$$\text{(ADJ)} \qquad \dot{\mathbf{p}}_0(t) = -\nabla_x H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t), q_0\right),$$

and

$$\text{(M)} \qquad H\left(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t), q_0\right) = \max_{a \in A} H(\mathbf{x}_0(t), \mathbf{p}_0(t), a, q_0).$$

(ii) If $q_0 = 0$, then

$$(4.16) \qquad \mathbf{p}_0(\cdot) \text{ is not identically } 0 \text{ on } [0, T_0].$$

(iii) Furthermore,

$$\text{(T)} \qquad \boxed{H\left(\mathbf{x}_0, \mathbf{p}_0, \alpha_0, q_0\right) = 0 \;\text{ on } [0, T_0].}$$

REMARKS.

(i) So for the free time problem, we have the transversality condition that $H(\mathbf{x}_0, \mathbf{p}_0, \alpha_0, q_0) = 0$ at T_0 and thus $H(\mathbf{x}_0, \mathbf{p}_0, \alpha_0, q_0) = 0$ on the entire interval $[0, T_0]$. This generalization of our earlier Theorem 1.3.4 is stronger than the corresponding assertion (4.13) for the fixed time problem.

(ii) But we for the free time problem, we must deal with an additional Lagrange multiplier q_0. We say the free time problem is **normal** if $q_0 = 1$; it is **abnormal** if $q_0 = 0$. (The abnormal case is analogous to the existence of the abnormal Lagrange multiplier $\gamma_0 = 0$ in the F. John conditions for finite dimensional optimization theory. See the Math 170 notes for more on this.) □

Most free time problems are normal, and a simple assumption ensuring this follows.

LEMMA 4.3.1. Suppose that the controllability assumption

$$(4.17) \qquad \max_{a \in A}\{\mathbf{f}(x, a) \cdot p\} > 0 \qquad (x, p \in \mathbb{R}^n, p \neq 0)$$

holds.

Then the associated free time, fixed endpoint control problem is normal.

Proof. If the problem were abnormal, then $q_0 = 0$, $\mathbf{p}_0 \not\equiv 0$, and (T) would assert

$$\max_{a \in A} H(\mathbf{x}_0(t), \mathbf{p}_0(t), a, q_0) = \max_{a \in A}\{\mathbf{f}(\mathbf{x}_0(t), a) \cdot \mathbf{p}_0(t)\} = 0$$

on $[0, T_0]$. This however contradicts (4.17). □

EXAMPLE. Here is an example of an abnormal problem with $n = 2$, $m = 1$ and $A = [-1, 1]$. The dynamics are

$$\text{(ODE)} \qquad \begin{cases} \dot{x}_1 = \alpha^2 \\ \dot{x}_2 = 1 \end{cases}$$

with the initial and terminal conditions

$$\mathbf{x}(0) = x^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \mathbf{x}(T) = x^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The goal is to maximize

$$P[\alpha] = \int_0^T \alpha \, dt.$$

Now the only admissible control is $\alpha_0(\cdot) \equiv 0$, which is therefore optimal, and $T_0 = 1$. The free time Hamiltonian is

$$H = p_1 a^2 + p_2 + qa.$$

Then (M) implies

$$0 = \frac{\partial H}{\partial a} = 2p_1^0 \alpha_0 + q_0 = q_0$$

and hence this problem is abnormal. Furthermore (ADJ) tells us that $\mathbf{p}_0(\cdot)$ is constant. If we take

$$\mathbf{p}_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix},$$

then $\mathbf{p}_0 \neq 0$ and the conditions (M),(T) of Theorem 4.3.2 hold. This example of course fails to satisfy (4.17). □

EXAMPLE. Let us check that Theorem 4.3.2 accords with our previous maximum principle for the time optimal linear problem, as developed in Section 4.2. We have

$$H(x, p, a, q) = (Mx + Na) \cdot p - q \qquad (x, p \in \mathbb{R}^n, a \in A, q \in \mathbb{R}).$$

Select the vector h as in Theorem 4.2.3, and consider the system

$$\begin{cases} \dot{\mathbf{p}}_0(t) = -M^T \mathbf{p}_0(t) \\ \mathbf{p}_0(0) = h, \end{cases}$$

the solution of which is

$$\mathbf{p}_0(t) = \mathbf{X}^{-T}(t) h.$$

We know from condition (M) in Theorem 4.2.3 that

$$h \cdot \mathbf{X}^{-1}(t) N \alpha_0(t) = \max_{a \in A} \{ h \cdot \mathbf{X}^{-1}(t) N a \}.$$

But since $\mathbf{p}_0(t)^T = h^T\mathbf{X}^{-1}(t)$, this says

$$\mathbf{p}_0(t)\cdot N\alpha_0(t) = \max_{a\in A}\{\mathbf{p}_0(t)\cdot Na)\},$$

Then

$$\mathbf{p}_0(t)\cdot(M\mathbf{x}_0(t)+N\alpha_0(t)) - q_0 = \max_{a\in A}\{\mathbf{p}_0(t)\cdot(M\mathbf{x}_0(t)+Na)\} - q_0,$$

and this agrees with (M) from Theorem 4.3.2. □

Another sort of control problem has the dynamics

$$\text{(ODE)}\qquad \begin{cases}\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t),\alpha(t)) & (0\le t\le T)\\ \mathbf{x}(0)=x^0, (T)=x^1\end{cases}$$

for fixed initial and terminal values x^0 and x^1 are given and a fixed terminal time $T>0$. This is a **fixed time, fixed endpoint problem**. The payoff is again

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

THEOREM 4.3.3. (Fixed time, fixed endpoint PMP) Suppose $\alpha_0(\cdot)$ is an optimal control for the fixed time, fixed endpoint control problem and $\mathbf{x}_0(\cdot)$ is the corresponding solution to (ODE). Let H be given by (4.14).

(i) Then there exists $\mathbf{p}_0 : [0,T]\to\mathbb{R}^n$ and

$$\text{(4.18)}\qquad q_0 = 0 \text{ or } 1,$$

such that for all $0\le t\le T$:

$$\text{(ODE)}\qquad \dot{\mathbf{x}}_0(t) = \nabla_p H\left(\mathbf{x}_0(t),\mathbf{p}_0(t),\alpha_0(t),q_0\right)$$

$$\text{(ADJ)}\qquad \dot{\mathbf{p}}_0(t) = -\nabla_x H\left(\mathbf{x}_0(t),\mathbf{p}_0(t),\alpha_0(t),q_0\right)$$

$$\text{(M)}\qquad H\left(\mathbf{x}_0(t),\mathbf{p}_0(t),\alpha_0(t),q_0\right) = \max_{a\in A} H(\mathbf{x}_0(t),\mathbf{p}_0(t),a,q_0)$$

(ii) Furthermore,

$$\text{(4.19)}\qquad H\left(\mathbf{x}_0,\mathbf{p}_0,\alpha_0,q_0\right) \text{ is constant on } [0,T].$$

REMARKS.

(i) There is no transversality condition (T), since the fixed time, fixed endpoint condition is too rigid to allow for any variations. As before, we call the problem **normal** if $q_0 = 1$ and **abnormal** if $q_0 = 0$.

(ii) We can deduce Theorem 4.3.3 from Theorem 4.3.2 by introducing a new variable and rewriting the dynamics as

$$\begin{cases} \dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{f}}(\bar{\mathbf{x}}, \alpha(t)) & (0 \le t \le T) \\ \bar{\mathbf{x}}(0) = \bar{x}^0, \ \bar{\mathbf{x}}(T) = \bar{x}^1, \end{cases}$$

for

$$\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ x_{n+1} \end{bmatrix}, \ \bar{\mathbf{f}} = \begin{bmatrix} \mathbf{f} \\ 1 \end{bmatrix}, \ \bar{x}^0 = \begin{bmatrix} x^0 \\ 0 \end{bmatrix}, \ \bar{x}^1 = \begin{bmatrix} x^1 \\ T \end{bmatrix}.$$

This gives a free time, fixed endpoint problem. Consequently there exist q_0 and $\bar{\mathbf{p}}_0 : [0, T_0] \to \mathbb{R}^{n+1}$ satisfying the conclusions of Theorem 4.3.2. We write

$$\bar{\mathbf{p}}_0 = \begin{bmatrix} \mathbf{p}_0 \\ p^0_{n+1} \end{bmatrix}.$$

Then (ADJ) implies p^0_{n+1} is constant on $[0, T]$. Hence we may deduce from (T) that $H(\mathbf{x}_0, \mathbf{p}_0, \alpha_0, q_0) = -p^0_{n+1}$ is constant on $[0, T]$. □

4.4. Applications

HOW TO USE THE PONTRYAGIN MAXIMUM PRINCIPLE

Step 1. Write down the Hamiltonian

$$H = \begin{cases} \mathbf{f}(x, a) \cdot p + r(x, a) & \text{for fixed time, free endpoint problems} \\ \mathbf{f}(x, a) \cdot p + q r(x, a) & \text{for fixed endpoint problems,} \end{cases}$$

and calculate

$$\frac{\partial H}{\partial x_i}, \ \frac{\partial H}{\partial p_i} \qquad (i = 1, \dots, n).$$

Step 2. Write down (ODE), (ADJ), (M) and, if appropriate, (T).

Step 3. Use the maximization condition (M) to compute, if possible, $\alpha_0(t)$ as a function of $\mathbf{x}_0(t), \mathbf{p}_0(t)$.

Step 4. Now try to solve the coupled equations (ODE), (ADJ) and (T), to find $\mathbf{x}_0(\cdot), \mathbf{p}_0(\cdot)$ (and q_0 for free time problems).

DEFINITION. If (M) does not uniquely determine $\alpha_0(\cdot)$ on some time interval (where Step 3 therefore fails), we call that part of the trajectory of $\mathbf{x}_0(\cdot)$ a **singular arc**. □

To simplify notation will mostly not write the subscripts "0" in the following examples:

4.4.1. Simple linear-quadratic regulator.

The dynamics for the simplest version of the linear-quadratic regulator read

$$\text{(ODE)} \qquad \begin{cases} \dot{x} = x + \alpha \\ x(0) = x^0 \end{cases}$$

for controls $\alpha : [0, T] \to \mathbb{R}$. Here $n = m = 1$. The values of the controls are unconstrained; that is, $A = \mathbb{R}$. The payoff is quadratic in the state and the control:

$$\text{(P)} \qquad P[\alpha(\cdot)] = -\int_0^T x^2 + \alpha^2 \, dt$$

We therefore have a fixed time problem, for

$$f = x + a, \quad r = -x^2 - a^2, \quad g = 0.$$

So the Hamiltonian is

$$H = f(x, a)p + r(x, a) = (x + a)p - x^2 - a^2.$$

Then

$$\frac{\partial H}{\partial p} = x + a, \quad \frac{\partial H}{\partial x} = p - 2x.$$

Consequently the equations for the PMP read

$$\text{(ADJ)} \qquad \dot{p} = -\frac{\partial H}{\partial x} = 2x - p$$

$$\text{(M)} \qquad H\left(x(t), p(t), \alpha(t)\right) = \max_{a \in \mathbb{R}} \{(x(t) + a)p(t) - x^2(t) - a^2\}$$

$$\text{(T)} \qquad p(T) = 0.$$

We start with (M), and compute the value of α by noting the (unconstrained) maximum occurs where $\frac{\partial H}{\partial a} = 0$. Since $\frac{\partial H}{\partial a} = p - 2a$, we see that

$$\text{(4.20)} \qquad \alpha = \frac{p}{2}.$$

We use this information to rewrite (ODE), (ADJ):

$$\text{(4.21)} \qquad \begin{cases} \dot{x} = x + \frac{p}{2} \\ \dot{p} = 2x - p, \end{cases}$$

with the initial condition $x(0) = x^0$ and the terminal condition $p(T) = 0$.

To solve (4.21), let us suppose that we can write

$$\text{(4.22)} \qquad p(t) = d(t)x(t) \qquad (0 \leq t \leq T),$$

for some function $d(\cdot)$ that we must find. To find an equation for $d(\cdot)$, we assume that (4.21) is valid and compute

$$\begin{aligned} \dot p &= \dot d x + d\dot x \\ 2x - p &= \dot d x + d\left(x + \frac{p}{2}\right) \\ (2-d)x &= \dot d x + d\left(x + \frac{dx}{2}\right). \end{aligned}$$

Cancelling the x, we discover that we should select $d(\cdot)$ to solve the **Riccati equation**

$$\text{(4.23)} \qquad \begin{cases} \dot d = 2 - 2d - \frac{d^2}{2} \\ d(T) = 0. \end{cases}$$

Conversely, we check that if $d(\cdot)$ solves this Riccati equation and (4.22) holds, then we have (4.21).

So we solve the terminal value problem for this nonlinear ODE, to get the function $d(\cdot)$. Recalling then (4.22) and (4.20), we set

$$\alpha_0(t) = \frac{1}{2} d(t) x_0(t) \qquad (0 \leq t \leq T),$$

to synthesize an optimal **feedback control**, where $x_0(\cdot)$ solves (ODE) for this control:

$$\begin{cases} \dot x_0 = x_0 + \frac{1}{2} d x_0 \\ x_0(0) = x^0. \end{cases}$$

REMARK. We can convert the Riccati equation (4.23) into a terminal value problem for a *linear* second-order ODE, by writing

$$d = \frac{2\dot b}{b},$$

where

$$\begin{cases} \ddot b = b - 2\dot b \\ b(T) = 1, \ \dot b(T) = 0. \end{cases}$$

□

4.4.2. Production and consumption.

Recall that for our production and consumption model on page 91, we have the dynamics

$$\text{(ODE)} \qquad \begin{cases} \dot x = \alpha x \\ x(0) = x^0, \end{cases}$$

where the control α takes values in $A = [0,1]$. We assume $x^0 > 0$, and for simplicity have taken the growth rate $\gamma = 1$. The payoff is

$$P[\alpha(\cdot)] = \int_0^T (1-\alpha)x\,dt. \tag{P}$$

This fits within our fixed time PMP setting, for $n = m = 1$ and

$$f = ax, \quad r = (1-a)x, \quad g = 0.$$

Thus the Hamiltonian is

$$H = f(x,a)p + r(x,a) = axp + (1-a)x = x + ax(p-1);$$

and so

$$\frac{\partial H}{\partial p} = ax, \quad \frac{\partial H}{\partial x} = 1 + a(p-1).$$

Consequently,

$$\dot{p} = -\frac{\partial H}{\partial x} = -1 - \alpha(p-1). \tag{ADJ}$$

Furthermore, we have

$$H(x(t),p(t),\alpha(t)) = \max_{0\le a\le 1}\{x(t) + ax(t)(p(t)-1)\}; \tag{M}$$

$$p(T) = 0. \tag{T}$$

We now carry out the maximization in (M), to learn how to compute an optimal control $\alpha(\cdot)$ in terms of the other functions:

$$\alpha(t) = \begin{cases} 1 & \text{if } p(t) > 1 \\ 0 & \text{if } p(t) < 1. \end{cases} \tag{4.24}$$

This follows since $x(\cdot)$ is positive on $[0,T]$.

We next use the above information to find $x(\cdot), p(\cdot)$, and the idea is to work backwards from the ending time. Since $p(T) = 0$, it must be that $p(t) < 1$ for some interval $[t_0, T]$. Thus (4.24) implies $\alpha = 0$ for $t_0 \le t \le T$. We now analyze the various equations above on this interval (when $\alpha = 0$):

$$\dot{x} = 0, \tag{ODE}$$

$$\dot{p} = -1. \tag{ADJ}$$

It follows that $p(t) = T - t$ for $t_0 \le t \le T$ and $p(t_0) = 1$ for

$$t_0 = T - 1.$$

Next, we study the equations on the time interval $0 \le t \le t_0$:

$$\dot{x} = \alpha x, \tag{ODE}$$

$$\dot{p} = -1 - \alpha(p-1). \tag{ADJ}$$

We see that $p(t) > 1$ if $t_1 \leq t \leq t_0$ for some time $t_1 < t_0$. But then (4.24) says $\alpha = 1$ for $t_1 \leq t \leq t_0$, and therefore

$$\dot{p} = -1 - (p-1) = -p.$$

Since $p(t_0) = 1$, it follows that $p(t) = e^{t_0 - t} > 1$ for $t_1 \leq t < t_0$. Consequently $t_1 = 0$ and $p(t) > 1$ for $0 \leq t < t_0$.

