Contents

1 Introduction

1.1 Motivation

We begin by introducing some famous questions in an effort to help motivate the course. The first such problem that we shall consider is finding geodesic paths; that is, minimal distances between points on a manifold. Dido's problem (also known as the isoperimetric inequality, circa 850 BCE) is the problem of finding the optimal shape to enclose a maximal area under a string of fixed length. Also of interest is the Brachistochrone problem, in which one is tasked with finding a path that optimizes the fall time of a ball travelling between two points, solely under the force of gravity.

In considering more applicable problems, such as the deflection of a beam under a load, we can exploit Bernoulli's principle. This principle states that the potential energy of a system is minimized at equilibrium. For example, one might be asked to find the function $y(x)$ that minimizes the cost function

$$J(y) = \int_0^L \frac{1}{2} u y''(x)^2 - p(x)y(x)ds$$

Exclusive to more modern motifs, we can consider applications in the field of robotics. Important for these questions will be determining the equations of motion, for which we will examine Hamilton's principle of minimizing the action.

After the notions of variational calculus have been properly examined, we will be able to consider applications in optimal control. An example of optimal control is the trajectory of a rocket. If we coordinatize the rocket's position in three space as (y_1, y_2, y_3), and let u_1, u_2 be the thrust angle and mass flow rate, our equations are governed by

$$\begin{aligned} \ddot{y}_1(t) &= \tfrac{c(\cos(u_1(t)))u_2(t)}{y_3(t)} & y_1(0) = 0, y_1(T) = y_{1f} \\ \ddot{y}_2(t) &= \tfrac{c(\sin(u_1(t)))u_2(t)}{y_3(t)} - g & y_2(0) = 0, y_2(T) = y_{2f} \\ \ddot{y}_3(t) &= -u_2(t) & y_3(0) = m_1 \end{aligned}$$

If we want to reach the destination that minimizes fuel, we would consider the cost integral $\int_0^T u_2(t)dt$. Similarily, minimizing time allows us to consider the cost function $\int_0^T dt$.

We can summarize our previous discussion in the following table

Problem	**Objective**	**Variable**
Geodesics	Length	curve in $\mathbb{R}^2$
Dido's Problem	Area	curve in $\mathbb{R}^2$
Brachistocrhone	Time	curve (function)
Steering	Time	angle
Beam	Potential Energy	deflection $y(x)$
Rocket	Time/Fuel Use	u(t)

In general, we can see that our purpose has been to find a function (or curve) to minimize some real-valued quantity. At this point in our academic career, when we have been tasked with minimizing a quantity, it has usually been to find a point that minimizes a curve. Our purpose for this course is to extend these ideas to find *functions* to minimize functions.

1.2 Review of Multivariate Calculus

1.2.1 Optimization in $\mathbb{R}^n$

In order to find optimal functions, it will be useful to define an object to optimize. We call this the *cost function*. One can think of our goal being to find a function that minimizes the cost (however we define it). Before we analysis the dynamics of more complex cost functionals, we shall first quickly review some fundamental results from multivariate calculus. We will make great use of these notions and their generalizations in the future.

Consider a real valued function $J : \mathbb{R} \to \mathbb{R}$, and define a subdomain $\mathcal{D} = [a, b] \subset \mathbb{R}$. Let's suppose that we wish to evaluate $\min_{y \in \mathcal{D}} J(y)$.

Definition 1.1. Given sets X, Y (for our purposes, some cartesian product $\mathbb{R}^m$), define

$$C^n(X, Y) = \left\{ f : X \to Y | f \text{ has continuous } n^{th} \text{ derivatives } \right\}$$

Note that in general, we shall take the codomain to be $\mathbb{R}$. In such cases where this is unambiguous, we may just denote this as $C^n(X, Y) = C^n(X)$.

Definition 1.2. The point $y' \in \mathcal{D}$ is said to **minimize** J over $\mathcal{D}$ if $J(y') \leq J(Y), \forall y \in \mathcal{D}$. Furthermore, the point y' is said to be a **local minimum** of J if $\exists \epsilon > 0$ such that $J(y') \leq J(y), \forall y \in B_\epsilon(y')$. Here we've defined

$$B_\epsilon(y') = \left\{ x \in \mathcal{D} \,\middle|\, d(x, y') < \epsilon \right\} = \left\{ x \in \mathcal{D} \,\middle|\, |x - y'| < \epsilon \right\}$$

where $d(x, y)$ is the Euclidean metric.

Definition 1.3. If $\lim_{v \to 0} \dfrac{J(y + v) - J(y)}{h}$ exists, we say that the limit point is the **derivative** of J at y. J is said to be **differentiable** at y if there exists D so that $J(y + v) = J(y) + Dv + R$ where $\lim_{v \to 0} \dfrac{R}{v} = 0$.

Example: Calculate $J(y) = y^2$ using the limit definition of the derivative

$$\begin{aligned}
\lim_{v \to 0} \frac{J(y + v) - J(y)}{v} &= \lim_{v \to 0} \frac{y^2 + 2yv + v^2 - y^2}{v} \\
&= \lim_{v \to 0} \frac{2yv + v^2}{v} \\
&= \lim_{v \to 0} 2y + v \\
&= 2y
\end{aligned}$$

Theorem 1.4. *Let $\mathcal{D} \subseteq \mathbb{R}$ and assume that J is differentiable. If $y' \in U$ is a local minimum, then $J'(y') = 0$.*

Proof. Assume that $v > 0$. By definition of a local minimum

$$\begin{aligned} J(y') &\leq J(y'+v), \text{ for sufficiently small } v \\ 0 &\leq \frac{J(y'+v) - J(y')}{v} \\ 0 &\leq \lim_{v \to 0} \frac{J(y'+v) - J(y')}{v} \\ 0 &\leq J'(y') \end{aligned}$$

Similarily, if we assume that $v < 0$, then we have that $0 \geq J'(y')$, and hence we can conclude that $J'(y') = 0$. □

Note that $J'(y') = 0$ does not necessarily imply that y' is a local minimum or maximum. Furthemore, on a closed interval $[a, b]$, $J(y)$ might not attain its extremal values on the interior points (it may occur at a or b), and that for a non-closed set, for example (a, b), the extremals of $J(y)$ might not exist.

We shall now extend these notions to the multivariate case.

Definition 1.5. Let $J : \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{D} \subseteq \mathbb{R}^n$. The point $y' \in \mathcal{D}$ minimizes J over $\mathcal{D}$ if $J(y') \leq J(y), \forall y \in \mathcal{D}$.

The point y' is a local minimum of J if $\exists \epsilon > 0$ such that $J(y') \leq J(y) \forall y \in B_\epsilon(y')$. In this case

$$B_\epsilon(y') = \left\{ x \in \mathcal{D} \;\middle|\; d(x, y') < \epsilon \right\} = \left\{ x \in \mathcal{D} \;\middle|\; \|x - y\|_2 < \epsilon \right\}$$

If J is differentiable at y', there is a vector $\nabla J(y')$ such that

$$J(y' + v) = J(y') + \nabla J(y') \cdot v + R$$

where $\lim_{\|v\| \to 0} \frac{|R|}{\|v\|} = 0$

Example: Consider the function $f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$. This function is continuous, and the partial derivatives exists as

$$\begin{aligned} f_x(0, 0) &= \lim_{v \to 0} \frac{f(v, 0) - f(0, 0)}{v} = \lim_{v \to 0} \frac{0 - 0}{v} \\ &= 0 \end{aligned}$$

Similarily, $f_y(0, 0) = 0$ and so $\nabla f(0, 0) = \vec{0}$.

Given an arbitrary vector $v = (v_1, v_2)$, the corresponding *directional derivative* with respect to v at $(0, 0)$ is

$$\begin{aligned}
\lim_{s\to 0} &= \frac{f(0+sv_1, 0+sv_2) - f(0,0)}{s} \\
&= \lim_{s\to 0} \frac{\frac{s^2 v_1 v_2}{\sqrt{s^2 v_1^2 + s^2 v_2^2}}}{s} \\
&= \lim_{s\to 0} \frac{s^2 v_1 v_2}{|s| s\sqrt{v_1^2 + v_2^2}} \\
&= \frac{v_1 v_2}{\sqrt{v_1^2 + v_2^2}}, \qquad\qquad s > 0
\end{aligned}$$

Theorem 1.6. *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be an open set, and assume J is differentiable. If $y' \in I$ is a local minimum, then $\nabla J(y') = \vec{0}$.*

Proof. Fix $i \in \{1, \ldots, n\}$ and let $e_i = (\underbrace{0, \ldots, 0, 1}_{i \text{ times}}, 0, \ldots, 0)$, the standard basis in $\mathbb{R}^n$. Define the function $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = J(y' + te_i)$. Now since J has a minimum at y', we immediately see that g must have a minimum at 0 since $g(0) = J(y')$. By Theorem 1.4 we then have that $0 = g'(0) = \frac{\partial f}{\partial x_i}(y')$. Since i was chosen arbitrarily, we can conclude that this must hold $\forall i = 1, \ldots, n$. Hence

$$\nabla J(y') = \left(\frac{\partial f}{\partial x_1}(y'), \ldots, \frac{\partial f}{\partial x_n}(y') \right) = \vec{0}$$

as required. □

Definition 1.7. A point at which $\nabla J(y) = y$ is called a **stationary point** or **critical point**

1.3 Convexity in $\mathbb{R}^n$

We recall from our studies in calculus that for a function f, not all points for which $f'(x) = 0$ necessarily gives rise to an extremal. A useful, albeit sometimes complicated, method for checking whether such critical points did indeed give rise to maximum/minimum values was checking the concavity of the function at the critical point. As we advance our studies to multi-dimensional surfaces and eventually functionals, it will be useful to consider a similar idea of concavity. We will see that this notion will allow us to stipulate over the existence (and uniqueness) of solutions to the minimization problem.

Definition 1.8. A set $\mathcal{D} \subseteq \mathbb{R}^n$ is **convex** if $(x, y) \in \mathcal{D}$ implies that $\alpha x + (1 - \alpha) \in \mathcal{D}$ for $0 < \alpha < 1$. Note that geometrically, this implies that any two points can be connected by a straight line. From here on, it will be assumed that a set $\mathcal{D}$ is always convex unless stated otherwise.

Definition 1.9. A function $J : \mathbb{R}^n \to \mathbb{R}$ is a **convex function** on $\mathcal{D} \subseteq \mathbb{R}^n$ if

$$J(\alpha x + (1-\alpha)y) = J(y + \alpha(x - y)) \leq \alpha J(x) + (1 - \alpha)J(y)$$

If the function satisfies the definition in a strict sense, we say that the function is **strictly convex**.

If we restrict ourselves to the interval $[a, b]] \subset \mathbb{R}$, we can gain some geometric intuition as to what a convex function is. Geometrically, a function is convex on the interval $[a, b]$ if the image of the function always lies beneath the secant line joining a to b.

Note that because we have decided to consider the problem of minimizing J over a domain $\mathcal{D}$, we want to consider convex functions which naively imply the existence of a minimum. If conversely, we wanted to consider the problem of maximizing J over $\mathcal{D}$, we would want to consider the case of **concave** function. A function f is said to be concave if $-f$ is convex.

Theorem 1.10. *Let $\mathcal{D} \subseteq \mathbb{R}^n$ and $J : \mathbb{R}^n \to \mathbb{R}$. Then if J is convex on $\mathcal{D}$, it has a directional derivative at every $y \in \mathcal{D}$.*

Proof. Recall that $\delta J(y; v) = \lim_{t \searrow 0} \dfrac{J(y + tv) - J(y)}{t}$ and define $f(t) = \dfrac{J(y + tv) - J(y)}{t}$, then

Claim 1 f is a monotonically increasing function.

For $0 < t < s$ we have

$$\begin{aligned} J(y + tv) &= J\left(\left(1 - \frac{t}{s}\right) y + \frac{t}{s}(y + sv)\right) \\ &\leq \left(1 - \frac{t}{s}\right) J(y) + \frac{t}{s} J(y + sv) \end{aligned}$$

We can now rearrange this last equation to find that

$$\underbrace{\frac{J(y + tv) - J(y)}{t}}_{f(t)} \leq \underbrace{\frac{J(y + sv) - J(y)}{s}}_{f(s)}$$

So we've shown that $f(t)$ is monotonically increasing.

Claim 2 f is bounded below.

Let us take $t_0 < 0 < t$. Then we have

$$\begin{aligned} y &= \frac{t}{t - t_0}(y + t_0 v) - \frac{t_0}{t - t_0}(y + tv) \\ &= \underbrace{\frac{t}{t - t_0}}_{\alpha}(y + t_0 v) + \left(\underbrace{1 - \frac{t}{t - t_0}}_{1-\alpha}\right)(y + tv) \end{aligned}$$

Thus we can conclude that

$$J(y) \leq \frac{t}{t - t_0} J(y + t_0 v) + \left(1 - \frac{t}{t - t_0}\right) J(y + tv)$$

Rearranging thus yields

$$\begin{aligned} (t - t_0) J(y) &\leq t J(y + t_0 v) + (\cancel{t} - t_0 - \cancel{t}) J(y + tv) \\ t(J(y) - J(y + t_0 v)) &\leq t_0 (J(y) - J(y + tv)) \\ \frac{J(y) - J(y + t_0 v)}{t_0} &\geq \frac{J(y) - J(y + tv)}{t} \\ f(t_0) &\leq f(t) \end{aligned}$$

Thus, $f(t)$ is bounded below as $t \to 0$.

Finally, since f is bounded from below and monotonically increasing, we can conclude that $\lim_{t \searrow 0} f(t)$ exists and so $\delta J(y; v)$ exists. □

Definition 1.11. If $\forall v$ and a fixed $y' \in \mathcal{D} \subseteq \mathbb{R}^n$, we have that $\delta J(y'; v)$ exists and $\delta J(y'; v) = 0$ then y' is a **stationary point** of J.

Theorem 1.12. *If $J : \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathcal{D}$ then every stationary point minimizes J on $\mathcal{D}$.*

Proof. By definition $J(y+\alpha v) - J(y) \leq \alpha J(y+v) + (1-\alpha)J(y) - J(y)$. Rearranging and taking the limit as $\alpha \searrow 0$ this equation yields

$$\lim_{\alpha \searrow 0} J(y+v) \geq \lim_{\alpha \searrow 0} J(y) + \frac{J(y+\alpha v) - J(y)}{\alpha}$$
$$J(y+v) \geq J(y) + \delta J(y; v)$$

If y is a stationary point $\delta J(y; v) = 0$ so $J(y+v) \geq J(y)$ for all v. □

Note that this theorem gives us a test for minimality without needing to invoke second derivatives.

Theorem 1.13. *Assume that $J \in C^2(\mathcal{D})$. Then J is convex at y if and only if the Hessian of J is non-negative on $\mathcal{D}$.*

Proof. To avoid tedious details, we will present only the outline of the proof. Notice that how we proceed is very similar to how we extended Theorem 1.4 to Theorem 1.6.