So the optimal control is

$$\alpha_0(t) = \begin{cases} 1 & (0 \leq t < T-1) \\ 0 & (T-1 < t \leq T). \end{cases}$$

This means that we should reinvest all of the output until the time $t_0 = T-1$, and thereafter consume all the output.

REMARK. The formulas (ODE) and (ADJ) from the PMP provide us with explicit differential equations for the optimal states and costates, but *we do not in general have a differential equation for the corresponding optimal control.* Indeed, the production/consumption example above has a "bang-bang" control, which is piecewise constant with a single jump, and so does not solve any differential equation.

However the next two applications illustrate that we can sometimes establish ODE also for the controls. The idea will be to try to eliminate $\mathbf{p}(\cdot)$ from the various equations. □

4.4.3. Ramsey consumption model.

For this example, $x(t) \geq 0$ represents the capital at time t in some economy and the control $c(t) \geq 0$ is the consumption at time t. Given an initial amount of capital x^0, we want to maximize the utility of the total consumption over a fixed time interval $[0, T]$, but are required to leave an amount x^1 of capital at time T .

We model the evolution of the economy by the equation

$$\text{(ODE)} \qquad \begin{cases} \dot{x} = f(x) - c & (0 \leq t \leq T) \\ x(0) = x^0, x(T) = x^1 \end{cases}$$

for some appropriate function $f : [0, \infty) \to [0, \infty)$. So this is a fixed time, fixed endpoint problem, which we assume is normal.

We wish to find an optimal consumption plan to maximize the payoff

$$\text{(P)} \qquad P[c(\cdot)] = \int_0^T \psi(c)\, dt,$$

where $\psi : [0, \infty) \to [0, \infty)$, the consumption **utility function**, satisfies

$$\psi' > 0, \ \psi'' < 0.$$

We will not analyze this problem completely, but will show that we can derive an ODE for an optimal consumption policy. The Hamiltonian is

$$H = (f(x) - c)p + \psi(c),$$

and therefore

$$\text{(ADJ)} \qquad \dot{p} = -f'(x)p;$$

$$\text{(M)} \qquad H(x(t), p(t), c(t)) = \max_{c \geq 0}\{(f(x(t)) - c)p(t) + \psi(c)\}.$$

Now (M) implies for each time t that either

$$\text{(4.25)} \qquad c(t) = 0$$

or

$$\text{(4.26)} \qquad \psi'(c(t)) = p(t), \ c(t) > 0.$$

We ignore for the moment the first possibility and therefore suppose (4.26) always holds. This and (ADJ) now imply

$$\psi''(c)\dot{c} = \dot{p} = -f'(x)p = -f'(x)\psi'(c).$$

We thus obtain the **Keynes-Ramsey consumption rule**

$$\text{(4.27)} \qquad \dot{c} = -\frac{\psi'(c)}{\psi''(c)} f'(x).$$

Then (ODE) and (4.27) provide us with a coupled system of equations for an optimal control $c_0(\cdot)$ and the corresponding state $x_0(\cdot)$.

REMARKS. (i) Our analysis of this problem is however incomplete, since we have ignored the **state constraint** that $x(\cdot) \geq 0$. If the consumption plan $c(\cdot)$ computed above forces $x(\cdot)$ to start to go negative at some time t, we are clearly in the case (4.25), rather than (4.26).

(ii) We have also not specified an initial condition for (4.27). This would need to be selected so that $x_0(T) = x^1$. □

4.4.4. Zermelo's navigation problem.

Let $\mathbf{x} = [x_1\, x_2]^T$ denote the location of a boat moving at fixed speed V through water that has a current (depending on position, but not changing in time) given by the vector field $\mathbf{v} : \mathbb{R}^2 \to \mathbb{R}^2$, $\mathbf{v} = [v_1\, v_2]^T$. We control the boat by changing the direction in which it is pointing, determined by the angle ξ from due east. The dynamics are therefore

$$\text{(ODE)} \qquad \begin{cases} \dot{x}_1 = V\cos\xi + v_1(x_1, x_2) \\ \dot{x}_2 = V\sin\xi + v_2(x_1, x_2). \end{cases}$$

How to we adjust the angle $\xi(\cdot)$ so as to steer the boat between two given points x^0, x^1 in the least time?

This is a free time, fixed endpoint problem for which the control is $\xi(\cdot)$. We assume the problem is normal, and so the Hamiltonian is

$$H(x,p,\xi) = (V\cos\xi + v_1)p_1 + (V\sin\xi + v_2)p_2 - 1.$$

Consequently

$$\text{(ADJ)} \qquad \begin{cases} \dot{p}_1 = -p_1\frac{\partial v_1}{\partial x_1} - p_2\frac{\partial v_2}{\partial x_1} \\ \dot{p}_2 = -p_1\frac{\partial v_1}{\partial x_2} - p_2\frac{\partial v_2}{\partial x_2}. \end{cases}$$

Furthermore, the maximization condition (M) implies

$$0 = \frac{\partial H}{\partial \xi} = V(-p_1\sin\xi + p_2\cos\xi).$$

Therefore

$$\xi = \arctan\frac{p_2}{p_1}, \tag{4.28}$$

$$\sin\xi = \frac{p_2}{(p_1^2+p_2^2)^{\frac{1}{2}}}, \quad \cos\xi = \frac{p_1}{(p_1^2+p_2^2)^{\frac{1}{2}}}. \tag{4.29}$$

For this problem, it turns out that we can eliminate the costates p_1, p_2 and so express the optimal dynamics in terms of x_1, x_2 and ξ. To do so, let us use (4.28) and (ADJ) to compute

$$\begin{aligned} \dot{\xi} &= \frac{1}{1+\left(\frac{p_2}{p_1}\right)^2}\frac{\dot{p}_2 p_1 - p_2\dot{p}_1}{p_1^2} \\ &= \frac{p_1}{p_1^2+p_2^2}\left(-p_1\frac{\partial v_1}{\partial x_2} - p_2\frac{\partial v_2}{\partial x_2}\right) - \frac{p_2}{p_1^2+p_2^2}\left(-p_1\frac{\partial v_1}{\partial x_1} - p_2\frac{\partial v_2}{\partial x_1}\right). \end{aligned}$$

Then (4.29) implies

$$\dot{\xi} = \sin^2\xi\frac{\partial v_2}{\partial x_1} + \sin\xi\cos\xi\left(\frac{\partial v_1}{\partial x_1} - \frac{\partial v_2}{\partial x_2}\right) - \cos^2\xi\frac{\partial v_1}{\partial x_2}. \tag{4.30}$$

REMARK. The equations (ODE) and (4.30) provide us with a coupled system for the optimal control $\xi_0(\cdot)$ and optimal state $\mathbf{x}_0(\cdot)$. However we do not have the initial condition for (4.30) and so must presumably rely on numerical simulations to find a trajectory (if there is one) that passes through the target point x^1 at some time $T_0 > 0$. □

4.4.5. Chaplygin's navigation problem.

Here is another navigation problem. A boat takes a given time T to move at constant speed V around a closed loop in the ocean. Assuming that the sea water is flowing from west to east at constant speed $v < V$, what is the shape of such a path that encloses the maximum area?

If $\mathbf{x} = [x_1 \, x_2]^T$ denotes the location of the boat, its motion is determined by the equations

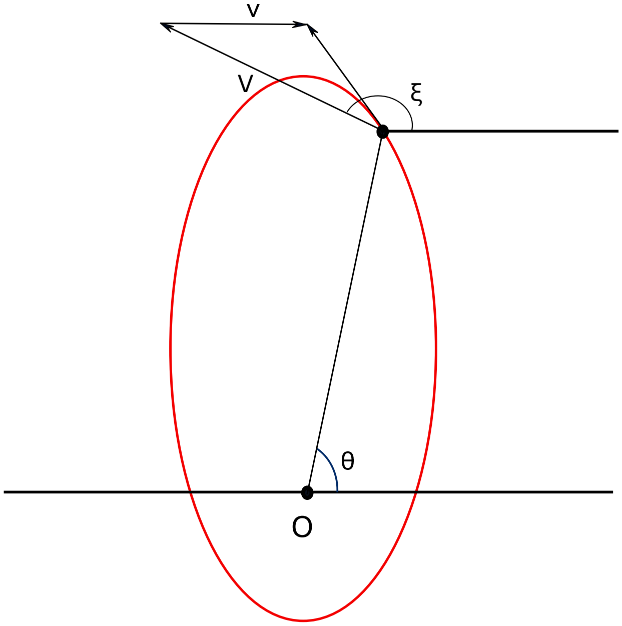
$$\text{(ODE)} \qquad \begin{cases} \dot{x}_1 = V \cos \xi + v \\ \dot{x}_2 = V \sin \xi, \end{cases}$$

where, as in the previous example, ξ is the angle from due east, as illustrated below.

We assume the path of the boat is a simple, closed curve, traversed counterclockwise. The area enclosed by the curve is

$$\text{(P)} \qquad P[\xi(\cdot)] = \frac{1}{2} \int x_1 dx_2 - x_2 dx_1 = \frac{1}{2} \int_0^T x_1 \dot{x}_2 - x_2 \dot{x}_1 \, dt.$$

So here $n = 2, m = 1$ and we have a fixed time and fixed endpoint problem, since the boat must begin and end its journey at some given point.



Chaplygin's problem

Assuming the problem to be normal, we see that the Hamiltonian is

$$\begin{aligned}H &= (V\cos\xi + v)p_1 + V\sin\xi\, p_2 + \frac{x_1}{2}V\sin\xi - \frac{x_2}{2}(V\cos\xi + v)\\ &= \left(p_1 - \frac{x_2}{2}\right)(V\cos\xi + v) + \left(p_2 + \frac{x_1}{2}\right)V\sin\xi.\end{aligned}$$

Therefore the adjoint dynamics are

$$\text{(ADJ)} \qquad \begin{cases}\dot p_1 = -\frac{V}{2}\sin\xi\\ \dot p_2 = \frac{1}{2}(V\cos\xi + v).\end{cases}$$

Using (ODE) and (ADJ), we see that

$$\dot p_1 + \frac{\dot x_2}{2} = 0,\ \dot p_2 - \frac{\dot x_1}{2} = 0.$$

Consequently

$$p_1 + \frac{x_2}{2} = a,\ p_2 - \frac{x_1}{2} = b$$

for appropriate constants a and b. Upon shifting coordinates if necessary, we may assume $a = b = 0$; so that

$$\text{(4.31)} \qquad p_1 + \frac{x_2}{2} = 0,\ p_2 - \frac{x_1}{2} = 0.$$

We next compute the optimal control angle from the maximization condition (M), by setting

$$0 = \frac{\partial H}{\partial \xi} = -\left(p_1 - \frac{x_2}{2}\right)V\sin\xi + \left(p_2 + \frac{x_1}{2}\right)V\cos\xi.$$

Then (4.31) implies

$$\text{(4.32)} \qquad x_1\cos\xi + x_2\sin\xi = 0.$$

We now switch to polar coordinates, by writing

$$\text{(4.33)} \qquad x_1 = r\cos\theta,\ x_2 = r\sin\theta,$$

where θ is the polar angle, as drawn. Then (4.32) tells us that

$$0 = \cos\theta\cos\xi + \sin\theta\sin\xi = \cos(\xi - \theta).$$

Therefore $\xi - \theta$ is an odd multiple of $\frac{\pi}{2}$; and so, from the picture,

$$\text{(4.34)} \qquad \xi = \theta + \frac{\pi}{2}.$$

So the optimal control is to steer at right angles to the polar angle θ.

We show next that the optimal path is an ellipse. To do so, we first differentiate (4.33):

$$\dot x_1 = \dot r\cos\theta - r\dot\theta\sin\theta;$$

$$\dot x_2 = \dot r\sin\theta + r\dot\theta\cos\theta.$$

This implies

$$\dot{r} = \dot{x}_1 \cos\theta + \dot{x}_2 \sin\theta.$$

Now use (4.34) to compute

$$\begin{aligned}\dot{r} - \frac{v}{V}\dot{x}_2 &= \dot{x}_1 \cos\theta + \dot{x}_2 \sin\theta - \frac{v}{V} V \sin\xi \\ &= (V\cos\xi + v)\cos\theta + V\sin\xi\sin\theta - v\sin\xi \\ &= (V\cos\xi + v)\sin\xi - V\sin\xi\cos\xi - v\sin\xi \\ &= 0.\end{aligned}$$

Hence for some constant γ, we have

$$r - ex_2 = \gamma,$$

where $e = \frac{v}{V}$. Thus the motion lies on the projection into $\mathbb{R}^2$ of the intersection in $\mathbb{R}^3$ of the cone $x_3 = r$ with the plane $x_3 = ex_2 + \gamma$. This is an ellipse, since $e < 1$.

REMARKS. If $v = 0$, the motion of the boat is a circle. We have thus in particular shown that *among smooth curves of fixed length, a circle encloses the maximum area.* (More precisely, we have shown that if there exists a smooth minimizer, it is a circle.) Compare this assertion with the isoperimetric problem discussed on page 25. □

4.4.6. Optimal harvesting.

A simple ODE model for the population x of, say, fish in a lake is

$$\dot{x} = \gamma x \left(1 - \frac{x}{k}\right),$$

where $\gamma > 0$ is the growth rate and $k > 0$ is the long term carrying capacity. As $t \to \infty$ all positive solutions converge to the equilibrium level k.

Suppose now that we continually harvest the populations:

$$\dot{x} = \gamma x \left(1 - \frac{x}{k}\right) - q\alpha x,$$

where $0 \leq \alpha \leq 1$ represents our fishing effort and $q > 0$ its effectiveness. Thus $h = q\alpha x$ is the harvest amount. The corresponding economic payoff over a given time period $[0, T]$ is therefore

$$P[\alpha(\cdot)] = \int_0^T ph - \theta\alpha \, dt$$

where the constant $p > 0$ is the fixed price for fish and the constant $\theta > 0$ represents a cost rate for our fishing efforts. We suppose the initial fish population is x^0, and we also require that after the fishing season is over, the population of fish in the lake should be restored to a prescribed level x^1.

We rescale and make various simplifying choices of the parameters, to reduce to the dynamics

$$\text{(ODE)}\qquad \begin{cases} \dot{x} = x(1-x) - q\alpha x \\ x(0) = x^0, \quad x(T) = x^1. \end{cases}$$

The payoff functional is

$$\text{(P)}\qquad P[\alpha(\cdot)] = \int_0^T (x-\theta)\alpha \, dt.$$

We will assume

$$\text{(4.35)}\qquad 0 < \theta < 1, \quad q > \frac{1}{2}.$$

We say that a control $\alpha(\cdot)$ is admissible if it satisfies the constraints $0 \leq \alpha(\cdot) \leq 1$ and the corresponding solution $x(\cdot)$ of the differential equation in (ODE) with $x(0) = x^0$ satisfies the terminal condition $x(T) = x^1$.

We wish to characterize an optimal fishing effort that maximizers the payoff, subject to the dynamics (ODE). This is a fixed time, fixed endpoint problem.