1. Show this first for $J : \mathbb{R} \to \mathbb{R}$
2. For $J : \mathbb{R}^n \to \mathbb{R}$, consider $\tilde{J}(t) = J(y + tv)$

□

Theorem 1.14. *Let $\mathcal{D} \subseteq \mathbb{R}^n$, and $J : \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathcal{D}$. Then J is convex on $\mathcal{D}$ if and only if $J(y+v) \geq J(y) + \nabla J(y) \cdot v, \quad \forall y, y+v \in \mathcal{D}$.*

Example: Determine whether the following functions are convex

1. $f(x_1, x_2) = x_1^3 + x_2$
2. $f(x_1, x_2) = (x_1 - x_2)^2$
3. $f(x_1, x_2) = (x_1^2 + x_2^2 - 2x_1)$

Solution:

1. Let $f(x_1, x_2) = x_1^3 + x_2$. Via (1.14) we want to determine whether or not $f(x_1 + v_1, x_2 + v_2) - f(x) - \nabla f(\vec{x}) \cdot \vec{v} \geq 0$. Let us make the appropriate substitutions to find

$$\begin{aligned} f(x_1 + v_1, x_2 + v_2) - f(x) - \nabla f(x) \cdot v &= (x_1 + v_1)^3 + (x_2 + v_2) - x_1^3 - x_2 - (3x_1^2, 1) \cdot (v_1, v_2) \\ &= (3x_1 + v_1)v_1^2 \end{aligned}$$

   Without too much trouble we can easily see that $\exists x_1, v_1$ such that $3x_1 + v_1 < 0$. We conclude that f is not convex

2. Consider $G(x_1, x_2) = (x_1 - x_2)^2$. Again, we can make the appropriate substitutions to see that

$$\begin{aligned} g(x_1 + v_1, x_2 + v_2) - g(x) - \nabla g(\vec{x}) \cdot \vec{v} &= (x_1 + v_1 - x_2 - v_2)^2 - (x_1 - x_2)^2 + 2(x_1 - x_2, x_2 - x_1) \cdot (v_1, v_2) \\ &= x_1^2 + x_2^2 + v_1^2 + v_2^2 + 2x_1v_1 - 2x_1v_1 - 2x_1v_2 - 2v_1v_2 - 2x_2v_1 + \\ &\quad 2x_2v_2 - \left[(x_1^2 - 2x_1x_2 + x_2^2) + 2x_1v_1 - 2x_2v_2 - 2x_1v_2 + 2x_2v_2\right] \\ &= v_1^2 + v_2^2 - 2v_1v_2 \qquad \text{after appropriate cancellation} \\ &= (v_1 - v_2)^2 \geq 0 \end{aligned}$$

   We notice that this quantity is always non-negative, albeit will be zero whenever $v_1 = v_2$. Thus g is convex though not strictly convex.

3. Take $h(x_1, x_2) = x_1^2 + x_2^2 - 2x_1$. We can proceed in a similar manner above to find that

$$h(x_1 + v_1, x_2 + v_2) - h(x_1, x_2) - \nabla h(x_1, x_2) \cdot (v_1, v_2) = v_1^2 + v_2^2 \geq 0$$

   Since there is no non-trivial variation (v_1, v_2) for which equality holds, we can conclude that h is strictly convex

1.4 Linear Spaces

Often introduced in introductory courses to linear algebra is the concept of a linear space (or vector space). These spaces consist of sets with the additional structure of linear addition and scaling, and together with certain closure properties allow for a rich theory. More rigourously, we define a vector space as follows

Definition 1.15. A **real linear space** is a triple $(Y, +, \cdot)$ such that $S \subset \mathbb{R}^n$, $+ : S \times S \to S$ is a binary operator, and $\cdot : \mathbb{R} \times S \to S$ is an external binary operator satisfying the following axioms:

1)	$\forall y_1, y_2 \in Y$	$y_1 + y_2 \in Y$	closure under addition
2)	$\forall \alpha \in \mathbb{R}, y \in Y$	$\alpha y \in Y$	closure under scalar multiplication
3)	$\forall y_1, y_2 \in Y$	$y_1 + y_2 = y_2 + y_1$	additive commutativity
4)	$\forall y_1, y_2, y_3 \in Y$	$(y_1 + y_2) + y_3 = y_1 + (y_2 + y_3)$	additive associativity
5)	$\exists \theta \in Y, \forall y \in Y$	$y + \theta = y$	additive identity
6)	$\forall y \in Y, \exists \bar{y} \in Y$	$y + \bar{y} = \theta$	additive inverse
7)	$\forall \alpha \in \mathbb{R}, y_1, y_2 \in Y$	$\alpha(y_1 + y_2) = \alpha y_1 + \alpha y_2$	distributivity
8)	$\forall y \in y$	$1y = y$	scalar identity

It is likely that the student has been exposed to many such vector spaces before, though without necessarily consciously aware of this fact. The following are some examples of linear spaces.

Examples:

1. Consider the space $\mathbb{R}^n$. This is a linear space under componentwise addition and scalar multiplication.That is, given a basis $\left\{\vec{b}_n\right\} \subset \mathbb{R}^n$, and vectors satisfying $\vec{v}_j = \sum_{i=1}^{n} a_{j,i}\vec{b}_i, \quad j = 1, 2$ we have that vector addition and scalar multiplication are defined as follows:

$$\vec{v}_1 + \vec{v}_2 = \sum_{i=1}^{n} a_{1,i}\vec{b}_i + \sum_{i=1}^{n} a_{2,i}\vec{b}_i = \sum_{i=1}^{n}(a_{1,i} + a_{2,i})\vec{b}_i$$

$$\alpha \in \mathbb{R}, \qquad \alpha\vec{v}_1 = \alpha \sum_{i=1}^{n} a_{1,i}\vec{b}_i = \sum_{i=1}^{n}(\alpha a_{1,i})\vec{b}_i$$

   The additive identity is the zero vector, defined as $\vec{\theta} = \sum_{i=1}^{n} 0\vec{b}_i$.

2. The set of real (or complex) $m \times n$ matrices $M_{m\times n}(\mathbb{R})$ is also a vector space. Let $(M)_{ij}$ denote the i^{th} row and j^{th} column of such a matrix. The additive identity is defined as $(\theta)_{ij} = 0$. The additive binary operator and scalar multiplication are done in a pointwise fashion, so that if $M_1, M_2 \in M_{m\times n}(\mathcal{F})$ and $\alpha \in \mathbb{R}$ then

$$(M_1 + M_2)_{ij} = (M_1)_{ij} + (M_2)_{ij} \qquad (\alpha M)_{ij} = \alpha M_{ij}$$

3. Let $C[a, b] = \{f : [a, b] \to \mathbb{R} | f \text{ is continuous on } [a, b]\}$. Since the image of f is contained in $\mathbb{R}$, we shall define addition in $C[a, b]$ to be pointwise. That is, for $f, g \in C[a, b], \alpha \in \mathbb{R}$ we have that

$$(f + g)(x) = f(x) + g(x), \qquad (\alpha f)(x) = \alpha f(x), \quad \forall x \in [a, b]$$

   The additive inverse is the constant function $\theta(x) = 0, \forall x \in [a, b]$.

Exercise: Determine whether the following are linear spaces

$$D = \left\{ y \in C[a, b] \,\middle|\, y(a) = y(b) = 0 \right\}$$
$$D = \left\{ y \in C[a, b] \,\middle|\, y(a) = 0, y(b) = 2 \right\}$$
$$D = \left\{ y \in C^2[a, b] \,\middle|\, y(a) = y(b) = 0, y'(a) = y'(b) = 0 \right\}$$

1.5 Gateaux Derivatives

Consider a function $J : \mathbb{R}^n \to \mathbb{R}$. Then the directional derivative is given by

$$\lim_{\epsilon \searrow 0} \frac{J(y + \epsilon v) - J(y)}{\epsilon}$$

If J is differentiable then $\delta J(y; v) = \nabla J(y) \cdot v$.

Definition 1.16. For $J : Y \to \mathbb{R}, Y$ a linear space and $y, v \in Y$, define

$$\delta J(y; v) = \lim_{\epsilon \searrow 0} \frac{J(y + \epsilon v) - J(y)}{\epsilon}$$

Assuming this limit exists, we refer to this as the **Gateaux derivative** (or a variation).

Note that y, v are taken to be fixed, and the limit with respect to ϵ assumes that $J(y + \epsilon v)$ is defined for all small valued ϵ.