We assume our problem is normal, and consequently the Hamiltonian is

$$H = (x(1-x) - qax)p + (x-\theta)a.$$

Hence the adjoint dynamics are

$$\text{(ADJ)}\qquad \dot{p} = -(1 - 2x - q\alpha)p - \alpha$$

and the maximization condition is

$$\text{(M)}\qquad H(x(t), p(t), \alpha(t)) = \max_{0 \leq a \leq 1} \{a(-qp(t)x(t) + x(t) - \theta)\}.$$

Equilibrium solutions. Let us first look for equilibrium solutions of the above, that is, solutions of the form $x(\cdot) \equiv x_*, p(\cdot) \equiv p_*, \alpha(\cdot) \equiv a_*$ for constants x_*, p_*, a_* with $0 < a_* < 1$. This will be an algebra execise. In this case, (ODE) and (ADJ) imply

$$\text{(4.36)}\qquad 1 - x_* - qa_* = 0, \quad (1 - 2x_* - qa_*)p_* + a_* = 0;$$

and, since $0 < a_* < 1$ solves (M), we must have

$$\text{(4.37)}\qquad -qp_*x_* + x_* - \theta = 0.$$

The two equations (4.36) give $a_* = p_*x_*$. Plugging this into (4.37) yields $a_* = \frac{x_*-\theta}{q}$. Using this back in (4.36) and simplifying, we find

$$\text{(4.38)}\qquad x_* = \frac{1+\theta}{2}, \quad p_* = \frac{1-\theta}{q(1+\theta)}, \quad a_* = \frac{1-\theta}{2q}.$$

Owing to (4.35), we have $0 < a_* < 1$, as required: we can interpret $\alpha(\cdot) \equiv a_*$ as a sustainable fishing policy. Note also, for future reference, that

$$x_* \text{ gives the maximum of the quadratic function } (x - \theta)(1 - x).$$

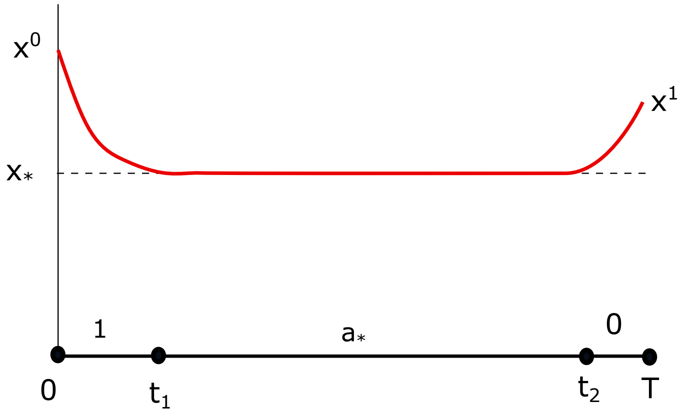
Most rapid approach path. We propose now to employ the constants x_*, p_*, a_* to build a general, nonequilibrium solution of our harvesting problem. To be specific, let us suppose $x^0 > x_*$ and $x^1 > x_*$. We now find, if we can, the first time $0 \leq t_1 \leq T$ so that the solution of (ODE) with $x(0) = x^0$ and $\alpha \equiv 1$ on the time interval $[0, t_1]$ satisfies

$$x(t_1) = x_*.$$

We also find, if possible, a time $t_1 \leq t_2 \leq T$ so that the solution of solution of (ODE) with $x(T) = x^1$ and $\alpha \equiv 0$ satisfies

$$x(t_2) = x_*.$$

(We assume that the values of $x(\cdot)$ are between 0 and 1, in appropriate units. Thus the fish population will rise if $\alpha \equiv 0$.)

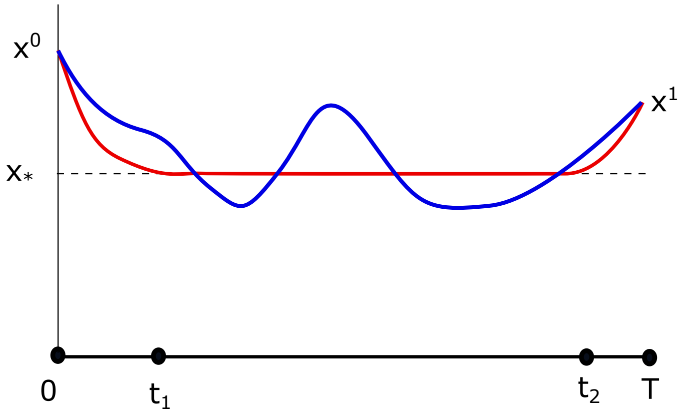


Optimal harvesting

We claim now that the optimal control is

$$\alpha_0(t) = \begin{cases} 1 & (0 \leq t \leq t_1) \\ a_* & (t_1 \leq t \leq t_2) \\ 0 & (t_2 \leq t \leq T). \end{cases}$$

To see this, consider another admissible control $\alpha : [0, T] \to [0, 1]$ and let $x(\cdot)$ denote the corresponding solution of (ODE).



Another harvesting plan

Now

$$P[\alpha(\cdot)] = \int_0^T (x-\theta)\alpha\, dt = \int_0^T (x-\theta)\left(\frac{x(1-x)-\dot{x}}{qx}\right) dt.$$

The part of integrand that involves $\dot{x}$ is a null Lagrangian, and consequently that part of the payoff depends only upon the fixed boundary values x^0 and x^1. That is,

$$\int_0^T \frac{x-\theta}{qx}\dot{x}\, dt = \frac{x^1 - \theta \log x^1}{q} - \frac{x^0 - \theta \log x^0}{q} = C;$$

and hence

$$P[\alpha(\cdot)] = \frac{1}{q}\int_0^T (x-\theta)(1-x)\, dt - C.$$

Let $x_0(\cdot)$ be the dynamics corresponding to $\alpha_0(\cdot)$. Since $\alpha_0 \equiv 1$ on $(0, t_1)$, we have $x(t) \geq x_0(t) \geq x_*$ on $[0, t_1]$; see the illustration. Since x_* gives the maximum of the quadratic $(x-\theta)(1-x)$, it follow that

$$\int_0^{t_1} (x_0-\theta)(1-x_0)\, dt \geq \int_0^{t_1} (x-\theta)(1-x)\, dt.$$

Similarly, $x(t) \geq x_0(t) \geq x_*$ on $[t_2, T]$, and consequently

$$\int_{t_2}^T (x_0-\theta)(1-x_0)\, dt \geq \int_{t_2}^T (x-\theta)(1-x)\, dt.$$

Furthermore,

$$\int_{t_1}^{t_2} (x_0-\theta)(1-x_0)\, dt \geq \int_{t_1}^{t_2} (x-\theta)(1-x)\, dt,$$

since $x_0(\cdot) \equiv x_*$ on this interval and x_* gives the maximum of the integrand. Hence

$$P[\alpha_0(\cdot)] \geq P[\alpha(\cdot)].$$

ECONOMIC INTERPRETATION. Observe that on $[t_1, t_2]$ we have a singular arc (see page 106), since the maximization condition (M) does not determine there the value of the optimal control.

This example, adapted from Mesterton-Gibbons [**MG**], is an instance of what economists call a **most rapid approach path**. It is optimal to move as quickly as possible to where $x = x_*$ and then to stay on this path, sometimes called a **turnpike**, as long as possible. □

REMARK. There are many more applications of the PMP discussed in the texts listed in the References. See in particular Kamien–Schwartz [**K-S**], Lee–Markus [**L-M**] and my old online lecture notes [**E**]. □

4.5. Proof of PMP

We present next a reasonably complete discussion of the ideas behind the derivation of the Pontryagin Maximum Principle.

4.5.1. Simple control variations.

Recall that the response $\mathbf{x}(\cdot)$ to a given control $\alpha(\cdot)$ is the unique solution of the system of differential equations:

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

We investigate in this section how certain simple changes in the control affect the response.

DEFINITION. Fix a time $s > 0$ and a control parameter value $a \in A$. Select $0 < \varepsilon < s$ and define then the modified control

$$\alpha_\varepsilon(t) = \begin{cases} a & \text{if } \ s - \varepsilon < t < s \\ \alpha(t) & \text{otherwise.} \end{cases}$$

We call $\alpha_\varepsilon(\cdot)$ a **simple** (or **needle**) **variation** of $\alpha(\cdot)$.

Let $\mathbf{x}_\varepsilon(\cdot)$ be the corresponding response to our system:

$$(ODE_\varepsilon) \qquad \begin{cases} \dot{\mathbf{x}}_\varepsilon(t) = \mathbf{f}(\mathbf{x}_\varepsilon(t), \alpha_\varepsilon(t)) & (t > 0) \\ \mathbf{x}_\varepsilon(0) = x^0. \end{cases}$$

We want to understand how our choices of s and a cause $\mathbf{x}_\varepsilon(\cdot)$ to differ from $\mathbf{x}(\cdot)$, for small $\varepsilon > 0$.

NOTATION. Define the matrix-valued function $\mathbf{A} : [0,\infty) \to \mathbb{M}^{n\times n}$ by

$$\mathbf{A}(t) = \nabla_x \mathbf{f}(\mathbf{x}(t), \alpha(t)).$$

So the (i,j)-th entry of $\mathbf{A}(t)$ is

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}(t), \alpha(t)) \qquad (i,j = 1,\dots,n).$$

□

We first quote a standard perturbation assertion for ordinary differential equations:

LEMMA 4.5.1. Let $\mathbf{y}_\varepsilon(\cdot)$ solve the initial-value problem:

$$\begin{cases} \dot{\mathbf{y}}_\varepsilon(t) = \mathbf{f}(\mathbf{y}_\varepsilon(t), \alpha(t)) \qquad (t \geq 0) \\ \mathbf{y}_\varepsilon(0) = x^0 + \varepsilon y^0 + o(\varepsilon). \end{cases}$$

Then

$$\mathbf{y}_\varepsilon(t) = \mathbf{x}(t) + \varepsilon \mathbf{y}(t) + o(\varepsilon) \quad \text{as } \varepsilon \to 0, \tag{4.39}$$

uniformly for t in compact subsets of $[0,\infty)$, where

$$\begin{cases} \dot{\mathbf{y}}(t) = \mathbf{A}(t)\mathbf{y}(t) \qquad (t \geq 0) \\ \mathbf{y}(0) = y^0. \end{cases} \tag{4.40}$$

NOTATION. We write

$$o(\varepsilon)$$

to denote any expression $\mathbf{g}_\varepsilon$ such that

$$\lim_{\varepsilon \to 0} \frac{\mathbf{g}_\varepsilon}{\varepsilon} = 0.$$

In words, if $\varepsilon \to 0$, then $\mathbf{g}_\varepsilon = o(\varepsilon)$ goes to zero "faster than ε". □

Returning now to the dynamics (ODE_ε), we establish

LEMMA 4.5.2. Assume that s is a point of continuity for the control $\alpha(\cdot)$. Then we have

$$\mathbf{x}_\varepsilon(t) = \mathbf{x}(t) + \varepsilon \mathbf{y}(t) + o(\varepsilon) \quad \text{as } \varepsilon \to 0, \tag{4.41}$$

uniformly for t in compact subsets of $[0,\infty)$, where

$$\mathbf{y}(t) = 0 \qquad (0 \leq t \leq s)$$

and

$$\begin{cases} \dot{\mathbf{y}}(t) = \mathbf{A}(t)\mathbf{y}(t) \qquad (t \geq s) \\ \mathbf{y}(s) = y^s, \end{cases} \tag{4.42}$$

for

$$y^s = \mathbf{f}(\mathbf{x}(s), a) - \mathbf{f}(\mathbf{x}(s), \alpha(s)). \tag{4.43}$$

NOTATION. We will sometimes write

$$\mathbf{y}(t) = \mathbf{Y}(t, s) y^s \qquad (t \geq s)$$

when (4.42) holds. □

Proof. Clearly $\mathbf{x}_\varepsilon(t) = \mathbf{x}(t)$ for $0 \leq t \leq s - \varepsilon$. For times $s - \varepsilon \leq t \leq s$, we have

$$\mathbf{x}_\varepsilon(t) - \mathbf{x}(t) = \int_{s-\varepsilon}^{t} \mathbf{f}(\mathbf{x}_\varepsilon(r), a) - \mathbf{f}(\mathbf{x}(r), \alpha(r))\, dr.$$

Thus

$$\mathbf{x}_\varepsilon(s) - \mathbf{x}(s) = [\mathbf{f}(\mathbf{x}(s), a) - \mathbf{f}(\mathbf{x}(s), \alpha(s))]\varepsilon + o(\varepsilon).$$

On the time interval $[s, \infty)$, $\mathbf{x}(\cdot)$ and $\mathbf{x}_\varepsilon(\cdot)$ both solve the same ODE, but with differing initial conditions given by

$$\mathbf{x}_\varepsilon(s) = \mathbf{x}(s) + \varepsilon y^s + o(\varepsilon),$$

for y^s defined by (4.43). According then to Lemma 4.5.1, we have

$$\mathbf{x}_\varepsilon(t) = \mathbf{x}(t) + \varepsilon \mathbf{y}(t) + o(\varepsilon) \qquad (t \geq s),$$

the function $\mathbf{y}(\cdot)$ solving (4.42). □

4.5.2. Fixed time problem.

Terminal payoff problem. We return to our usual dynamics

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

and introduce the terminal payoff functional

$$\text{(P)} \qquad P[\alpha(\cdot)] = g(\mathbf{x}(T)),$$

to be maximized. So for now we are taking the running payoff $r \equiv 0$.

We assume that $\alpha_0(\cdot)$ is an optimal control for this problem, corresponding to the optimal response $\mathbf{x}_0(\cdot)$. The control theory Hamiltonian is

$$H(x, p, a) = \mathbf{f}(x, a) \cdot p,$$

and our task is find $\mathbf{p}_0 : [0, T] \to \mathbb{R}^n$ such that (ADJ), (T) and (M) hold.

We reintroduce the function $\mathbf{A}(\cdot) = \nabla_x \mathbf{f}(\mathbf{x}_0(\cdot), \alpha_0(\cdot))$ and the control variation $\alpha_\varepsilon(\cdot)$, as in the previous section. We now define $\mathbf{p}_0 : [0,T] \to \mathbb{R}$ to be the unique solution of the terminal-value problem

$$\begin{cases} \dot{\mathbf{p}}_0(t) = -\mathbf{A}^T(t)\mathbf{p}_0(t) \qquad (0 \le t \le T) \\ \mathbf{p}_0(T) = \nabla g(\mathbf{x}_0(T)). \end{cases} \tag{4.44}$$

This gives (ADJ) and (T), and so our goal is to establish the maximization principle (M).

The main point is that $\mathbf{p}_0(\cdot)$ helps us calculate the variation of the terminal payoff:

LEMMA 4.5.3. Assume $0 < s < T$ is a point of continuity for $\alpha_0(\cdot)$. Then we have

$$\frac{d}{d\varepsilon} P[\alpha_\varepsilon(\cdot)]|_{\varepsilon=0} = [\mathbf{f}(\mathbf{x}_0(s), a) - \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s))] \cdot \mathbf{p}_0(s). \tag{4.45}$$

Proof. According to Lemma 4.5.2,

$$P[\alpha_\varepsilon(\cdot)] = g(\mathbf{x}_\varepsilon(T)) = g(\mathbf{x}_0(T) + \varepsilon \mathbf{y}(T) + o(\varepsilon)),$$

where $\mathbf{y}(\cdot)$ satisfies (4.42) for

$$y^s = \mathbf{f}(\mathbf{x}_0(s), a) - \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s)).$$

Thus

$$\frac{d}{d\varepsilon} P[\alpha_\varepsilon(\cdot)]|_{\varepsilon=0} = \nabla g(\mathbf{x}_0(T)) \cdot \mathbf{y}(T). \tag{4.46}$$

On the other hand, (4.42) and (4.44) imply

$$\begin{aligned} \frac{d}{dt}(\mathbf{p}_0(t) \cdot \mathbf{y}(t)) &= \dot{\mathbf{p}}_0(t) \cdot \mathbf{y}(t) + \mathbf{p}_0(t) \cdot \dot{\mathbf{y}}(t) \\ &= -\mathbf{A}^T(t)\mathbf{p}_0(t) \cdot \mathbf{y}(t) + \mathbf{p}_0(t) \cdot \mathbf{A}(t)\mathbf{y}(t) \\ &= 0. \end{aligned}$$

Hence

$$\nabla g(\mathbf{x}_0(T)) \cdot \mathbf{y}(T) = \mathbf{p}_0(T) \cdot \mathbf{y}(T) = \mathbf{p}_0(s) \cdot \mathbf{y}(s) = \mathbf{p}_0(s) \cdot y^s.$$

Since $y^s = \mathbf{f}(\mathbf{x}_0(s), a) - \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s))$, this identity and (4.46) imply (4.45). □

THEOREM 4.5.1. (PMP with no running costs) There exists a function $\mathbf{p}_0 : [0,T] \to \mathbb{R}^n$ satisfying the adjoint dynamics (ADJ), the maximization principle (M) and the terminal condition (T).

Proof. The adjoint dynamics and terminal condition are both in (4.44). To confirm (M), fix $0 < s < T$ and $a \in A$, as above. Since the mapping $\varepsilon \mapsto P[\alpha_\varepsilon(\cdot)]$ for $0 \leq \varepsilon \leq 1$ has a maximum at $\varepsilon = 0$, we deduce from Lemma 4.5.3 that

$$0 \geq \frac{d}{d\varepsilon} P[\alpha_\varepsilon(\cdot)] = [\mathbf{f}(\mathbf{x}_0(s), a) - \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s)] \cdot \mathbf{p}_0(s).$$

Hence

$$\begin{aligned} H(\mathbf{x}_0(s), \mathbf{p}_0(s), a) &= \mathbf{f}(\mathbf{x}_0(s), a) \cdot \mathbf{p}_0(s) \\ &\leq \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s)) \cdot \mathbf{p}_0(s) = H(\mathbf{x}_0(s), \mathbf{p}_0(s), \alpha_0(s)) \end{aligned}$$

for all $a \in A$ and each time $0 < s < T$ that is a point of continuity for $\alpha_0(\cdot)$. This proves the maximization condition (M). □

General payoff problem. We next extend our analysis, to cover the case that the payoff functional includes also a running payoff:

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

The control theory Hamiltonian is now

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

and we must manufacture a costate function $\mathbf{p}_0(\cdot)$ satisfying (ADJ), (M) and (T).

Adding a new variable. The trick is to introduce another variable and thereby convert to the previous case. We consider the function $x_{n+1} : [0, T] \to \mathbb{R}$ given by

$$\text{(4.47)} \qquad \begin{cases} \dot{x}_{n+1}(t) = r(\mathbf{x}(t), \alpha(t)) & (0 \leq t \leq T) \\ x_{n+1}(0) = 0, \end{cases}$$

where $\mathbf{x}(\cdot)$ solves (ODE). It follows that

$$x_{n+1}(T) = \int_0^T r(\mathbf{x}(t), \alpha(t))\, dt.$$

Introduce next the new notation

$$\bar{x} = \begin{bmatrix} x \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix}, \ \bar{x}^0 = \begin{bmatrix} x^0 \\ 0 \end{bmatrix}, \ \bar{p} = \begin{bmatrix} p \\ p_{n+1} \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \\ p_{n+1} \end{bmatrix},$$

$$(4.48) \quad \bar{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ x_{n+1}(t) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \\ x_{n+1}(t) \end{bmatrix}, \ \bar{\mathbf{f}}(\bar{x}, a) = \begin{bmatrix} \mathbf{f}(x,a) \\ r(x,a) \end{bmatrix} = \begin{bmatrix} f_1(x,a) \\ \vdots \\ f_n(x,a) \\ r(x,a) \end{bmatrix}$$

and

$$(4.49) \qquad \bar{g}(\bar{x}) = g(x) + x_{n+1}.$$

Then (ODE) and (4.47) produce the dynamics

$$(\overline{\text{ODE}}) \qquad \begin{cases} \dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{f}}(\bar{\mathbf{x}}(t), \alpha(t)) & (0 \le t \le T) \\ \bar{\mathbf{x}}(0) = \bar{x}^0. \end{cases}$$

Consequently our control problem transforms into a new problem with no running payoff and the terminal payoff functional

$$\bar{P}[\alpha(\cdot)] = \bar{g}(\bar{\mathbf{x}}(T)).$$

THEOREM 4.5.2. (PMP for fixed time, free endpoint problem) There exists a function $\mathbf{p}_0 : [0,T] \to \mathbb{R}^n$ satisfying the adjoint dynamics (ADJ), the maximization principle (M) and the terminal condition (T).

Proof. We apply Theorem 4.5.1, to obtain $\bar{\mathbf{p}}_0 : [0,T] \to \mathbb{R}^{n+1}$ satisfying ($\overline{\text{M}}$) for the Hamiltonian

$$\bar{H}(\bar{x}, \bar{p}, a) = \bar{\mathbf{f}}(\bar{x}, a) \cdot \bar{p}.$$

Also the adjoint equations ($\overline{\text{ADJ}}$) hold, with the terminal transversality condition

$$\bar{\mathbf{p}}_0(T) = \nabla \bar{g}(\bar{\mathbf{x}}_0(T)).$$

But $\bar{\mathbf{f}}$ does not depend upon the variable x_{n+1}, and consequently the last equation in ($\overline{\text{ADJ}}$) reads

$$\dot{p}_0^{n+1}(t) = -\frac{\partial \bar{H}}{\partial x_{n+1}} = 0.$$

Since $\frac{\partial \bar{g}}{\partial x_{n+1}} = 1$, we deduce that

$$p_{n+1}^0(t) \equiv 1.$$

As the last component of the vector function $\bar{\mathbf{f}}$ is r, we then conclude from (8.11) that

$$\bar{H}(\bar{x}, \bar{p}, a) = \mathbf{f}(x,a) \cdot p + r(x,a) = H(x,p,a).$$

Therefore

$$\mathbf{p}_0(t) = \begin{bmatrix} p_1^0(t) \\ \vdots \\ p_n^0(t) \end{bmatrix}$$

satisfies (ADJ), (M) for the Hamiltonian H. □

4.5.3. Multiple control variations.

Proving the PMP for the free time, fixed endpoint problem is much more difficult, since the result of a simple variation as above may produce a response $\mathbf{x}_\varepsilon(\cdot)$ that never hits the target point x^1. We consequently need to introduce more complicated control variations, discussed in this section.

DEFINITION. Let us select times $0 < s_1 < s_2 < \cdots < s_N$, numbers $\lambda_1, \dots, \lambda_N > 0$, and also control parameters $a_1, a_2, \dots, a_N \in A$. Write

$$\alpha_\varepsilon(t) = \begin{cases} a_k & \text{if } s_k - \lambda_k\varepsilon \le t < s_k \quad (k = 1, \dots, N) \\ \alpha(t) & \text{otherwise,} \end{cases} \tag{4.50}$$

for $\varepsilon > 0$ taken so small that the intervals $[s_k - \lambda_k\varepsilon, s_k]$ do not overlap.

This is called a **multiple variation** of the control $\alpha(\cdot)$.

We assume for this section that $\mathbf{x}(\cdot)$ solves

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \alpha(t)) & (0 \le t \le T) \\ \mathbf{x}(0) = x^0. \end{cases} \tag{4.51}$$

for some piecewise continuous control $\alpha(\cdot)$, and that $\mathbf{x}_\varepsilon(\cdot)$ is the response to $\alpha_\varepsilon(\cdot)$:

$$\begin{cases} \dot{\mathbf{x}}_\varepsilon(t) = \mathbf{f}(\mathbf{x}_\varepsilon(t), \alpha_\varepsilon(t)) & (0 \le t \le T) \\ \mathbf{x}_\varepsilon(0) = x^0. \end{cases} \tag{4.52}$$

NOTATION. (i) Define

$$y^{s_k} = \mathbf{f}(\mathbf{x}_0(s_k), a_k)) - \mathbf{f}(\mathbf{x}_0(s_k), \alpha_0(s_k))$$

for $k = 1, \dots, N$.

(ii) As before, set $\mathbf{A}(\cdot) = \nabla_x \mathbf{f}(\mathbf{x}_0(\cdot), \alpha_0(\cdot))$ and write

$$\mathbf{y}(t) = \mathbf{Y}(t,s) y^s \qquad (t \ge s)$$

to denote the solution of

$$\begin{cases} \dot{\mathbf{y}}(t) = \mathbf{A}(t)\mathbf{y}(t) & (t \ge s) \\ \mathbf{y}(s) = y^s, \end{cases}$$

where $y^s \in \mathbb{R}^n$ is given.

Suitably modifying the proof of Lemma 4.5.2, we can establish

LEMMA 4.5.4. We have

$$\mathbf{x}_\varepsilon(t) = \mathbf{x}(t) + \varepsilon \mathbf{y}(t) + o(\varepsilon) \quad \text{as } \varepsilon \to 0, \tag{4.53}$$

uniformly for t in compact subsets of $[0,\infty)$, where

$$\begin{cases} \mathbf{y}(t) = 0 & (0 \le t \le s_1) \\ \mathbf{y}(t) = \sum_{k=1}^{m} \lambda_k \mathbf{Y}(t, s_k) y^{s_k} & (s_m \le t \le s_{m+1},\ m = 1, \dots, N-1) \\ \mathbf{y}(t) = \sum_{k=1}^{N} \lambda_k \mathbf{Y}(t, s_k) y^{s_k} & (s_N \le t). \end{cases}$$

DEFINITION. The **cone of variations** at time T is the set

$$K(T) = \Big\{ \sum_{k=1}^{N} \lambda_k \mathbf{Y}(T, s_k) y^{s_k} \mid N = 1, 2, \dots,$$
$$\lambda_k > 0,\ a_k \in A,\ 0 < s_1 \le \cdots \le s_N < T \Big\}.$$

Observe that $K(T)$ is a convex cone in $\mathbb{R}^n$, which according to Lemma 4.5.4 comprises all changes in the state $\mathbf{x}(T)$, up to order ε, that we can effect by multiple variations of the control $\alpha(\cdot)$.

4.5.4. Fixed endpoint problem.

We turn now to the free time, fixed endpoint problem, characterized by the constraint

$$\mathbf{x}(T) = x^1,$$

where $T = T[\alpha(\cdot)]$ is the first time that $\mathbf{x}(\cdot)$ hits the given target point. The payoff functional is

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

Adding a new variable. As before, we introduce the function $x_{n+1} : [0,T] \to \mathbb{R}$ defined by (4.47) and recall the notation (4.48), (4.49), with

$$\bar{g}(\bar{x}) = x_{n+1}.$$

Our problem is therefore to find controlled dynamics satisfying

$$(\overline{\text{ODE}}) \qquad \begin{cases} \dot{\bar{\mathbf{x}}}(t) = \bar{\mathbf{f}}(\bar{\mathbf{x}}(t), \alpha(t)) & (0 \le t \le T) \\ \bar{\mathbf{x}}(0) = \bar{x}^0, \end{cases}$$

and maximizing

$$(\overline{\text{P}}) \qquad \bar{g}(\bar{\mathbf{x}}(T)) = x_{n+1}(T),$$

T being the first time that $\mathbf{x}(T) = x^1$. In other words, the first n components of $\bar{\mathbf{x}}(T)$ are prescribed, and we want to maximize the $(n+1)$-th component.

We assume that $\alpha_0(\cdot)$ is an optimal control for this problem, corresponding to the optimal trajectory $\mathbf{x}_0(\cdot)$ and the time T_0; our task is to construct the corresponding costate $\mathbf{p}_0(\cdot)$, satisfying (ADJ) and the maximization principle (M).

Using the cone of variations. Our program for building the costate depends upon our taking multiple variations, as in the previous section, and understanding the resulting cone of variations $K = K(T_0)$.

Let K^0 denote the (perhaps empty) interior of K. Put

$$e^{n+1} = [0 \cdots 0\, 1]^T \in \mathbb{R}^{n+1}.$$

Here is the key observation:

LEMMA 4.5.5. We have

$$e^{n+1} \notin K^0. \tag{4.54}$$

Proof. 1. If (4.54) were false, there would exist $n+1$ linearly independent vectors $z^1, \dots, z^{n+1} \in K$ such that

$$e^{n+1} = \sum_{k=1}^{n+1} \lambda_k z^k$$

with constants $\lambda_k > 0$ and

$$z^k = \mathbf{Y}(T_0, s_k)\bar{y}^{s_k}$$

for appropriate times $0 < s_1 < s_1 < \cdots < s_{n+1} < T_0$, where

$$\bar{y}^{s_k} = \bar{\mathbf{f}}(\bar{\mathbf{x}}(s_k), a_k)) - \bar{\mathbf{f}}(\bar{\mathbf{x}}(s_k), \alpha(s_k)) \quad (k = 1, \dots, n+1).$$

2. We will next construct a control $\alpha_\varepsilon(\cdot)$, having the multiple variation form (4.50), with corresponding response $\bar{\mathbf{x}}_\varepsilon(\cdot) = [\mathbf{x}_\varepsilon(\cdot)^T\, x^\varepsilon_{n+1}(\cdot)]^T$ satisfying

$$\mathbf{x}_\varepsilon(T_0) = x^1 \tag{4.55}$$

and

$$x^\varepsilon_{n+1}(T_0) > x^0_{n+1}(T_0). \tag{4.56}$$

This will be a contradiction to the optimality of the control $\alpha_0(\cdot)$.

3. Introduce for small $\eta > 0$ the closed and convex set

$$C = \left\{ z = \sum_{k=1}^{n+1} \lambda_k z^k \mid 0 \le \lambda_k \le \eta \right\}.$$

Since the vectors $z^1, \dots, z^{n+1}$ are independent, C has an interior.

Now define for small $\varepsilon > 0$ the mapping

$$\mathbf{\Phi}^{\varepsilon} : C \to \mathbb{R}^{n+1} \tag{4.57}$$

by setting

$$\mathbf{\Phi}^{\varepsilon}(z) = \bar{\mathbf{x}}_{\varepsilon}(T_0) - \bar{\mathbf{x}}_0(T_0)$$

for $z = \sum_{k=1}^{n+1} \lambda_k z^k$, where $\bar{\mathbf{x}}_{\varepsilon}(\cdot)$ solves (4.52) for the control $\alpha_{\varepsilon}(\cdot)$ defined by (4.50).

We assert that if $\mu, \eta, \varepsilon > 0$ are small enough, then we can solve the nonlinear equation

$$\mathbf{\Phi}^{\varepsilon}(z) = \mu e^{n+1} = [0 \cdots 0 \, \mu]^T \tag{4.58}$$

for some $z \in C$. To see this, note that

$$\begin{aligned} |\Phi^{\varepsilon}(z) - z| &= |\bar{\mathbf{x}}_{\varepsilon}(T_0) - \bar{\mathbf{x}}_0(T_0) - z| \\ &= o(|z|) \\ &< |z - \mu e^{n+1}| \qquad \text{for all } z \in \partial C. \end{aligned}$$

We now apply the topological theorem from Appendix F, to find a point $z \in C$ satisfying (4.58). Then

$$\bar{\mathbf{x}}_{\varepsilon}(T_0) = \bar{\mathbf{x}}_0(T_0) + \mu e^{n+1},$$

and hence (4.55), (4.56) hold. This gives the desired contradiction, provided $e^{n+1} \in K^0$. □

Existence of the costate.

THEOREM 4.5.3. (PMP for free time, fixed endpoint problem) There exists a function $\mathbf{p}_0 : [0, T_0] \to \mathbb{R}^n$ and a number

$$q_0 = 0 \text{ or } 1$$

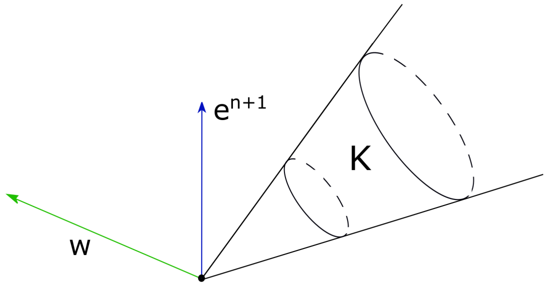
satisfying the statements (ADJ), (M) and (T) of the free time, fixed endpoint PMP.