Examples:

1. Consider the function $J(y) = y(a)^3$. Choose $Y = C[a,b]$ so that $J : C[a,b] \to \mathbb{R}$. Then

$$\begin{aligned}
\frac{J(y+\epsilon v) - J(y)}{\epsilon} &= \frac{(y(a)+\epsilon v(a))^3 - y(a)^3}{\epsilon} \\
&= \frac{(y(a)+\epsilon v(a))(y(a)^2 + 2\epsilon v(a)y(a) + \epsilon^2 v(a)^2) - y(a)^3}{\epsilon} \\
&= v(a)y(a)^2 + 2\epsilon v(a)^2 y(a) + \epsilon^2 v(a)^3 + 2y(a)^2 v(a) + \epsilon y(a)v(a)^2 \\
\lim_{\epsilon\to 0} \frac{J(y+\epsilon v) - J(y)}{\epsilon} &= 3v(a)y(a)^2
\end{aligned}$$

2. Consider the function $J(y) = \int_0^5 1 + y'(x)^2 dx$. Choose $C^1[0,5] = Y$ such that $J : C^1[0,5] \to \mathbb{R}$. Then

$$\begin{aligned}
J(y+\epsilon v) &= \int_0^5 1 + \left(y'(x) + \epsilon v'(x)\right)^2 dx \\
J(y) &= \int_0^5 1 + y'(x)^2 dx \\
\lim_{\epsilon\to 0} \frac{J(y+\epsilon v) - J(y)}{\epsilon} &= \lim_{\epsilon\to 0} \frac{2y'\epsilon v' + \epsilon^2 v'^2}{\epsilon} dx \\
&= \lim_{\epsilon\to 0} \int_0^5 2v'y' + \epsilon v'^2 dx \\
&= \int_0^5 2v'y' dx
\end{aligned}$$

For all $y, v \in C^1[0,5]$ we have that $\delta J(y;v)$ exists and so $\delta J(y;v) = \int_0^5 2v'(x)y'(x)dx$.

The observant student may notice that the integrand in each case looks very similar to a derivative. In fact, we note that if $\frac{\partial}{\partial\epsilon} J(y+\epsilon v)$ exists for all sufficiently small $\epsilon > 0$ and is continuous at $\epsilon = 0$ then

$$\begin{aligned}
\lim_{\epsilon\to 0} \frac{J(y+\epsilon v) - J(y)}{\epsilon} &= \frac{\partial}{\partial\epsilon} J(y+\epsilon v)\Big|_0 \\
\lim_{\epsilon\to 0} \frac{f(\epsilon) - f(0)}{\epsilon} &= f'(0) \qquad \text{where} f(\epsilon) = J(y+\epsilon v)
\end{aligned}$$

Examples:

1. Note that taking our previous examples into consideration with this new result yields

$$\begin{aligned}
f'(\epsilon) = \frac{\partial}{\partial\epsilon} J(y+\epsilon v) &= \frac{\partial}{\partial\epsilon} \int_0^5 1 + (y' + \epsilon v')^2 dx \\
&= \int_0^5 \frac{\partial}{\partial\epsilon} (y' + \epsilon v')^2 dx \\
&= \int_0^5 2(y' + \epsilon v')v' dx
\end{aligned}$$

Notice that $f'(\epsilon)$ is a continous function of ϵ so that $\delta J(y;v)$ exists and

$$\delta J(y;v) = \frac{\partial}{\partial \epsilon} J(y+\epsilon v)\bigg|_{\epsilon=0} = \int_0^5 2y'(x)v'(x)dx$$

2. Consider the function $J(y) = \int_a^b \rho(x)\sqrt{1+y'(x)^2}dx$ for $\rho(x) \in C[a,b]$. Based off of our previous work, we can deduce the directional derivative as follows

$$\begin{aligned}
\delta J(y;v) = \frac{\partial}{\partial \epsilon} J(y+\epsilon v)\bigg|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon}\int_a^b \rho\sqrt{1+(y'+\epsilon v')^2}dx \\
&= \int_a^b \rho \frac{\partial}{\partial \epsilon}\sqrt{1+(y'+\epsilon v')^2}dx \\
&= \int_a^b \rho\left(\frac{1}{2}\left(1+(y'+\epsilon v')^2\right)^{-\frac{1}{2}}\left(2(y'+\epsilon v')v'\right)\right)dx \\
&= \int_a^b \rho\frac{y'v'}{\sqrt{1+y'^2}}dx \qquad \text{by continuity}
\end{aligned}$$

3. In the most general case, consider $J(y) = \int_a^b f(x,y(x),y'(x))dx$. Due to the presence of the first derivative as a parameter for f, we allow our domain to be $J : C^1[a,b] \to \mathbb{R}$. Furthermore, let us assume that f has continuous partial derivatives so that it is differentiable. Then we can easily see the following:

$$\begin{aligned}
J(y+\epsilon v) &= \int_a^b f(x, y+\epsilon v, y'+\epsilon v')dx \\
\frac{\partial J}{\partial \epsilon}(y+\epsilon v) &= \int_a^b \frac{\partial}{\partial \epsilon} f(x, y+\epsilon v, y'+\epsilon v')dx \\
&= \int_a^b f_x(x, y+\epsilon v, y'+\epsilon v')\cdot 0 \\
&\quad + f_y(x, y+\epsilon v, y'+\epsilon v')\cdot v \\
&\quad + f_z(x, y+\epsilon v, y'+\epsilon v')\cdot v' dx \qquad \text{by differentiabilty} \\
J(y+\epsilon v)\bigg|_{\epsilon=0} &= \int_a^b f_y(x,y,y')v + f_z(x,y,y')v'dx
\end{aligned}$$

Theorem 1.17. *Consider $J : C^1[a,b] \to \mathbb{R}$ where $J(y) = \int_a^b f(x,y(x),y'(x))dx$, and f has continuous partial derivatives. The Gateaux derivative is defined for all $y, v \in C^1[a,b]$ and*

$$\delta J(y;v) = \int_a^b f_y(x,y(x),y'(x))v(x) + f_z(x,y(x),y'(x))v'(x)dx$$

1.6 Formal Problem Statements

1.6.1 Geodesics on a Sphere:

Consider the 2-dimensional sphere (embedded in $\mathbb{R}^3$) of radius R. We can parameterize this space as

$$S^2 = (R\cos\theta\sin\phi, R\sin\theta\sin\phi, R\cos\phi) \qquad \text{for } \theta \in [0, 2\pi), \phi \in [0, \pi)$$

Points on the sphere are identified by (θ, ϕ). Note that this is a much simpler coordinate system than classical Cartesian coordinates, as $\sqrt{x^2 + y^2 + z^2} = R$. Without loss of generality, we can always either define our coordinate system (or subject our sphere to a isometric rotation) such that we can always take $\phi = 0$. That is, consider the points $A : (0, \theta_B), B : (\phi_B, \theta_B)$, so that the curve joining A to B can be represented by $(\phi(t), \theta(t)), t \in [0, 1]$, with initial conditions give by

$$\begin{array}{ll} \phi(0) = 0 & \phi(1) = \phi_B \\ \theta(0) = \theta_B & \theta(1) = \theta_B \end{array}$$

Furthermore, we shall assume $\phi, \theta \in C^1[0, 1]$ so that the curve is smooth. Now we recall that the length of curve $Y(t)$ is given by

$$\begin{aligned} S &= \int_0^1 |Y'(t)|dt, & |Y'(t)| &= \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} \\ &= R\int_0^1 \sqrt{\sin^2(\phi(t))\theta'(t)^2 + \phi'(t)^2}dt \end{aligned}$$

Define $J(\theta, \phi) = R\int_0^1 \sqrt{\sin^2(\phi(t))\theta'(t)^2 + \phi'(t)^2}dt$, and our goal will be to consequently minimize J. Notice that $J : C^1[0, 1] \times C^1[0, 1] \to \mathbb{R}$. Then our goal is to minimize $J(\theta, \phi)$ on the domain

$$\mathcal{D} = \left\{ (\theta, \phi) \in Y \;\middle|\; \begin{array}{ll} \phi(0) = 0 & \phi(1) = \phi_B \\ \theta(0) = \theta_B & \theta(1) = \theta_B \end{array} \right\}$$

However, we note that $\mathcal{D}$ is not a linear space, as it is not closed under pointwise addition. To resolve this issue, we will minimize J over all of θ such that $\theta'(t) = 0$ so that $\theta(t) = \theta_B, \forall t$. Then

$$J = R\int_0^1 |\phi'(t)|dt$$

We can always assume that $\phi'(t) > 0$, as this is again simply a coordinate orientation. Thus

$$\begin{aligned} J &= R\int_0^1 \phi'(t)dt \\ &= R[\phi(1) - \phi(0)] \\ &= R\phi_B \end{aligned}$$

Thus our solution is the curve $\theta = \theta_B, \phi = \phi(t)$ such that $\phi(0) = 0, \phi(1) = \phi_B$. The curves are all segments of great circles.