Proof. 1. Since $e_{n+1} \notin K^0$ according to Lemma 4.5.5, there exists a nonzero vector $w \in \mathbb{R}^{n+1}$ such that

$$w \cdot z \leq 0 \qquad \text{for all } z \in K \tag{4.59}$$

and

$$w \cdot e_{n+1} = w_{n+1} \geq 0. \tag{4.60}$$



Separating e^{n+1} from K

Let $\bar{\mathbf{p}}_0(\cdot)$ solve ($\overline{\text{ADJ}}$), with the terminal condition

$$\bar{\mathbf{p}}_0(T_0) = w.$$

Then

$$p_0^{n+1}(t) = w_{n+1} \geq 0 \quad (0 \leq t \leq T_0). \tag{4.61}$$

Fix any time $0 \leq s < T_0$, any control value $a \in A$, and set

$$\bar{y}^s = \bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), a) - \bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), \alpha_0(s)).$$

Now solve

$$\begin{cases} \dot{\bar{\mathbf{y}}}(t) = \bar{\mathbf{A}}(t)\bar{\mathbf{y}}(t) & (s \leq t \leq T_0) \\ \bar{\mathbf{y}}(s) = \bar{y}^s; \end{cases}$$

so that as before

$$0 \geq w \cdot \bar{\mathbf{y}}(T_0) = \bar{\mathbf{p}}_0(T_0) \cdot \bar{\mathbf{y}}(T_0) = \bar{\mathbf{p}}_0(s) \cdot \bar{\mathbf{y}}(s) = \bar{\mathbf{p}}_0(s) \cdot \mathbf{y}^s.$$

Therefore

$$\bar{\mathbf{p}}_0(s) \cdot [\bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), a) - \bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), \alpha_0(s))] \leq 0;$$

and then

$$\begin{aligned} \bar{H}(\bar{\mathbf{x}}_0(s), \bar{\mathbf{p}}_0(s), a) &= \bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), a) \cdot \bar{\mathbf{p}}_0(s) \\ &\leq \bar{\mathbf{f}}(\bar{\mathbf{x}}_0(s), \alpha_0(s)) \cdot \bar{\mathbf{p}}_0(s) \\ &= \bar{H}(\bar{\mathbf{x}}_0(s), \bar{\mathbf{p}}_0(s), \alpha_0(s)), \end{aligned} \tag{4.62}$$

for the Hamiltonian

$$\bar{H}(\bar{x}, \bar{p}, a) = \bar{\mathbf{f}}(\bar{x}, a) \cdot \bar{p}.$$

2. We now must address two situations, according to whether

$$w_{n+1} > 0 \tag{4.63}$$

or

$$w_{n+1} = 0. \tag{4.64}$$

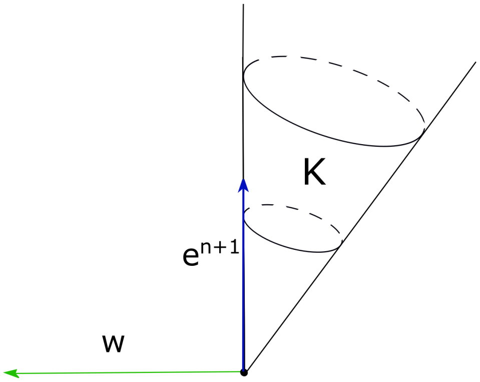
When (4.63) holds, we can divide $\mathbf{p}_0(\cdot)$ by w_{n+1}, to reduce to the case that

$$q_0 = p^0_{n+1} \equiv 1.$$

Then (4.62) provides the maximization principle (M). If instead (4.64) holds, we have an abnormal problem, for which

$$q_0 = p^0_{n+1} \equiv 0.$$

□



An abnormal problem

REMARK. We have not proved the condition (T) that

$$H(\mathbf{x}_0(t), \mathbf{p}_0(t), \alpha_0(t)) = 0 \quad (0 \leq t \leq T_0)$$

for the free time, fixed endpoint problem. This is in fact rather subtle: see the book [**L-M**] of Lee and Markus for details. For more advanced students, I recommend the book of Bressan and Piccoli [**B-P**], which provides a fully rigorous and detailed proof of the PMP. □

Chapter 5

DYNAMIC PROGRAMMING

5.1. Hamilton-Jacobi-Bellman equation

We next show how to use value functions in optimal control theory, within the context of **dynamic programming**. This approach depends upon the mathematical idea that it is sometimes easier to solve a given problem by incorporating it within a larger class of problems.

We want to adapt some version of this insight to the vastly complicated setting of control theory. For this, fix a terminal time $T > 0$ and then look at the controlled dynamics

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

with the associated payoff functional

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

for the free endpoint problem.

The new idea is to embed this into a larger family of similar problems, by varying the starting times and starting points:

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

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

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

5.1.1. Derivation.

DEFINITION. For $x \in \mathbb{R}^n$, $0 \le t \le T$, we define the **value function** $v : \mathbb{R}^n \times [0, T] \to \mathbb{R}$ to be the greatest payoff possible if we start at $x \in \mathbb{R}^n$ at time t. In other words,

$$\boxed{v(x,t) = \sup_{\alpha(\cdot) \in \mathcal{A}} P_{x,t}[\alpha(\cdot)]}$$

for $x \in \mathbb{R}^n, 0 \le t \le T$.

REMARK. Then

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

since if we start at time T at the point x, we must immediately stop and so collect the payoff $g(x)$. □

Our task in this section is to show that the value function v so defined satisfies a certain nonlinear partial differential equation. Our derivation will be based upon the reasonable principle that *it is better to act optimally from the beginning, rather than to act arbitrarily for a while and then later act optimally.* We will convert this philosophy of life into mathematics.

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

THEOREM 5.1.1. Assume that the value function v is a continuously differentiable function of the variables (x, t). Then v solves the **Hamilton–Jacobi–Bellman** partial differential equation

$$\text{(HJB)} \qquad \boxed{\frac{\partial v}{\partial t}(x,t) + \max_{a \in A} \{\mathbf{f}(x,a) \cdot \nabla_x v(x,t) + r(x,a)\} = 0}$$

for $x \in \mathbb{R}^n, 0 \le t < T$, with the terminal condition

$$\text{(5.1)} \qquad v(x,T) = g(x) \qquad (x \in \mathbb{R}^n).$$

DEFINITION. The **PDE Hamiltonian** is

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

where $x, p \in \mathbb{R}^n$. Hence we can write (HJB) as

$$\frac{\partial v}{\partial t} + H(x, \nabla_x v) = 0 \qquad \text{in } \mathbb{R}^n \times [0, T].$$

□

Proof. 1. Let $x \in \mathbb{R}^n$, $0 \leq t < T$, and note that, as always in this course,

$$\mathcal{A} = \{\alpha(\cdot) : [0, T] \to A \mid \alpha(\cdot) \text{ is piecewise continuous}\}.$$

Pick any parameter $a \in A$ and let $\varepsilon > 0$ be so small that $t + \varepsilon \leq T$. Suppose we start at x at time t, and use the constant control

$$\alpha(s) = a$$

for times $t \leq s \leq t + \varepsilon$. The dynamics then arrive at the point $\mathbf{x}(t + \varepsilon)$. Suppose now that we switch to an optimal control (assuming it exists) and employ it for the remaining times $t + \varepsilon \leq s \leq T$.

What is the payoff of this procedure? Now for $t \leq s \leq t + \varepsilon$, we have

$$\text{(5.2)} \qquad \begin{cases} \dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), a) \\ \mathbf{x}(t) = x. \end{cases}$$

The payoff for this time period is $\int_t^{t+\varepsilon} r(\mathbf{x}(s), a)\, ds$. Furthermore, the payoff incurred from time $t + \varepsilon$ to T is $v(\mathbf{x}(t + \varepsilon), t + \varepsilon)$, according to the definition of the payoff function $v(\cdot)$. Hence the total payoff is

$$\int_t^{t+\varepsilon} r(\mathbf{x}(s), a)\, ds + v(\mathbf{x}(t + \varepsilon), t + \varepsilon).$$

But the greatest possible payoff if we start from (x, t) is $v(x, t)$. Therefore

$$\int_t^{t+\varepsilon} r(\mathbf{x}(s), a)\, ds + v(\mathbf{x}(t + \varepsilon), t + \varepsilon) \leq v(x, t).$$

2. Next rearrange (5.5) and divide by $\varepsilon > 0$:

$$\frac{v(\mathbf{x}(t + \varepsilon), t + \varepsilon) - v(x, t)}{\varepsilon} + \frac{1}{\varepsilon} \int_t^{t+\varepsilon} r(\mathbf{x}(s), a)\, ds \leq 0.$$

Hence

$$\frac{\partial v}{\partial t}(x, t) + \nabla_x v(\mathbf{x}(t), t) \cdot \dot{\mathbf{x}}(t) + r(\mathbf{x}(t), a) \leq o(1),$$

as $\varepsilon \to 0$. But recall that $\mathbf{x}(\cdot)$ solves (5.2). We employ this above and send $\varepsilon \to 0$, to discover

$$\frac{\partial v}{\partial t}(x, t) + \mathbf{f}(x, a) \cdot \nabla_x v(x, t) + r(x, a) \leq 0.$$

This inequality holds for all control parameters $a \in A$, and consequently

$$\max_{a\in A}\left\{\frac{\partial v}{\partial t}(x,t)+\mathbf{f}(x,a)\cdot\nabla_x v(x,t)+r(x,a)\right\}\le 0. \tag{5.3}$$

3. We next demonstrate that in fact the maximum above equals zero. To see this, suppose $\alpha_0(\cdot)$, $\mathbf{x}_0(\cdot)$ are optimal for the problem above. Let us utilize the optimal control $\alpha_0(\cdot)$ for $t \le s \le t+\varepsilon$. The payoff is

$$\int_t^{t+\varepsilon} r(\mathbf{x}_0(s),\alpha_0(s))\,ds$$

and the remaining payoff is $v(\mathbf{x}_0(t+\varepsilon),t+\varepsilon)$. Consequently, the total payoff is

$$\int_t^{t+\varepsilon} r(\mathbf{x}_0(s),\alpha_0(s))\,ds+v(\mathbf{x}_0(t+\varepsilon),t+\varepsilon)=v(x,t).$$

Rearrange and divide by ε:

$$\frac{v(\mathbf{x}_0(t+\varepsilon),t+\varepsilon)-v(x,t)}{\varepsilon}+\frac{1}{\varepsilon}\int_t^{t+\varepsilon} r(\mathbf{x}_0(s),\alpha_0(s))\,ds=0.$$

Let $\varepsilon \to 0$ and suppose $\alpha_0(t)=a_0\in A$. Then

$$\frac{\partial v}{\partial t}(x,t)+\nabla_x v(x,t)\cdot\mathbf{f}(x,a_0)+r(x,a_0)=0.$$

This and (5.3) confirm that v solves the Hamilton-Jacobi-Bellman PDE. □

REMARK. Dynamic programming applies as well to free time, fixed endpoint optimal control problems. To be specific, let us suppose that we are required to steer from a point $x \in \mathbb{R}^n$ to the origin under the dynamics

$$\text{(ODE)}\qquad \begin{cases}\dot{\mathbf{x}}(s)=\mathbf{f}(\mathbf{x}(s),\alpha(s)) & (0\le s\le T)\\ \mathbf{x}(0)=x,\ x(T)=0,\end{cases}$$

so as to maximize the payoff

$$\text{(P)}\qquad P_x[\alpha(\cdot)]=\int_0^T r(\mathbf{x}(s),\alpha(s))\,ds.$$

Here the terminal time T is free.

The corresponding value function is

$$\boxed{v(x)=\sup_{\alpha(\cdot)\in\mathcal{A}} P_x[\alpha(\cdot)]}\qquad (x\in\mathbb{R}^n).$$

Arguing as above, we discover that if the value function v is continuously differentiable, it solves the **(stationary) Hamilton-Jacobi-Bellman equation**

$$\text{(HJB)} \qquad \boxed{\max_{a\in A}\{\mathbf{f}(x,a)\cdot\nabla v(x)+r(x,a)\}=0}$$

for $x\in\mathbb{R}^n\setminus\{0\}$ and

$$v(0)=0.$$

□

5.1.2. Optimality.

HOW TO USE DYNAMIC PROGRAMMING

For fixed time optimal control problems as in Section 5.1.1, we carry out these steps to synthesize an optimal control:

Step 1: Try to solve the HJB equation, with the terminal condition (5.1), and thereby find the value function v.

Step 2: Use the value function v and the Hamilton–Jacobi–Bellman PDE to design an optimal control $\alpha_0(\cdot)$, as follows. Define for each point $y\in\mathbb{R}^n$ and each time $0\le s\le T$,

$$a(y,s)\in A$$

to be a parameter value where the maximum in HJB is attained at the point (y,s). In other words, select $a(y,s)\in A$ so that

$$\text{(5.4)} \qquad \frac{\partial v}{\partial t}(y,s)+\mathbf{f}(y,a(y,s))\cdot\nabla_x v(y,s)+r(y,a(y,s))=0.$$

Step 3: Next, solve the following ODE, assuming $a(\cdot)$ is sufficiently regular to do so:

$$\text{(5.5)} \qquad \begin{cases}\dot{\mathbf{x}}_0(s)=\mathbf{f}(\mathbf{x}_0(s),a(\mathbf{x}_0(s),s)) & (t\le s\le T)\\ \mathbf{x}_0(t)=x.\end{cases}$$

Step 4: Finally, define the **optimal feedback control**

$$\text{(5.6)} \qquad \alpha_0(s)=a(\mathbf{x}_0(s),s) \quad (t\le s\le T),$$

so that we can rewrite (5.5) as

$$\begin{cases}\dot{\mathbf{x}}_0(s)=\mathbf{f}(\mathbf{x}_0(s),\alpha_0(s)) & (t\le s\le T)\\ \mathbf{x}_0(t)=x.\end{cases}$$

In particular, if the state of system is y at time s, we use the control which at time s takes on a parameter value $a=a(y,s)\in A$ for which the maximum in HJB is obtained.

THEOREM 5.1.2. For each starting time $0 \leq t < T$ and initial point $x \in \mathbb{R}^n$, the control $\alpha_0(\cdot)$ defined by (5.5) and (5.6) is optimal.

Proof. We have

$$P_{x,t}[\alpha_0(\cdot)] = \int_t^T r(\mathbf{x}_0(s), \alpha_0(s))\, ds + g(\mathbf{x}_0(T)).$$

Then (5.4) and (5.6) imply

$$\begin{aligned} P_{x,t}[\alpha_0(\cdot)] &= \int_t^T \left(-\frac{\partial v}{\partial t}(\mathbf{x}_0(s), s) - \mathbf{f}(\mathbf{x}_0(s), \alpha_0(s)) \cdot \nabla_x v(\mathbf{x}_0(s), s) \right) ds + g(\mathbf{x}_0(T)) \\ &= -\int_t^T \frac{\partial v}{\partial t}(\mathbf{x}_0(s), s) + \nabla_x v(\mathbf{x}_0(s), s) \cdot \dot{\mathbf{x}}_0(s)\, ds + g(\mathbf{x}_0(T)) \\ &= -\int_t^T \frac{d}{ds} v(\mathbf{x}_0(s), s)\, ds + g(\mathbf{x}_0(T)) \\ &= -v(\mathbf{x}_0(T), T) + v(\mathbf{x}_0(t), t) + g(\mathbf{x}_0(T)) \\ &= v(x, t) \\ &= \sup_{\alpha(\cdot) \in \mathcal{A}} P_{x,t}[\alpha(\cdot)]. \end{aligned}$$

Hence $\alpha_0(\cdot)$ is optimal, as asserted. □

REMARKS.

(i) Notice that v acts here as a calibration function that we use to establish optimality.

(ii) We can similarly design optimal controls for free time problems by solving the stationary HJB equation. □

5.2. Applications

Applying dynamic programming is usually quite tricky, as it requires us to solve a nonlinear PDE and this is often very difficult. The main hope, as we will see in the following examples, is *to try to guess the form of v, to plug this guess into the HJB equation and then to adjust various constants and auxiliary functions, to ensure that we have an actual solution.* (Alternatively, we could compute the solution of the terminal-value problem for HJB numerically.)

To simplify notation will not write the subscripts "0" in the subsequent examples.

5.2.1. General linear-quadratic regulator. For this important problem, we are given matrices $M, B, D \in \mathbb{M}^{n\times n}$, $N \in \mathbb{M}^{n\times m}$, $C \in \mathbb{M}^{m\times m}$; and assume

$$B, C, D \text{ are symmetric,}$$

with

$$B, D \succeq 0,\ C \succ 0.$$

In particular, C is invertible.

We take the linear dynamics

$$\text{(ODE)} \qquad \begin{cases} \dot{\mathbf{x}}(s) = M\mathbf{x}(s) + N\alpha(s) & (t \le s \le T) \\ \mathbf{x}(t) = x, \end{cases}$$

for which we want to minimize the quadratic cost functional

$$\int_t^T \mathbf{x}(s)^T B\mathbf{x}(s) + \alpha(s)^T C\alpha(s)\, ds + \mathbf{x}(T)^T D\mathbf{x}(T).$$

So we must maximize the payoff

$$\text{(P)} \qquad P_{x,t}[\alpha(\cdot)] = -\int_t^T \mathbf{x}(s)^T B\mathbf{x}(s) + \alpha(s)^T C\alpha(s)\, ds - \mathbf{x}(T)^T D\mathbf{x}(T).$$

The control values are unconstrained, meaning that the control parameter values can range over all of $A = \mathbb{R}^m$.

We employ dynamic programming to design an optimal control. To carry out this plan, we first figure out the structure of the HJB equation

$$\begin{cases} \frac{\partial v}{\partial t} + \max_{a\in\mathbb{R}^m} \{\mathbf{f}\cdot\nabla_x v + r\} = 0 & \text{in } \mathbb{R}^n \times [0,T] \\ \qquad\qquad v = g & \text{on } \mathbb{R}^n \times \{t = T\}, \end{cases}$$

for

$$\mathbf{f} = Mx + Na,\ r = -x^T Bx - a^T Ca,\ g = -x^T Dx.$$

We rewrite the PDE as

$$\text{(5.7)} \qquad \frac{\partial v}{\partial t} + \max_{a\in\mathbb{R}^m}\{(\nabla v)^T Na - a^T Ca\} + (\nabla v)^T Mx - x^T Bx = 0,$$

and note that we have the terminal condition

$$\text{(5.8)} \qquad v(x,T) = -x^T Dx.$$

To compute the maximum above, define

$$Q(a) = (\nabla v)^T Na - a^T Ca,$$

and solve

$$\frac{\partial Q}{\partial a_j} = \sum_{i=1}^n v_{x_i} n_{ij} - 2a_i c_{ij} = 0 \quad (j = 1, \dots, n).$$

Then $(\nabla_x v)^T N = 2a^T C$, and thus $2Ca = N^T \nabla_x v$, since C is symmetric. Therefore

$$a = \frac{1}{2} C^{-1} N^T \nabla_x v. \tag{5.9}$$

This is the point at which the maximum in HJB occurs, which we now insert into (5.7):

$$\frac{\partial v}{\partial t} + \frac{1}{4} (\nabla v)^T N C^{-1} N^T \nabla v + (\nabla v)^T M x - x^T B x = 0. \tag{5.10}$$

Solving the HJB equation. Our task now is to solve this nonlinear PDE, with the terminal condition (5.8). We *guess* that our solution has the form

$$v(x,t) = x^T K(t) x \tag{5.11}$$

for some appropriate symmetric $n \times n$-matrix valued function $K(\cdot)$ for which

$$K(T) = -D. \tag{5.12}$$

Let us compute

$$\frac{\partial v}{\partial t} = x^T \dot{K}(t) x, \quad \nabla_x v = 2K(t)x. \tag{5.13}$$

We now insert our guess $v = x^T K(t) x$ into (5.10), to discover that

$$x^T \{\dot{K}(t) + K(t) N C^{-1} N^T K(t) + 2K(t) M - B\} x = 0.$$

Since

$$\begin{aligned} 2x^T K M x &= x^T K M x + [x^T K M x]^T \\ &= x^T K M x + x^T M^T K x, \end{aligned}$$

the foregoing becomes

$$x^T \{\dot{K} + K N C^{-1} N^T K + K M + M^T K - B\} x = 0.$$

This identity will hold provided $K(\cdot)$ satisfies the **matrix Riccati equation**

$$\boxed{\dot{K}(t) + K(t) N C^{-1} N^T K(t) + K(t) M + M^T K(t) = B} \tag{R}$$

on the interval $[0, T]$.

In summary, once we solve the Riccati equation (R) with the terminal condition (5.12), we can then use (5.9) and (5.13) to construct the optimal feedback control

$$\alpha_0(t) = C^{-1} N^T K(t) \mathbf{x}_0(t).$$

□

5.2.2. Rocket railway car. In view of our discussion on page 98, the rocket railway problem is actually quite easy to solve. However it is also instructive to see how dynamic programming applies.

The equations of motion are

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

for $n = 2$, and

$$P_x[\alpha(\cdot)] = -\text{ time to reach the origin } = -\int_0^T 1\,dt = -T.$$

Then the value function $v(x)$ is *minus* the least time it takes to get to the origin from the point $x = [x_1\, x_2]^T$; and the corresponding stationary HJB equation is

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

for

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

Therefore

$$\max_{|a| \le 1} \left\{ x_2 \frac{\partial v}{\partial x_1} + a \frac{\partial v}{\partial x_2} - 1 \right\} = 0;$$

and consequently the Hamilton-Jacobi-Bellman equation is

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

Solution of HJB equation. We introduce the regions

$$I := \{(x_1, x_2) \mid x_1 > -\tfrac{1}{2} x_2 |x_2|\},$$
$$II := \{(x_1, x_2) \mid x_1 < -\tfrac{1}{2} x_2 |x_2|\},$$

and define

$$(5.14) \qquad v(x) = \begin{cases} -x_2 - 2\left(x_1 + \frac{1}{2} x_2^2\right)^{\frac{1}{2}} & \text{in Region I} \\ x_2 - 2\left(-x_1 + \frac{1}{2} x_2^2\right)^{\frac{1}{2}} & \text{in Region II.} \end{cases}$$

We could have derived this formula for v using the ideas in the next example, but for now let us just show that v really solves HJB.

In Region I we compute

$$\frac{\partial v}{\partial x_1} = -\left(x_1 + \frac{x_2^2}{2}\right)^{-\frac{1}{2}}, \ \frac{\partial v}{\partial x_2} = -1 - \left(x_1 + \frac{x_2^2}{2}\right)^{-\frac{1}{2}} x_2;$$

and check that there

$$\frac{\partial v}{\partial x_2} < 0.$$

Hence in Region I we have

$$x_2 \frac{\partial v}{\partial x_1} + \left| \frac{\partial v}{\partial x_2} \right| = -x_2 \left(x_1 + \frac{x_2^2}{2} \right)^{-\frac{1}{2}} + \left[1 + x_2 \left(x_1 + \frac{x_2^2}{2} \right)^{-\frac{1}{2}} \right] = 1.$$

This confirms that our HJB equation holds in Region I, and a similar calculation holds in Region II, owing to the symmetry condition

$$v(-x) = v(x) \quad (x \in \mathbb{R}^2).$$

Now let Γ denote the boundary between Regions I and II. Since

$$\frac{\partial v}{\partial x_2} \begin{cases} < 0 & \text{in Region I} \\ > 0 & \text{in Region II} \end{cases}$$

and

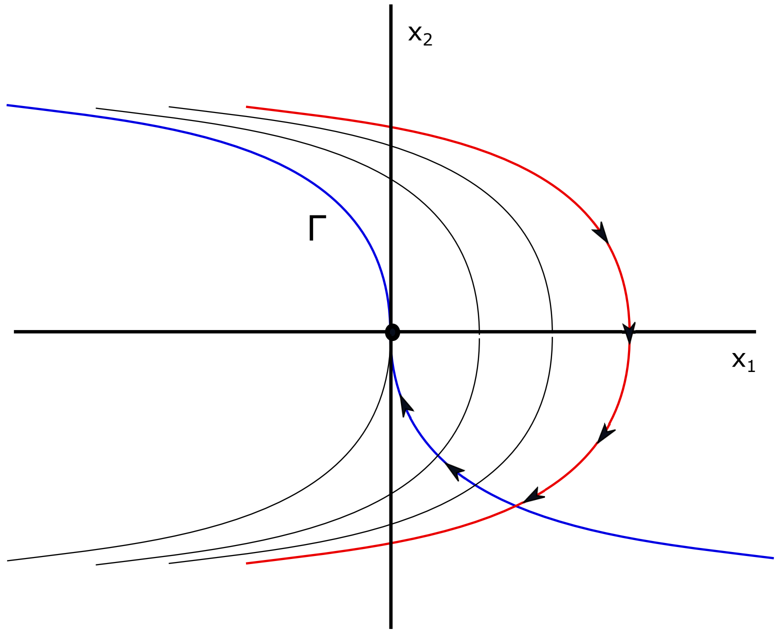
$$\frac{\partial v}{\partial x_2} = 0 \quad \text{on } \Gamma,$$

our function v defined by (5.14) does indeed solve the nonlinear HJB partial differential equation.

GEOMETRIC INTERPRETATION. It is in fact easy to find optimal trajectories for this problem. Indeed, when $\alpha = \pm 1$, then

$$\frac{d}{dt} \left(x_1 \mp \frac{1}{2}(x_2)^2 \right) = x_2 \mp x_2(\pm 1) = 0.$$

Thus any solution of (ODE) moves along a parabola of the form $x_1 = \frac{x_2^2}{2} + C$ when $\alpha = 1$, and along a parabola of the form $x_1 = -\frac{x_2^2}{2} + C$ when $\alpha = -1$. Since we know from the PMP that an optimal control changes sign at most once (see page 98), an optimal trajectory must move along one such family of parabolas, and change to the other family of parabolas at most once, at the switching curve Γ given by the formula $x_1 = -\frac{1}{2}|x_2|x_2$. The picture illustrates a typical optimal trajectory.



Optimal path for rocket railway car

So we did not need to invoke the full majesty of dynamic programming to solve this simple problem, but the next example will build upon these ideas. □

5.2.3. Fuller's problem, chattering controls. We take the same equations of motion as in the previous example

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

but make a simple change in the payoff functional (which will have a profound effect on optimal controls and trajectories). So let us now take

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

Thus

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

and the stationary HJB equation for the value function

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

is now

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

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

$$\text{(5.15)} \qquad v(-x) = v(x) \qquad (x \in \mathbb{R}^2)$$

and

$$\text{(5.16)} \qquad v(\lambda^2 x_1, \lambda x_2) = \lambda^5 v(x_1, x_2) \qquad (x \in \mathbb{R}^2, \lambda > 0).$$

To verify (5.16), suppose that $\mathbf{x} = [x_1\, x_2]^T$ is an optimal trajectory starting at the point $x^0 = [x_1^0\, x_2^0]^T$, corresponding to an optimal control α. Then if $\lambda > 0$ and we start instead at the point $x_\lambda^0 = [\lambda^2 x_1^0 \;\lambda x_2^0]^T$, optimal trajectories and control are

$$\mathbf{x}_\lambda(t) = [\lambda^2 x_1(\tfrac{t}{\lambda})\; \lambda x_2(\tfrac{t}{\lambda})]^T, \;\; \alpha_\lambda(t) = \alpha(\tfrac{t}{\lambda}).$$

Since

$$P[\alpha_\lambda(\cdot)] = -\frac{1}{2}\int_0^{T_\lambda} (x_\lambda^1)^2 dt = -\frac{\lambda^4}{2}\int_0^{T_\lambda} (x_1)^2(\tfrac{t}{\lambda}) dt = \lambda^5 P[\alpha(\cdot)],$$

the scaling identity (5.16) holds.

Now (5.16) and the previous example suggest that the optimal switching should occur on the boundary Γ between two regions of the form

$$I := \{(x_1, x_2) \mid x_1 > -\beta x_2 |x_2|\},$$
$$II := \{(x_1, x_2) \mid x_1 < -\beta x_2 |x_2|\},$$

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

Still motivated by the previous example, we look a function v for which the scaling symmetry (5.16) holds,

$$\text{(5.17)} \qquad \frac{\partial v}{\partial x_2} < 0 \text{ in Region I}, \qquad \frac{\partial v}{\partial x_2} = 0 \text{ on } \Gamma,$$

and v solves the *linear* PDE

$$\text{(5.18)} \qquad x_2 \frac{\partial v}{\partial x_1} - \frac{\partial v}{\partial x_2} = \frac{1}{2}(x_1)^2 \quad \text{in Region I.}$$

In view (5.16), let us start by looking for a particular solution of (5.18) having the polynomial form

$$v = A x_2^5 + B x_1 x_2^3 + C x_1^2 x_2.$$

If we plug this guess into (5.18) 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 (5.18) is

$$v = -\frac{1}{15}x_2^5 - \frac{1}{3}x_1x_2^3 - \frac{1}{2}x_1^2x_2 + f\left(x_1 + \frac{1}{2}x_2^2\right).$$

In order to satisfy the scaling condition (5.16), we take f to be homogeneous and so have

$$v = -\frac{1}{15}x_2^5 - \frac{1}{3}x_1x_2^3 - \frac{1}{2}x_1^2x_2 - \gamma\left(x_1 + \frac{1}{2}x_2^2\right)^{\frac{5}{2}} \tag{5.19}$$

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

$$\frac{\partial v}{\partial x_2} = -\frac{1}{3}x_2^4 - x_1x_2^2 - \frac{1}{2}x_1^2 - \frac{5\gamma}{2}\left(x_1 + \frac{1}{2}x_2^2\right)^{\frac{3}{2}} x_2. \tag{5.20}$$

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

$$\frac{\partial v}{\partial x_2} = x_2^4\left[-\frac{1}{3} + \beta - \frac{1}{2}\beta^2 - \frac{5\gamma}{2}\left(-\beta + \frac{1}{2}\right)^{\frac{3}{2}}\right]; \tag{5.21}$$

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

$$\frac{\partial v}{\partial x_2} = x_2^4\left[-\frac{1}{3} - \beta - \frac{1}{2}\beta^2 + \frac{5\gamma}{2}\left(\beta + \frac{1}{2}\right)^{\frac{3}{2}}\right]. \tag{5.22}$$

Consequently, we need to select β, γ so that

$$\begin{cases} -\frac{1}{3} + \beta - \frac{1}{2}\beta^2 - \frac{5\gamma}{2}\left(-\beta + \frac{1}{2}\right)^{\frac{3}{2}} = 0 \\ -\frac{1}{3} - \beta - \frac{1}{2}\beta^2 + \frac{5\gamma}{2}\left(\beta + \frac{1}{2}\right)^{\frac{3}{2}} = 0. \end{cases} \tag{5.23}$$

Solving each equation for γ, we see (5.23) implies

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

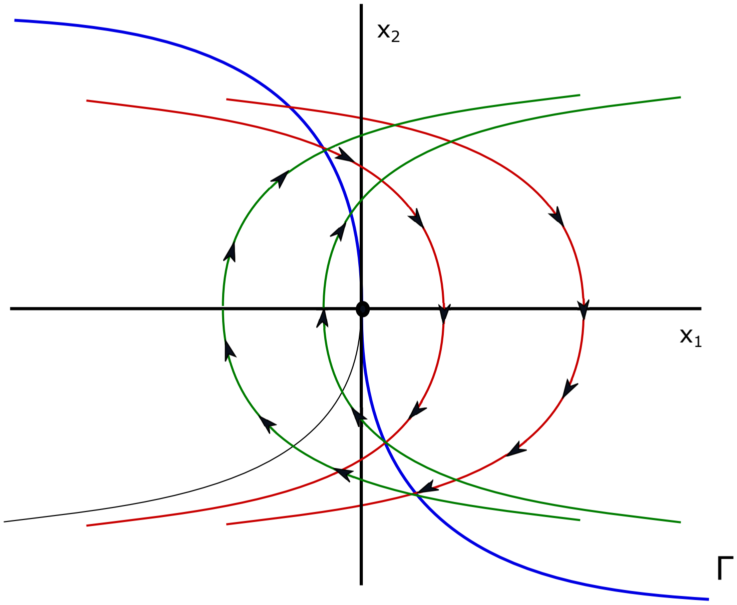
Since

$$\phi(0) < 0,\ \phi(\tfrac{1}{2}) > 0,$$

there exist $0 < \beta < \frac{1}{2}$ such that $\phi(\beta) = 0$. We can then find $\gamma > 0$ so that β, γ solve (5.23). A further calculation confirms that for these choices, $\frac{\partial v}{\partial x_2} < 0$ within Region I.

Using (5.15) to extend our definition of v to all of $\mathbb{R}^2$, we have (at last) found our solution of the stationary HJB equation.

GEOMETRIC INTERPRETATION. Optimal trajectories for Fuller's problem are more interesting than for the rocket-railway car.



Part of an optimal path for Fuller's problem

As before, a solution of (ODE) 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 $x_1 = -\beta|x_2|x_2$.

But since $0 < \beta < \frac{1}{2}$, 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**. □

CLOSING REMARKS. Our detailed discussion of Fuller's problem illustrates how difficult it can be to find an explicit solution of an HJB equation, if this is even possible.

In fact, there are not so many optimal control problems for which exact formulas can be had, using either the Pontryagin maximum principle or dynamic programming. (See the old book of Athans and Falb [**A-F**] for an extensive discussion of various solvable engineering control problems.) Designing optimal controls is an important, but highly nonlinear and infinite dimensional undertaking, and it is not surprising that exactly solvable problems have been named after the researchers who found them.