1.6.2 Brachistochrone

The Brachistochrone problem is the task of minimizing the travel time of a bead on a wire. Without loss of generality, let us coordinatize our endpoints as $(0,0)$ and (a,b). Let s be the distance of the curve, $v = \frac{ds}{dt}$, so that total time is

$$T = \int_0^L \frac{ds}{v} \tag{1}$$

Now we recall that the arclength of the curve is given by

$$s(x) = \int_0^x \sqrt{1 + y'(r)^2} dr \tag{2}$$

so that $\frac{s}{x} = \sqrt{1 + y'(x)^2}$. Using some fundamental physics, we can easily see that $\dot{v} = g\cos\alpha$ and that $\dot{y} = v\cos\alpha$. Substituting these two equations tells us that $\dot{v}v = g\dot{y}$. We recognize that the left hand size can be written as $\frac{1}{2}\frac{d}{dt}v^2 = \dot{v}v$, hence we can integrate the entire equation to get

$$\int \dot{v}v dt = \int g\dot{y} dt \Rightarrow \frac{1}{2}v(t)^2 = gy(t) + C$$

Now since our bead starts at rest, we have $v = 0$ when $y = 0$, hence $C = 0$. Thus $v(t) = \sqrt{2gy(t)}$. We can reject the negative root of why as we defined our coordinate system such that the bead falls under a positive force g in the positive y direction. We can substitute this as well as (2) into (1), to get that

$$T(y) = \int_0^a \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gy(x)}} dx$$

Let us take $T : C^1[0,a]\backslash_0 \to \mathbb{R}$. Thus our task is to minimize $T(y)$ in $\mathcal{D}$ where

$$\mathcal{D} = \left\{ y \in C^1[0,a] \;\middle|\; T(y) \text{ defined}, y(0) = 0, y(a) = b \right\}$$

Thus we can calculate $T(y)$ as follows

$$T(y) = \int_0^a \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gx}} dx$$
$$|T(y)| \leq \max \sqrt{1 + y'^2} \int_0^a \frac{1}{\sqrt{2gx}} \leq M\sqrt{\frac{2}{g}}a < \infty$$

Thus we want to find $\min_{y \in \mathcal{D}} T(y)$ where $\mathcal{D} = \{y \in C^1[0,a], y(0) = 0, y(a) = b\}$.

1.7 Convex Functional

Up until this point, we have been primarily considering the convexity on a functions between real spaces. However, we notice that our cost functional has $\mathcal{D} \subseteq C^1(\mathbb{R})$ as a domain, and hence requires some more

special attention. In particular, it would be useful if we could characterize the convexity of a integral-based cost functional in terms of the convexity of the integrand.

Recall that a differentiable function $J : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$J(y + v) \geq J(y) + \delta J(y; v)$$

The function is said to be strictly convex if the inequality holds strictly. This then leads us into the following definition:

Definition 1.18. A functional $J : Y \to R$ defined on a set $\mathcal{D} \subseteq Y$ is [strictly] **convex** on $\mathcal{D}$ if, whenever $y, y + v \in \mathcal{D}$, we have

1. $\delta J(y, v)$ is defined
2. $J(y + v) \geq J(y) + \delta J(y; v)$ (where for strict convexity, the inequality is taken to be strict

Proposition 1.19. *Assume that $J : Y \to R$ is [strictly] convex on $\mathcal{D} \subset Y$. If $y_0 \in \mathcal{D}$ satisfies $\delta J(y; v) = 0$ for all v such that $y_0 + v \in \mathcal{D}$ then y_0 minimizes J on $\mathcal{D}$ [uniquely].*

Proof. Let $v = y - y_0$ for any other $y \in \mathcal{D}$. Then

$$\begin{aligned} J(y) - J(y_0) &= J(y_0 + v) - J(y_0) \\ &\geq \delta J(y; v) = 0 \end{aligned}$$

with strict convexity applying if $y = y_0$. □

Recall from before the following theorem:

Theorem 1.20. *Consider $J : C^1[a, b] \to \mathbb{R}$ where $J(y) = \int_a^b f(x, y(x), y'(x))dx$ and f has continuous partial derivatives $\forall y, v \in C^1[a, b]$. Then*

$$\delta J(y; v) = \int_a^b f_y[y]v(x) + f_z[y]v'(x)dx$$

where $f[y] = f(x, y(x), y'(x))$

Example Let $J(y) = \int_1^2 \frac{y'(x)^2}{x} dx$. Then we consider the set of admissable functions to be $J : C^1[a, b] \to \mathbb{R}$ and can calculate the Gateaux derivative as follows:

$$\delta J(y; v) = \int_1^2 \frac{2y'(x)}{x} v'(x)dx$$

Our goal is then to minimize $J(y)$ over $\mathcal{D} = \left\{ y \in C^1[1, 2] \,\middle|\, y(1) = 0, y(2) = 3 \right\}$. If $y \in \mathcal{D}$, then $y + v \in \mathcal{D}$ if and only if $v \in \mathcal{A}$ where $A = \left\{ v \in C^1[1, 2] | v(1) = 0, v(2) = 0 \right\}$. In order to exploit our previous proposition, we must show that J is convex on $\mathcal{D}$.

$$\begin{aligned} J(y + v) - J(y) - \delta J(y; v) &= \int_1^2 \frac{(y' + v')^2 - y'^2 - 2y'v'}{x} dx \\ &= \int_1^2 \frac{v'(x)^2}{x} dx \\ &\geq 0 \end{aligned}$$

Hence J is convex on $\mathcal{D}$. Equality only holds if $v'(x) \equiv 0$. Furthermore, this implies that $v(x)$ is constant, which tells use that $v(x) \equiv 0$ by the endpoint conditions. However, notice that J is not *strictly* convex over all of $C^1[1,2]$. This is because we no longer have to impose the restriction given by $\mathcal{A}$ allowing v to be a non-zero function. Now if $\delta J(y;v) = 0$ for all v satisfying $y + v \in \mathcal{D} \subset Y$, then we say that $y \in \mathcal{A}$ is a stationary point (function) of J on $\mathcal{D}$.

We say that y_0 minimize J on $\mathcal{D}$ if the following hold:

1. J is [strictly] convex on $\mathcal{D} \subset Y$
2. $y_0 \in \mathcal{D}$ satisfies $\delta J(y_0; v) = 0$ and v is such that $y_0 + v \in \mathcal{D}$

Theorem 1.21. *Let $D \subset \mathbb{R}^2$ and for given y_a, y_b define*

$$\mathcal{D} = \left\{ y \in C^1[a,b] \;\middle|\; y(a) = y_a, y(b) = y_b, (y(x), y'(x)) \in D \right\}$$

and $J(y) = \int_a^b f[y]dx$ for f having continuous partial derivatives on $[a,b] \times D$. If J is [strictly] convex on $\mathcal{D}$ and y solves

$$\frac{d}{dx} f_z[y(x)] = f_y[y(x)]$$

then y minimizes J on $\mathcal{D}$ [uniquely].

Proof. Any $y \in \mathcal{D}$ for which $\delta J(y;v) = 0$ for all v satisfying $y + v \in \mathcal{D}$ minimizes J on $\mathcal{D}$ [uniquely].

$$\begin{aligned}
\delta J(y;v) &= \int_a^b f_y[y(x)]v(x) + f_z[y(x)]v'(x)dx \\
&= \int_a^b f_y[y(x)]v(x)dx + \cancel{f_z[y(x)]v(x)\Big|_{x=a}^{x=b}}^{0} - \int_a^b \frac{d}{dx}\left(f_z[y(x)]\right) v(x)dx && \text{by parts} \\
&= \int_a^b \left(f_y[y(x)] - \frac{\partial}{\partial x} f_z[y(x)] \right) v(x)dx && v(a) = 0 = v(b)
\end{aligned}$$

Thus if $y(x)$ solves the required differential equation, then $\delta J(y;v) = 0$ for all admissable v. □

Note: The differential equation that needs to be solved is called the **Euler-Lagrange** equation.

Example: (Continued)

We then have that $J(y) = \int_a^2 \frac{y'(x)^2}{x} dx, \mathcal{D} = \left\{ y \in C^1[1,2] | y(1) = 0, y(2) = 3 \right\}$ and J is strictly convex on $\mathcal{D}$. The Euler-Lagrange equation is then

$$\begin{aligned}
\frac{d}{dx} f_z[y(x)] &= 0 \\
f_z[y(x)] &= c \\
\frac{2y'}{x} &= c \\
y'(x) &= \frac{c}{2} x \\
y(x) &= \frac{c}{4} x^2 + k
\end{aligned}$$

Since $y(1) = 0, y(2) = 3$ then $y(x) = x^2 - 1$. This is the *unique* minimizing function for J on $\mathcal{D}$

We know now that if y solves the Euler-Lagrange equation ,then $\delta J(y; v) = 0$ for all admissable v. This is the only way for $\delta J(y; v) = 0, \forall v \in A$.

Lemma 1.22 (duBois-Raymond)**.** *If $h \in C[a, b]$ and $\int_a^b h(x)v'(x)dx = 0, \forall v \in \mathcal{A}$ where*

$$\mathcal{A} = \left\{ v \in C^1[a, b] \;\middle|\; v(a) = v(b) = 0 \right\}$$

then h is constant on all $[a, b]$ except possibly a zero measure set.

Proof. Define $c = \dfrac{1}{b-a} \displaystyle\int_a^b h(x)dx$ so that c is the average value of h over $[a, b]$. Then $\displaystyle\int_a^b \Big(h(x) - c\Big)\, dx = 0$ by the fundamental theorem of calculus. Now define $g(x) = \displaystyle\int_a^x \Big(h(t) - c\Big)\, dt$ and notice that $g \in C^1[a, b]$ since $h \in C[a, b]$. Furthermore, $g(a) = g(b) = 0$ so in particular, $g \in \mathcal{A}$. Thus by our original hypothesis we have that

$$\begin{aligned}
0 &= \int_a^b \Big(h(x) - c\Big)\, h(x)dx \\
&= \int_a^b \Big(h(x) - c\Big)\, h(x)dx + c\int_a^b \Big(h(x) - c\Big)\, dx \qquad\qquad \text{since } \int_a^b \Big(h(x) - c\Big)\, dx = 0 \\
&= \int_a^b (h(x) - c)^2\, dx \geq 0
\end{aligned}$$

The only way this is true is if the integrand is identically zero for all but a measure zero set, that is $h(x) - c = 0$. But then $h(x) = c$ which is precisely what we wanted to show. □

Proposition 1.23. *If $g, h \in C[a, b]$ and $\int_a^b g(x)v(x) + h(x)v'(x)dx = 0, \forall v \in \mathcal{A}$, then $h \in C^1[a, b]$ and $h'(x) = g(x)$*

Proof. Define $G(x) = \int_a^x g(t)dt$, then

$$\begin{aligned}
\int_a^b g(x)v(x) + h(x)v'(x)dx &= \cancel{G(x)v(x)\Big|_a^b}^{\,0} + \int_a^b (h(x) - G(x))v'(x)dx \\
&= \int_a^b [h(x) - G(x)]v'(x)dx \\
&\Rightarrow h(x) - \int_a^x g(t)dt = \text{constant} \qquad\qquad \text{by duBois-Raymond} \\
\frac{d}{dx}h(x) &= g(x)
\end{aligned}$$

□

Note that this is precisely the Euler-Lagrange equation when we define $h(x)$ and $g(x)$ appropriately to the Gateaux derivative of J.

1.8 Brachistochrone

Recall that we can model the Brachistochrone problem as $T(y) = \int_0^a \frac{\sqrt{1+y'(x)^2}}{\sqrt{2gx}}dx$, where we notice that $T : [0, a] \to \mathbb{R}$. Define $\mathcal{D} = \left\{ y \in C^1[0, a] \;\middle|\; y(0) = 0, y(a) = b \right\}$. Let us begin by showing that T is convex. Then

$$\begin{aligned}
\delta T(y; v) &= \int_0^a f_z[y]v'(x)dx \\
&= \int_0^a \frac{y'(x)}{\sqrt{1+y'(x)^2}} \frac{v'(x)}{\sqrt{2gx}} dx \\
g(z) &= \sqrt{1+z^2} \\
g'(z) &= \frac{z}{\sqrt{1+z^2}} \\
g''(z) &= \frac{1}{(1+z^2)^{\frac{3}{2}}}
\end{aligned}$$

$$T(y+v) - T(y) - \delta T(y; v) = \int_0^a \frac{1}{\sqrt{2gx}} \left(g(y'+v') - g(y') - g'(y')v' \right) dx$$

By Taylor's Remainder Theorem, for each $y'(x), v'(x)$ there exists c such that

$$g(y'(x) + v'(x)) = g(y'(x)) + g'(y'(x))v'(x) + \frac{1}{2}g''(c)v'(x)^2$$

$$\begin{aligned}
T(y+v) - T(y) - \delta T(y; v) = &= \int_0^a \frac{1}{\sqrt{2gx}} \frac{1}{2} g''(c)v'(x)^2 dx \\
&\geq 0 \quad \text{with equality iff } v'(x) = 0 \Leftrightarrow v(x) \equiv 0
\end{aligned}$$

Hence we can conclude that T is strictly convex. Thus the unique, minimizing solution function will be given by the solution to the Euler-Lagrange equation

$$\frac{d}{dx} f_z[y] - f_y[y] = 0$$

Since f varies independent of y, we can see that this simplifies to $f_z[y] = \text{constant}$. Hence $\frac{y'}{\sqrt{x}\sqrt{1+y'^2}} = \text{constant}$. Now if this constant is equal to zero, this implies that y is also constant. In this case, our point lies at $(a, 0)$ and the solution is the vertical line, so that the bead drops straight downward. Otherwise

$$\begin{aligned}
\frac{y'(x)}{\sqrt{x}\sqrt{1+y'(x)^2}} &= \frac{1}{c} \\
\frac{y'(x)^2}{1+y'(x)^2} &= \frac{x}{c^2} \\
y'(x) &= \sqrt{\frac{x}{c^2 - x}}
\end{aligned}$$

Using a change of variables, $x(\theta) = \frac{c^2}{2}(1-\cos\theta) = c^2\sin\left(\frac{\theta}{2}\right)$. Then

$$\frac{dy}{d\theta} = \frac{dy}{dx}\frac{dx}{d\theta}$$
$$c^2 - x = c^2 - \frac{c^2}{2}(1-\cos\theta) = \frac{c^2}{2}(1+cos\theta) \to \frac{dx}{d\theta} = \frac{c^2}{2}\sin\theta$$

So from our result above

$$\begin{aligned}
\frac{dy}{dx} &= \sqrt{\frac{1-\cos\theta}{1+\cos\theta}} \\
\frac{dy}{d\theta} &= \sqrt{\frac{1-\cos\theta}{1+\cos\theta}}\frac{c^2\sin\theta}{2} \\
&= \frac{c^2}{2}\sqrt{\frac{(1-\cos\theta)(1-\cos^2\theta)}{1+\cos\theta}} \\
&= \frac{c^2}{2}(1-\cos\theta) \\
y(\theta) &= c^2(\theta-\sin\theta), \quad y(0) = 0 \\
x(\theta) &= \frac{c^2}{2}(1-\cos\theta)
\end{aligned}$$

From section 8.8, we have the following calculation. Let $y = \frac{\tilde{y}(x)}{2}$, then

$$T(y) = \frac{1}{\sqrt{2g}}\int_0^b \frac{\sqrt{1+y'^2}}{\sqrt{y}}dx$$
$$\tilde{T}(\tilde{y}) = \frac{1}{\sqrt{g}}\int_0^b \sqrt{\tilde{y}^{-2}+\tilde{y}'^2}dx$$

Now $\tilde{T}$ is strictly convex, so let $y_0(x)$ be the cycloid, then $\tilde{y}_0(x) = \sqrt{2y_0(x)}$ is stationary for $\tilde{T}$.

1.9 Determining Convexity

Recall that we've been using the shorthand notation $f[y] = f(x, y(x), y'(x))$ so that we can write our cost integrals compactly as $J(y) = \int_a^b f[y]dx$ for $J : C^1[a,b] \to \mathbb{R}$ on $\mathcal{D} = \left\{y \in C^1[a,b] \,\middle|\, y(a) = y_a, y(b) = y_b\right\}$.