It is therefore essential to turn to computational methods for most optimal control problems, and indeed for optimization problems in general. I therefore strongly recommend that students take subsequent classes (mostly offered in engineering) on computational techniques for optimization, with the hope that their understanding the optimization theory from Math 170 and 195 will make the algorithms and software packages from these courses more understandable. □

APPENDIX

A. Notation

$\mathbb{R}^n$ denotes n-dimensional Euclidean space, a typical point of which is the column vector

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

To save space we will often write the corresponding row vector

$$x = [x_1 \cdots x_n]^T.$$

If $x, y \in \mathbb{R}^n$, we define

$$x \cdot y = \sum_{i=1}^{n} x_i y_i = x^T y, \qquad |x| = (x \cdot x)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

B. Linear algebra

Throughout these notes A denotes a real $m \times n$ matrix and A^T denotes its transpose:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad A^T = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{mn} \end{bmatrix}.$$

Recall the rule

$$(AB)^T = B^T A^T.$$

We write

$$\mathbb{M}^{n\times m}$$

for the space of all real $m \times n$ matrices.

An $n \times n$ matrix A is **symmetric** if $A = A^T$, and a symmetric matrix A is **nonnegative definite** if

$$y^T Ay = \sum_{i,j=1}^{n} a_{ij} y_i y_j \geq 0 \quad \text{for all } y \in \mathbb{R}^n.$$

We then write $A \succeq 0$. We say A is **positive definite** if

$$y^T Ay = \sum_{i,j=1}^{n} a_{ij} y_i y_j > 0 \quad \text{for all } y \in \mathbb{R}^n;$$

and write $A \succ 0$.

C. Multivariable chain rule

Let $f : \mathbb{R}^n \to \mathbb{R}$, $f = f(x) = f(x_1, \dots, x_n)$. Then we write

$$\frac{\partial f}{\partial x_k}$$

for the k-th partial derivative, $k = 1, \dots, n$. We likewise write

$$\frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{\partial}{\partial x_l}\left(\frac{\partial f}{\partial x_k}\right) \qquad (k, l = 1, \dots, n),$$

and recall that $\frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{\partial^2 f}{\partial x_l \partial x_k}$ if f is twice continuously differentiable.

The **gradient** ∇f is the vector

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

and the **Hessian matrix** of second partial derivatives $\nabla^2 f$ is the symmetric $n \times n$ matrix

$$\nabla^2 f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f}{\partial x_1 \partial x_n} & \dots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The chain rule tells us how to compute the partial derivatives of composite functions, made from simpler functions. For this, assume that we are given a function

$$f : \mathbb{R}^n \to \mathbb{R},$$

which we write as $f(x) = f(x_1, \dots, x_n)$. Suppose also we have functions

$$g_1, \dots, g_n : \mathbb{R}^m \to \mathbb{R}$$

so that $g_i(y) = g_i(y_1, \dots, y_m)$ for $i = 1, \dots, n$.

NOTATION. We define $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^n$ by

$$\mathbf{g} = \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix}.$$

The gradient matrix of $\mathbf{g}$ is

$$\nabla \mathbf{g} = \begin{bmatrix} (\nabla g_1)^T \\ \vdots \\ (\nabla g_n)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \dots & \frac{\partial g_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial y_1} & \dots & \frac{\partial g_n}{\partial y_m} \end{bmatrix}.$$

This is an $n \times m$ matrix-valued function. □

We now build the composite function $h : \mathbb{R}^m \to \mathbb{R}$ by setting $x_i = g_i(y)$ in the definition of f; that is, we define

$$h(y) = f(\mathbf{g}(y)) = f(g_1(y), g_2(y), \dots, g_n(y)).$$

THEOREM. (Multivariable chain rule) We have

$$\boxed{\frac{\partial h}{\partial y_k}(y) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{g}(y)) \frac{\partial g_i}{\partial y_k}(y)} \qquad (k = 1, \dots, m).$$

In matrix notation, this says

$$\nabla h = (\nabla \mathbf{g})^T \nabla f.$$

EXAMPLE. We use the chain rule to prove the useful formula

$$\nabla \left(|\mathbf{g}|^2 \right) = 2 (\nabla \mathbf{g})^T \mathbf{g}$$

where $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$.

To prove this, we compute that

$$\frac{1}{2} \frac{\partial}{\partial x_k} |\mathbf{g}|^2 = \sum_{i=1}^{n} g_i \frac{\partial g_i}{\partial x_k} = k\text{-th entry of } (\nabla \mathbf{g})^T \mathbf{g}.$$

□

D. Divergence Theorem

If $U \subset \mathbb{R}^n$ is an open set, with smooth boundary ∂U, we let

$$\boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}.$$

denote the *outward pointing* unit normal vector field along ∂U. Then $|\boldsymbol{\nu}| = 1$ along ∂U.

Assume also that $\mathbf{h} : U \to \mathbb{R}^n$, written

$$\mathbf{h} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix},$$

is a vector field. Its divergence is

$$\operatorname{div} \mathbf{h} = \nabla \cdot \mathbf{h} = \sum_{i=1}^{n} \frac{\partial h_i}{\partial x_i}.$$

THEOREM. (Divergence Theorem) We have

$$\boxed{\int_U \operatorname{div} \mathbf{h} \, dx = \int_{\partial U} \mathbf{h} \cdot \boldsymbol{\nu} \, dS.}$$

The expression on the right is an integral with respect to ($n-1$ dimensional) surface area over the boundary of U.

E. Implicit Function Theorem

Assume for this section that

$$f : \mathbb{R}^2 \to \mathbb{R}$$

is continuously differentiable. We will write $f = f(x, y)$.

THEOREM. (Implicit Function Theorem) Assume that $f(x_0, y_0) = 0$ and

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0.$$

(i) Then there exist $\varepsilon > 0$ and a continuously differentiable function

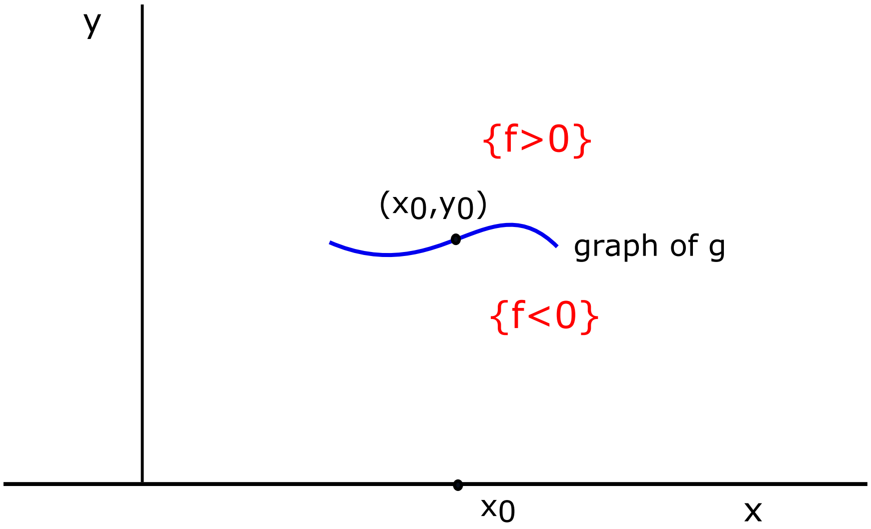
$$g : (x_0 - \varepsilon, x_0 + \varepsilon) \to \mathbb{R}$$

such that

$$f(x, g(x)) = 0 \qquad (x_0 - \varepsilon < x < x_0 + \varepsilon).$$

(ii) Furthermore, for every solution of $f(x,y)=0$ with (x,y) sufficiently close to (x_0,y_0), we have

$$y=g(x).$$



Implicit Function Theorem (if $\frac{\partial f}{\partial y}(x_0,y_0)>0$)

F. Solving a nonlinear equation

THEOREM. Let C denote a closed, bounded, convex subset of $\mathbb{R}^n$ and assume p lies in the interior of C. Suppose $\mathbf{\Phi}:C\to\mathbb{R}^n$ is a continuous vector field that satisfies the strict inequalities

$$|\mathbf{\Phi}(x)-x|<|x-p| \qquad \text{for all } x\in\partial C$$

Then there exists a point $x\in C$ such that

$$\mathbf{\Phi}(x)=p.$$

Proof. 1. Suppose first that C is the unit ball $B(0,1)$ and $p=0$. Squaring the inequality $|\mathbf{\Phi}(x)-x|<|x|$, we deduce that

$$\mathbf{\Phi}(x)\cdot x>0 \qquad \text{for all } x\in\partial B(0,1).$$

Then for small $t>0$, the continuous mapping

$$\mathbf{\Psi}(x):=x-t\mathbf{\Phi}(x)$$

maps $B(0,1)$ into itself, and hence has a fixed point x according to Brouwer's Fixed Point Theorem. Then $\mathbf{\Phi}(x)=0=p$.

2. For the general case, we can always assume after a translation that $p = 0$, so that 0 belongs to the interior of C. We introduce a nonnegative gauge function $\rho : \mathbb{R}^n \to [0, \infty)$ such that $\rho(\lambda x) = \lambda \rho(x)$ for all $\lambda \geq 0$ and

$$C = \{x \in \mathbb{R}^n \mid \rho(x) \leq 1\}.$$

We next map C onto $B(0,1)$ by the continuous function

$$\mathbf{a}(x) = \frac{\rho(x)}{|x|} x = y.$$

Define

$$\mathbf{\Psi}(y) = \frac{\rho(y)}{|y|} \mathbf{\Phi} \left(\frac{|y|}{\rho(y)} y \right).$$

Then the inequality $|\mathbf{\Phi}(x) - x| < |x|$ implies

$$|\mathbf{\Psi}(y) - y| < 1 \qquad \text{for all } y \in \partial B(0,1).$$

Consequently the first part of the proof shows that there exits $y \in B(0,1)$ such that $\mathbf{\Psi}(y) = 0$. And then

$$x = \frac{|y|}{\rho(y)} y \in C$$

satisfies $\mathbf{\Phi}(x) = 0 = p$. □

EXERCISES

Some of these problems are from Kamien–Schwartz [**K-S**] and Bressan–Piccoli [**B-P**].

1. This and the next few problems provide practice for solving certain ODE. Find the general solution of

$$y' = a(x)y + b(x).$$

(Hint: Multiply by $e^{A(x)}$ for an appropriate function A.)

2. (a) If y satisfies

$$y' = a(x)y + b(x),$$

what ODE does $z = y^q$ solve?

(b) Use (a) to solve *Bernoulli's equation*

$$y' = a(x)y + b(x)y^p.$$

3. Find an implicit formula for the general solution of the separable ODE

$$y' = a(x)b(y),$$

where $b > 0$.

4. Assume y_1, y_2 both solve the linear, second-order ODE

$$Ly = y'' + b(x)y' + c(x)y = 0.$$

Find the first-order ODE satisfied by the *Wronskian* $w = y_2'y_1 - y_1'y_2$.

5. (a) Find two linearly independent solutions y_1, y_2 of the constant coefficient ODE

$$Ly = ay'' + by' + cy = 0.$$

where $a > 0$. (Hint: Try $y = e^{\lambda x}$. What if λ is a repeated root of the corresponding polynomial?)

(b) Find two linearly independent solutions y_1, y_2 of *Euler's equation*

$$Ly = ax^2y'' + bxy' + cy = 0.$$

6. If y_1 solves

$$Ly = y'' + b(x)y' + c(x)y = 0,$$

find another, linearly independent solution of the form $y_2 = zy_1$. (Hint: Show $w = z'$ satisfies a first-order linear ODE.)

7. Write down the Euler-Lagrange equations for the following Lagrangians:
(a) e^{z^2}
(b) $z^4 + y^2$
(c) $\sin(xyz)$
(d) y^3x^4.

8. Give an alternate proof that

$$L = a(y)z$$

is a null Lagrangian, by observing $I[y(\cdot)] = \int_0^1 L(y, y')\,dx$ depends only upon the values taken on by $y(\cdot)$ at the endpoints $0, 1$.

9. Compute (E-L) for

$$\int_0^1 a(x, y) + b(x, y)y'\,dx$$

and show that it is not a differential equation, but rather an implicit algebraic formula for y as a function of x.

10. (Discrete version of Euler-Lagrange equations) Let $h > 0$ and define $x_k = kh$ for $k = 0, \dots, N+1$. Consider the problem of finding $\{y_k\}_{k=1}^N$ to minimize

$$\sum_{k=0}^{N} L\left(x_k, y_k, \tfrac{y_{k+1}-y_k}{h}\right) h,$$

where $y_0 = y^0$, $y_{N+1} = y_1$ are prescribed. What algebraic conditions do minimizers satisfy? What is the connection with the Euler-Lagrange equation?

11. Assume
$$I[y(\cdot)] = \int_a^b L(x, y, y', y'')\,dx$$
for $L = L(x, y, z, u)$. Derive the corresponding Euler-Lagrange equation for extremals.

12. Show that the extremals of
$$\int_a^b (y')^2 e^{-y'}\,dx$$
are linear functions.

13. Find the extremals of
$$\int_0^1 (y')^2 + 10xy\,dx,$$
subject to $y(0) = 1, y(1) = 2$.

14. Find the extremals of
$$\int_0^T e^{-\lambda t}(a(\dot{x})^2 + bx^2)\,dt,$$
satisfying $x(0) = 0, x(T) = c$.

15. Assume $c > 0$. Find the minimizer of
$$\int_0^T e^{-\lambda t}(c(\dot{x})^2 + dx)\,dt,$$
subject to $x(0) = 0, x(T) = b$.

16. Find the minimizer of
$$\int_0^1 \frac{(y')^2}{2} - x^2 y\,dx,$$
where $y(0) = 1$ and $y(1)$ is free.

17. Consider the free endpoint problem of minimizing
$$I[y(\cdot)] = \int_a^b L(x, y, y')\,dx + g(y(b))$$
subject to $y(a) = y^0$, where $g : \mathbb{R} \to \mathbb{R}$ is a given function. What is the transversality condition satisfied at $x = b$ by a minimizer $y_0(\cdot)$?

18. Assume $y_0(\cdot)$ minimizes the functional
$$I[y(\cdot)] = \frac{1}{2}\int_a^b (y'')^2\,dx$$
over the admissible class
$$\mathcal{A} = \{y : [a, b] \to \mathbb{R} \mid y(a) = 0, y(b) = 1\}.$$

What additional boundary conditions does $y_0(\cdot)$ satisfy?

19. Give a different proof of Theorem 1.3.4 by rewriting

$$j(\sigma) = \frac{1}{1+\sigma} \int_0^T L\left(\tfrac{s}{1+\sigma}, x_0(s), (1+\sigma)\dot{x}_0(s)\right) ds$$

and then computing $\frac{dj}{d\sigma}(0)$.

20. Find the minimizer of

$$\int_0^1 (\dot{x})^2 \, dt,$$

subject to $x(0) = 0, x(1) = 2$ and $\int_0^1 x \, dt = b$.

21. Find a minimizer of $I[y(\cdot)] = \int_0^\pi (y')^2 \, dx$ over the admissible class

$$\mathcal{A} = \{y : [0,\pi] \to \mathbb{R} \mid y(0) = y(\pi) = 0, \int_0^\pi y^2 \, dx = 1\}.$$

22. Suppose $I[y(\cdot)] = \int_a^b L(x,y,y') \, dx$ and the admissible class is

$$\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid y(a) = 0, y(b) = 0, \ \phi(x) \le y(x) \ (a \le x \le b)\}$$

for some given function ϕ.

(a) Show that if $y_0(\cdot) \in \mathcal{A}$ is a minimizer, then

$$-\left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)' + \frac{\partial L}{\partial y}(x, y_0, y_0') \ge 0 \quad (a \le x \le b).$$

(b) Show that

$$-\left(\frac{\partial L}{\partial z}(x, y_0, y_0')\right)' + \frac{\partial L}{\partial y}(x, y_0, y_0') = 0$$

in the region $\{a \le x \le b \mid \phi(x) < y_0(x)\}$ where $y_0(\cdot)$ does not hit the constraint.

23. Let $I[\mathbf{x}(\cdot)] = \int_0^T |\dot{\mathbf{x}}| \, dt$ for functions belonging to the admissible class

$$\mathcal{A} = \{\mathbf{x} : [0,T] \to \mathbb{R}^n \mid \mathbf{x}(0) = A, \mathbf{x}(T) = B\},$$

where $A, B \in \mathbb{R}^n$ are given. Show that the graph of a minimizer $\mathbf{x}_0(\cdot)$ is a line segment from A to B.

24. For each Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ write out the Euler–Lagrange equation, the Hamiltonian H and the Hamiltonian differential equations:

(a) $L = \frac{m}{2}|v|^2 - W(x)$

(b) $L = \frac{m}{2}|v|^2 + \mathbf{b}(x) \cdot v$.