Definition 1.24. The function $f : \mathbb{R}^3 \to \mathbb{R}$ is [strongly] **pointwise convex** on $S \subset \mathbb{R}^3$ if f, f_y, f_z are continuous on S and

$$f(x, y+v, z+w) - f(x,y,z) - f_y(x,y,z)\cdot v - f_z(x,y,z) \geq 0$$

where the strong condition is satisfied if and only if $v = 0$ or $w = 0$.

Theorem 1.25. *Suppose that $J : C^1[a,b] \to \mathbb{R}$ is defined by $J(y) = \int_a^b f[y]dx$ for f [strongly] pointwise convex on $[a,b] \times D, D \subset \mathbb{R}^2$. Then J is [strictly] convex on any set $\mathcal{D}$, where*

$$\mathcal{D} = \left\{y \in C^1[a,b] | y(a) = y_a, y(b) = y_b, (y(x), y'(x)) \in D\right\}$$

Proof.

$$J(y+v) - J(y) - \delta J(y+v) = \int_a^b f(x, y+v, y'+v') - f(x,y,y') - f_y(x,y,y')v - f_z(x,y,y')v'dx$$
$$\geq 0 \quad \text{by p.w. convexity of } f$$

Thus J is a convex function on $\mathcal{D}$. If f is strongly pointwise convex, then equality will hold if and only if $v = 0$ or $v' = 0$; that is, $v(x)v'(x) = \frac{1}{2}\frac{d}{dx}(v(x)^2) = \vec{0}$, or $v(x) =$ constant. However, for $y, y+v \in \mathcal{D}$, this will only hold if $v = \vec{0}$. Thus J is strictly convex. □

It is important to note that f might be non-convex as a function but still pointwise convex. For example, consider $f(x,y,z) = -x^2 + y^2 + z^2$. f is not convex in $\mathbb{R}^3$ but is strongly pointwise convex. Furthermore, J can be convex even if f is not pointwise convex. For example, consider

$$J(y) = \int_0^1 y(x)^2 y'(x)dx, \qquad \mathcal{D} = \left\{ y \in C^1[0,1] \,\middle|\, y(1) = 2, y(0) = 0 \right\}$$

Then $f(x,y,z) = y^2y'$. Now

$$f(x, y+v, z+w) - f(x,y,z) - f_y(x,y,z)v - f_z(x,yz)w = v^2 z + ((v+y)^2 - y^2)w$$

This is not a convex function. However

$$J(y) = \int_0^1 \frac{1}{3}\frac{d}{dx}y^3(x)dx = \frac{1}{3}\left(y(1)^3 - y(0)^3\right) = 8/3$$

Since $J(y)$ is constant, it is necessarily convex.

Proposition 1.26. *If $f = f(x,z)$ and f, f_{zz} are continuous on $[a,b] \times I$ and for each $x \in [a,b]$ we have $f_{zz}(x,z) > 0$ (except possibly on a zero measure set), then $J = \int_a^b f(x, y'(x))dx$ is strictly convex on $\mathcal{D} = \left\{ y \in C^1[a,b] | y' \in I, y(a) = y_a, y(b) = y_b \right\}$*

Proof. We note that it is sufficent to show that f is pointwise convex. That is,

$$f(x, z+w) - f(x,z) - f_z(x,z)w \geq 0$$

Define $g(z) = f(x,z)$, so that by Taylor's Remainder Theorem, $g(z+w) = g(z) + g'(z)w + \frac{1}{2}g''(c_{z,w})w^2$ for some $c_{z,w}$. Thus if $g''(z) = f_{zz}(x,z) > 0$ we have

$$f(x, z+w) - f(x,z) - f_z(x,z)w = \frac{1}{2}f_{zz}(c)w^2 \geq 0$$

Hence f is pointwise convex and is equal to zero only if $w = 0$. Thus by our previous proposition, J is convex. Note that if $f_{zz} = 0$ for some points in x, this argument fails, but

$$J(y+v) - J(y) - \delta J(y; v) = \int_a^b \frac{1}{2}f_{zz}(x,c)v'(x)^2 dx \geq 0$$

This will be equal to 0 if $v'(x) = \vec{0}$ so $v(x) = \vec{0}$. □

Example: Show that the following cost function is convex on the given domain, then find a minimizing function.

$$J(y) = \int_0^{\frac{\pi}{4}} y'(x)^4 \sec^3(x) dx, \qquad \mathcal{D} = \left\{ y \in C^1\left[0, \frac{\pi}{4}\right] \,\middle|\, y(0) = 0, y\left(\frac{\pi}{4}\right) = 1 \right\}$$

We note that f has no dependency on y, so we impose our previous proposition. We can compute the derivatives of f as follows

$$\begin{aligned} f(x,z) &= z^4 \sec^3(x) \\ f_z(x,z) &= 4z^3 \sec^3(x) \\ f_{zz} &= 12z^2 \sec^3(x) \end{aligned}$$

We note that $f_{zz} > 0$ for $x \in [0, \frac{\pi}{4}]$, hence there is only one point on which $f_{zz} = 0$ corresponding to $z = 0$. We can conclude that J is strictly convex on $\mathcal{D}$. To find the minimizing function, we then solve the Euler-Lagrange equations

$$\begin{aligned} \frac{d}{dx} f_z[y] - f_y[y] &= 0 \\ f_z[y] &= 4c^3 \\ \frac{4y'(x)^3}{\cos^3(x)} &= 4c^3 \\ y'(x)^3 &= c^3 \cos^3(x) \\ y(x) &= c\sin(x) + d \end{aligned}$$

We can solve for these constants by applying the boundary conditions, to find that $y(x) = \sqrt{2}\sin x$, the unique minimizing function on $\mathcal{D}$.

1.10 Simplifications

Our goal is to minimize the cost function $J(y) = \int_a^b f[y(x)]dx$ on the set of admissable functions $D = \{y \in C^1[a,b], y(a) = y_a, y(b) = y_b\}$. Up to this point, we have seen that if J is convex, we can utilize the Euler-Lagrange equation

$$\frac{d}{dx} f_z[y(x)] = f_y[y(x)]$$

to find a minimizing solution. However, the Euler-Lagrange equation differentiates f with respect to the variables x, y, y'. In any instance where our function f does not explicitly depend on one of these three quantities, the EL equations then become greatly simplified. Let us consider each of these cases individually.

Case 1 Suppose firstly that $f[y]$ has no dependence on $y(x)$; that is, we can write $f(x, y'(x))$. In this case we notice that $f_y[y(x)] = 0$ and so the EL equation becomes

$$\frac{d}{dx} f_z[y(x)] = 0 \quad \Rightarrow f_z[y(x)] = \text{constant} \tag{3}$$

For example, let $f(x, y'(x)) = xy'(x)^2$. Then $f_z(x, y(x)) = \frac{d}{dy'}(xy'^2) = 2xy'$. Then by (3) we have that $2xy'(x) = c$ for some constant c, which we can solve for $y(x)$ as follows

$$\begin{aligned} 2xy'(x) &= c \\ y'(x) &= \frac{c}{2x} \\ y(x) &= \frac{c}{2}\log(x) + d \end{aligned}$$

where the constants c, d are to be determined by the boundary conditions imposed by $\mathcal{D}$.

Case 2 Let's assume that $f[y]$ does not explicitly depend on $y'(x)$. This is a very rare case (as it almost trivializes the problem) but we will consider it nonetheless. In this case we have that $f_z[y(x)] = 0$, causing the Euler-Lagrange equation to become

$$f_y[y(x)] - \frac{d}{dx} f_z[y(x)] = f_y[y(x)] = 0 \tag{4}$$

For example, let $f(x, y(x)) = y^2 - y\sin x$. In this case, $f_y[y(x)] = 2y - \sin x$ and (4) allows us to very quickly solve for $y(x)$

$$\begin{aligned} f_y[y(x)] &= 0 \\ 2y - \sin x &= 0 \\ y &= \frac{\sin x}{2} \end{aligned}$$

Notice that since we assumed that f had no dependency on $y'(x)$, we did not need to solve a differential equation. Indeed, so long as f is not transcendental in $y(x)$, we should be able to easily find the minimizing function for $J(y) = \int_a^b f[y(x)]dx$. While we're at it, let's check the pointwise convexity of f. We find that

$$\begin{aligned} f(x, y+v) - f(x, y) - f_y(x, y)\cdot v &= (y+v)^2 - \sin x(y+v) - y^2 + y\sin x - 2yv + v\sin x \\ &= v^2 \geq 0 \end{aligned}$$

Since equality holds if and only if $v \equiv 0$ we have that f is strongly pointwise convex. Hence we can conclude that J is strictly convex and that $y(x) = \frac{1}{2}\sin x$ is the unique minimizing function.