25. Assume that the matrix function G, whose (i,j)-th entry is g_{ij}, is symmetric and positive definite. What is the Hamiltonian corresponding to the Lagrangian
$$L = \frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(x) v_i v_j ?$$

26. Show that
$$L = \frac{m}{2}|v|^2 + qv \cdot \mathbf{A}(x)$$
is the Lagrangian corresponding to the Hamiltonian
$$H = \frac{1}{2m}|p - q\mathbf{A}(x)|^2.$$

27. If $\mathbf{x} : [0,1] \to U$ is a curve and $\mathbf{z} = \mathbf{y}(\mathbf{x})$ is its image under the coordinate patch $\mathbf{y} : U \to \mathbb{R}^l$, show that the length of the curve $\mathbf{z}$ in $\mathbb{R}^l$ is
$$\int_0^1 \Big(\sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \dot{x}_i \dot{x}_j \Big)^{\frac{1}{2}} dt.$$

28. Compute explicitly the Christoffel symbols Γ^k_{ij} for the hyperbolic plane. Check that the ODE given in the text for the geodesics is correct.

29. Suppose
$$\int_a^b \frac{\partial^2 L}{\partial z^2}(x, y_0, y_0')\zeta^2 \, dx \geq 0$$
for all functions ζ such that $\zeta(a) = \zeta(b) = 0$. Explain carefully why this implies
$$\frac{\partial^2 L}{\partial z^2}(x, y_0, y_0') \geq 0 \quad (a \leq x \leq b).$$

30. Assume $y(\cdot) > 0$ solves the linear, second-order ODE
$$y'' + b(x)y' + c(x)y = 0.$$
Find the corresponding *Riccati equation*, which is the nonlinear, first-order ODE that $w = \frac{y'}{y}$ solves.

31. What does the Weierstrass condition say about the possible values of $y_0'(\cdot)$ for minimizers of
(a) $\int_a^b ((y')^2 - 1)^2 \, dx$
(b) $\int_a^b \frac{1}{1+(y')^2} \, dx$
subject to given boundary conditions?

32. Derive this useful calculus formula, which we used in the proof of Theorem 2.3.1:
$$\frac{d}{dx}\left(\int_a^{g(x)} f(x,t)\,dt\right) = \int_a^{g(x)} \frac{\partial f}{\partial x}(x,t)\,dt + f(x,g(x))g'(x).$$
(Hint: Define $F(x,y) = \int_a^y f(x,t)\,dt$; so that $\int_a^{g(x)} f(x,t)\,dt = F(x,g(x))$. Apply the chain rule.)

33. Suppose that $y(\cdot)$ is an extremal of $I[\,\cdot\,]$, the second variation of which satisfies for some constant $\gamma > 0$ the estimate
$$\int_a^b A(w')^2 + 2Bww' + Cw^2\,dx \geq \gamma \int_a^b (w')^2 + w^2\,dx$$
for all $w : [a,b] \to \mathbb{R}$ with $w(a) = w(b) = 0$. Use a Taylor expansion to show directly that $y(\cdot)$ is a weak local minimizer; this means that there exists $\delta > 0$ such that
$$I[y] \leq I[\bar{y}]$$
for all admissible $\bar{y}$ satisfying $\max_{[a,b]}\{|y-\bar{y}| + |y'-\bar{y}'|\} \leq \delta$.

34. Show that for each $l > 0$ the function $y(\cdot) = 0$ is a strong local minimizer of
$$\int_0^l \frac{(y')^2}{2} + \frac{y^2}{2} - \frac{y^4}{4}\,dx$$
for
$$\mathcal{A} = \{y : [0,l] \to \mathbb{R} \mid y(0) = y(l) = 0\}.$$

35. Assume that $(y,z) \mapsto L(x,y,z)$ is convex for each $a \leq x \leq b$. Show that each extremal $y(\cdot) \in \mathcal{A}$ is in fact a minimizer of $I[\,\cdot\,]$, for the admissible set
$$\mathcal{A} = \{y : [a,b] \to \mathbb{R} \mid y(a) = y^0, y(b) = y^1\}.$$
(Hint: Recall that if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then
$$f(\hat{x}) \geq f(x) + \nabla f(x) \cdot (\hat{x} - x)$$
for all $\hat{x}$.)

36. A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex provided
$$f(\theta x + (1-\theta)\hat{x}) < \theta f(x) + (1-\theta) f(\hat{x})$$
if $x \neq \hat{x}$ and $0 < \theta < 1$.

Suppose that $(y,z) \mapsto L(x,y,z)$ is strictly convex for each $a \leq x \leq b$. Show that there exists at most one minimizer $y_0(\cdot) \in \mathcal{A}$ of $I[\,\cdot\,]$.

37. Explain carefully why

$$I[y(\cdot)] = \int_0^1 y\,dx$$

does not have a minimizer over the admissible set

$$\mathcal{A} = \{y : [0,1] \to \mathbb{R} \mid y(\cdot) \text{ is continuous}, \ y \geq 0, y(0) = 0, y(1) = 1\}.$$

38. Write down the Euler-Lagrange PDE for the following Lagrangians:
(a) $\frac{|z|^2}{2} + \mathbf{b}(x) \cdot z$
(b) $\frac{|z|^p}{p} \quad (1 \leq p < \infty)$
(c) $\left(1 + |z|^2\right)^{\frac{1}{2}} + F(y)$.

39. Write down the Euler-Lagrange PDE for these functionals:
(a) $I[u] = \int_U \frac{|\nabla u|^2}{2} + F(x,u)\,dx$
(b) $I[u] = \frac{1}{2}\int_U \sum_{k,l=1}^n a_{kl}(x) \frac{\partial u}{\partial x_k}\frac{\partial u}{\partial x_l}\,dx$.

40. Use the Divergence Theorem to give another proof that

$$L = z \cdot \mathbf{b}(x) f(y) + F(y)\operatorname{div}\mathbf{b},$$

is a null Lagrangian, where $F' = f$.
(Hint: $L(x, u, \nabla u) = \operatorname{div}(F(u)\mathbf{b})$.)

41. Derive the 4-th order Euler-Lagrange PDE satisfied by minimizers of

$$I[u(\cdot)] = \frac{1}{2}\int_U (\Delta u)^2\,dx,$$

subject to the boundary conditions $u = g$ on ∂U. What is the transversality condition on ∂U for a minimizer?

42. Consider the system of PDE

$$\begin{cases} -\Delta u_1 = u_2 \\ -\Delta u_2 = 2u_1, \end{cases}$$

for the unknown $\mathbf{u} : \mathbb{R}^2 \to \mathbb{R}^2$, $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$.

Show that this problem is variational. This means to find a Lagrangian function $L : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{M}^{2\times 2} \to \mathbb{R}$ so that the two PDE above are the Euler-Lagrange equations for $I[\mathbf{u}(\cdot)] = \int_{\mathbb{R}^2} L(x, \mathbf{u}, \nabla\mathbf{u})\,dx$.

43. A system of two *reaction-diffusion equations* has the form

$$\begin{cases} -a_1\Delta u_1 = f_1(u_1, u_2) \\ -a_2\Delta u_2 = f_2(u_1, u_2). \end{cases}$$

Under what conditions on the functions $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$ is this system variational?

44. Consider the dynamics

$$\begin{cases} \dot{x}(t) = \alpha(t) \quad (0 \leq t \leq 1) \\ x(0) = x_0 \end{cases}$$

and payoff functional

$$P[\alpha(\cdot)] = \int_0^1 |x(t)|\, dt.$$

(a) Suppose $A = \{-1, 1\}$. If $|x_0| \geq 1$, describe an optimal control that minimizes $P[\alpha(\cdot)]$. Explain why an optimal control does not exist if $|x_0| < 1$.

(b) Suppose instead that $A = \{-1, 0, 1\}$. Explain why an optimal control exists.

45. Consider the problem of reaching the origin in least time, when the dynamics are

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 + \alpha, \end{cases}$$

where $|\alpha| \leq 1$.

(a) Check that $\mathbf{X}(t) = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$.

(b) Show that an optimal control $\alpha_0(\cdot)$ is periodic in time.

46. Use the maximum principle to find an optimal control for the linear time optimal problem with dynamics

$$\begin{cases} \dot{x}_1 = x_2 + \alpha_1 \quad |\alpha_1| \leq 1 \\ \dot{x}_2 = -x_1 + \alpha_2 \quad |\alpha_2| \leq 1 \end{cases}$$

47. Write down (ODE), (ADJ), (M) and (T) for the fixed time, free endpoint problems with $n = m = 1$ and

(a) $f = (x^2 + a)^2 - x^4, A = [-1, 1], r = 2ax, g = \sin x$

(b) $f = x^2 a, A = [0, 2], r = a^2 + x^2, g = 0.$

48. Write down the equations (ADJ), (M) and (T) for the fixed time, free endpoint problem corresponding to the dynamics

$$\begin{cases} \dot{x}_1 = \sin(x_1 + \alpha x_2) \\ \dot{x}_2 = \cos(\alpha x_1 + x_2), \end{cases}$$

where $0 \le \alpha \le 1$, with the payoff

$$P[\alpha(\cdot)] = \int_0^T \alpha^4 + (x_1 x_2)^2 \, dt + (x_1(T))^4.$$

49. Consider the variational problem of minimizing

$$I[\mathbf{x}(\cdot)] = \int_0^T L(\mathbf{x}(t), \dot{\mathbf{x}}(t)) \, dt,$$

for $L = L(x, v)$, over the admissible class

$$\mathcal{A} = \{\mathbf{x} : [0,T] \to \mathbb{R}^n \mid x(\cdot) \text{ is continuous and piecewise} \\ \text{continuously differentiable}, \mathbf{x}(0) = x^0, \mathbf{x}(T) = x^1\}.$$

(a) Show how to interpret this as a fixed time, fixed endpoint optimal control problem.

(b) If $\mathbf{x}_0$ is a minimizer, show that (M) implies $\mathbf{p}_0 = \nabla_v L(\mathbf{x}_0, \dot{\mathbf{x}}_0)$.

(c) Show that (ADJ) implies the Euler-Lagrange equations.

50. Use the free time, fixed endpoint PMP to give a new proof of Theorem 1.3.4 for a Lagrangian $L = L(x, v)$. In particular, explain what condition (T) says for the variational problem.

51. Solve the linear-quadratic regulator problem of minimizing

$$\int_0^T x^2 + \alpha^2 \, dt + \frac{x^2(T)}{2}$$

for the dynamics

$$\begin{cases} \dot{x} = x + \alpha \\ x(0) = x^0 \end{cases}$$

with controls $\alpha : [0,T] \to \mathbb{R}$.

52. Assume that a function $z : [0,T] \to \mathbb{R}$ is given and we wish to minimize

$$\int_0^T (x - z)^2 + \alpha^2 \, dt,$$

where $A = \mathbb{R}$ and

$$\begin{cases} \dot{x} = x + \alpha \\ x(0) = x^0. \end{cases}$$

Show

$$\alpha = \frac{1}{2} dx + \frac{e}{2}$$

is an optimal feedback control, where $d(\cdot)$ and $e(\cdot)$ solve

$$\begin{cases} \dot{d} = 2 - 2d - \frac{1}{2}d^2, & d(T) = 0 \\ \dot{e} = -2z - \left(1 + \frac{d}{2}\right)e, & e(T) = 0. \end{cases}$$

(Hint: Assume PMP applies and write down the equations for x and p. Look for a solution of the form $p = dx + e$.)

53. Find explicit formulas for the optimal state $x_0(\cdot)$ and costate $p_0(\cdot)$ for the production and consumption model discussed on page 108. Show that $H(x_0, p_0, \alpha_0)$ is constant on the interval $[0, T]$, as asserted by the PMP.

54. How does our analysis of the Ramsey consumption model break down if we drop the requirement that $x(T) = x^1$?

55. Use the PMP to solve the problem of maximizing

$$P[\alpha(\cdot)] = \int_0^2 2x - 3\alpha + \alpha^2\, dt,$$

where

$$\begin{cases} \dot{x} = x + \alpha, & 0 \leq \alpha \leq 2 \\ x(0) = 0. \end{cases}$$

56. Use the PMP to find a control to minimize the payoff

$$P[\alpha(\cdot)] = \frac{1}{4}\int_0^1 \alpha^4\, dt,$$

for the dynamics

$$\begin{cases} \dot{x} = x + \alpha \\ x(0) = 1,\ x(1) = 0, \end{cases}$$

where $A = \mathbb{R}$.

57. Assume that the matrices B, C, D are symmetric and that the matrix Riccati equation

$$\begin{cases} \dot{K} + KNC^{-1}N^TK + KM + M^TK = B \quad (0 \leq t \leq T) \\ \qquad\qquad\qquad\qquad K(T) = -D \end{cases}$$

has a unique solution $K(\cdot)$. Show that $K(t)$ is symmetric for each time $0 \leq t \leq T$.

58. Apply the PMP to solve the general linear-quadratic regulator problem, introduced on page 137. In particular, show how to solve (ADJ) for the costate $\mathbf{p}_0(\cdot)$ in terms of the matrix Riccati equation.

59. This exercise discusses the infinite horizon problem:

$$\begin{cases} \dot{x} = -x + \alpha \quad (t \geq 0) \\ x(0) = x, \end{cases}$$

where $A = \mathbb{R}$ and

$$P_x[\alpha(\cdot)] = \int_0^\infty x^2 + \alpha^2 \, dt$$

Define

$$v(x) = \inf_{\alpha(\cdot)} P_x[\alpha(\cdot)].$$

(a) Derive the HJB equation for v.

(b) Solve this equation and design an optimal feedback control.

60. Use dynamic programming to solve the tracking problem of minimizing

$$\int_0^T |\mathbf{x}(t) - \mathbf{z}(t)|^2 + |\alpha(t)|^2 \, dt$$

for the dynamics

$$\begin{cases} \dot{\mathbf{x}}(t) = M\mathbf{x}(t) + N\alpha(t) \qquad (0 \leq t \leq T) \\ \mathbf{x}(0) = 0, \end{cases}$$

where $\mathbf{z} : [0, T] \to \mathbb{R}^n$ is given. Here $M \in \mathbb{M}^{n \times n}$, $N \in \mathbb{M}^{n \times m}$, and $A = \mathbb{R}^m$.

Bibliography

[A-F] M. Athans and P. L. Falb, *Optimal Control: An Introduction to the Theory and its Applications*, Dover, 2007

[B-M] J. Ball and F. Murat, Remarks on rank-one convexity and quasiconvexity, in *Ordinary and Partial Differential Equations*, B. D. Sleeman and R. J. Jarvis, eds., Pitman Research Notes, Pitman, 1991.

[B-P] A. Bressan and B. Piccoli, *Introduction to the Mathematical Theory of Control*, AIMS Series on Applied Math, Vol 2, American Institute of Mathematical Sciences, 2007

[C-Z] F. H. Clarke and V. Zeidan, Sufficiency and the Jacobi condition in the calculus of variations, Canad. J. Math (38) 1986, 1199–1209.

[E] L. C. Evans, *An Introduction to Mathematical Optimal Control Theory*, Version 0.2, lecture notes available at `math.berkeley.edu/~evans/control.course.pdf`

[F-R] W. Fleming and R. Rishel, *Deterministic and Stochastic Optimal Control*, Springer, 1975

[G] T. Gilbert, Lost mail: the missing envelope in the problem of the minimal surface of revolution, American Math Monthly (119) 2012, 359–372

[K-S] M. Kamien and N. Schwartz, *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, 2nd ed, Dover, 2012

[K] M. Kot, *A First Course in the Calculus of Variations*, Student Mathematical Library, Vol 72, American Math Society, 2014

[L-M] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, 1967

[L] M. Levi, *Classical Mechanics with Calculus of Variations and Optimal Control*, Student Mathematical Library, Vol 69, American Math Society, 2014

[M-S] J. Macki and A. Strauss, *Introduction to Optimal Control Theory*, Springer, 1982

[M] Z. A. Melzak, *Mathematical Ideas, Modeling & Applications*, Wiley–Interscience, 1976

[MG] M. Mesterton-Gibbons, *A Primer on the Calculus of Variations and Optimal Control Theory*, Student Mathematical Library, Vol 50, American Math Society, 2009

[S] D. R. Smith, *Variational Methods in Optimization*, Prentice Hall, 1974

[T] J. R. Taylor, *Classical Mechanics*, University Science Books, 2005