Case 3 Now assume that $f[y]$ has no dependency on x. If $y \in C^2[a, b]$ then

$$\frac{d}{dx} f(y(x), y'(x)) = f_y(y(x), y'(x))y'(x) + f_z(y(x), y'(x))y''(x)$$

Hence we can see that

$$\frac{d}{dx}\Big(f[y(x)] - y'(x)f_z[y(x)]\Big) = f_y[y]y' + f_z[y]y'' - y''f_z[y] - y'\frac{d}{dx}f_z[y] \tag{5}$$

$$= y'\underbrace{\left(f_y[y] - \frac{d}{dx}f_z[y]\right)}_{\text{the EL equation}} = 0 \tag{6}$$

$$\Rightarrow \qquad f[y(x)] - y'(x)f_z[y(x)] = \text{constant} \tag{7}$$

It can be shown that, conversely, if $y(x)$ is a function satisfying (7) on an interval where $y'(x) \neq 0$, then y satisfies the original Euler Lagrange equation.

For example, consider the function $f[y] = y(y')^2$. In this case we have that

$$f[y] - y'f_z[y] = y(y')^2 - y'2yy' = -y(y')^2$$

Hence we are left with the differential equation $y(y')^2 = c^2$ for some constant c. Assume that $y > 0, c > 0$ then

$$\begin{aligned} y(y')^2 = c^2 \quad &\Rightarrow \quad \frac{dy}{dx} = \frac{c}{\sqrt{y}} \\ \sqrt{y}dy &= cdx \\ \frac{2}{3}y^{3/2} &= cx + d \end{aligned}$$

where c, d are constants to be determined by the boundary conditions.

Note that this alternate formulation for finding a minimizing function can be quite useful. If we compare our previous work to the differential equation formed by the standard form of the Euler-Lagrange equation we get

$$(y')^2 - \frac{d}{dx}(2yy') = 0$$

1.11 Higher Order Derivatives

Conceivably, there may be times when the cost function is dependent on higher order derivatives, for example, taking

$$J(y) = \int_a^b f(x, y, y', y'')dx$$

And example of such a physical scenario is the potential energy of a beam under small deflections. In this case, we find that

$$U(y) = \int_0^L \mu y''(x)^2 - \rho(x)y(x)dx$$

In order to solve such minimizing problems, let us consider a similar technique to the derivation of the Euler-Lagrange equation.

We start by assuming that f has continuous partial derivatives so that

$$\delta J(y; v) = \int_a^b f_y[y]v(x) + f_z[y]v'(x) + f_w[y]v''(x)dx$$

Take $\mathcal{D} = \left\{ y \in C^2[a,b] \;\middle|\; y(a) = y'(a) = y(b) = y'(b) = 0 \right\}$. Integrating $\delta J(y;v)$ (by parts) then gives us

$$\begin{aligned}
\delta J(y;v) &= \int_a^b f_y[y]v dx + \cancelto{0}{f_z[y]v\Big|_a^b} - \int_a^b \frac{d}{dx}(f_z[y])v dx + \cancelto{0}{f_w[y]v'\Big|_a^b} - \int_a^b \frac{d}{dx} f_w[x]v' dx \\
&= \int_a^b f_y[y]v - \frac{d}{dx} f_z[y]v dx - \cancelto{0}{\frac{d}{dx} f_w[y]v\Big|_a^b} + \int_a^b \frac{d^2}{dx^2} f_w[y] y dx \\
&= \int_a^b \left[f_y[y] - \frac{d}{dx} f_z[y] + \frac{d^2}{dx^2} f_w[y] \right] v(x) dx
\end{aligned}$$

Thus we conclude that $y \in \mathcal{D}$ is a stationary function for J on $\mathcal{D}$ if and only if it solves the extended Euler-Lagrange equation

$$f_y[y] - \frac{d}{dx} f_z[y] + \frac{d^2}{dx^2} f_w[y] = 0 \tag{8}$$

Example: Consider the problem of a beam under a small deflection. We can express the potential energy of this system by

$$U(y) = \int_0^L \mu y''(x)^2 - \rho(x) y(x) dx$$

In the event that the beam is clamped, we have the terminal point conditions $y(0) = y'(0) = y(L) = y'(L) = 0$, over $\mathcal{D} = \left\{ y \in C^2[0,L] | \text{ end points satisfied } \right\}$. Define our integrand as $f(x,y,z,w) = \mu w^2 - \rho(x) y$. Our next step is to check the cost functional.

$$\begin{aligned}
U(y+v) - U(y) - \delta U(y;v) &= \int_0^L \mu(y'' + v'')^2 - \rho(x)(y+v) - \mu y''^2 + \rho(x) y + \rho(x) v - \mu^2 y'' v'' dx \\
&= \mu \int_0^L v''^2 dx \geq 0
\end{aligned}$$

We note that equality holds if and only if $v''(x) = 0, \forall x \in [0,L]$. However, the admissable variations on the clamped beam in turn implies that $v(x) \equiv 0, \forall x \in [0,L]$. We conclude that U is strictly convex. The Euler-Lagrange equation then becomes

$$\begin{aligned}
f_y - \frac{d}{dx} f_z + \frac{d^2}{dx^2} f_2 &= 0 \\
-\rho(x) + \frac{d^2}{dx^2}\left(2\mu y''\right) &= 0 \\
\frac{d^4 y}{dx^4} &= \frac{\rho(x)}{2\mu}
\end{aligned}$$

Suppose that the load on the beam is distributed uniformly such that $\rho(x) = 2\mu F$. Then our solution is a fourth order polynomial

$$y(x) = a + bx + cx^2 + dx^3 + \frac{F}{24} x^4$$

Imposing our endpoint conditions we can find that our unique solution is given by

$$y(x) = \frac{F}{24} x^2 (x - L)^2$$

1.12 Free Endpoint Problems

Physically, we may not always want to consider problems in which we have boundary conditions at the initial and terminal position. In such a case it seems reasonable to assume that we would still be able to minimize our cost function; however, we may need some additional analysis in order to uniquely specify minimizing solutions.

Consider the cost function with free terminal point

$$J(y) = \int_a^b f(x, y, y')dx, \qquad y(a) = y_a$$

We want to minimize $J(y)$ over the set $\mathcal{D} = \{y \in C^1[a,b] | y(a) = y_a\}$. In order to consider the general case, let us assume that f is pointwise convex. Then

$$\begin{aligned} J(y+v) - J(y) - \delta J(y; v) &= \int_a^v -f_y[y]v - f_z[y]v'dx \\ &\geq 0 \qquad \text{by pointwise convexity} \end{aligned}$$

If f is strongly pointwise convex, we have equality if and only if

$$v(x)v'(x) = 0, \qquad \forall x \in [a,b] \tag{9}$$

$$\frac{1}{2}\frac{d}{dx}v(x)^2 = 0 \tag{10}$$

$$v(x) = \text{constant} \tag{11}$$

For $y + v \in \mathcal{D}, v(a) = 0$ so $v(x) \equiv 0$. Hence J is strictly convex.

Now note however that we cannot directly apply the Euler-Lagrange equation. We've changed the assumptions that we used to derive the EL equations. Let us go back to our original derivation and see what modifications (if any) must be made.

$$\begin{aligned} \delta J(y; v) &= \int_a^b f_y[y]v + f_z[y]v'dx \\ &= \int_a^b f_y[y(x)]v(x)dx + f_z[y(x)]v(x)\Big|_{x=a}^b - \int_a^b \frac{d}{dx}f_z[y(x)]v(x)dx \\ &= \int_a^b \underbrace{\left(f_y[y(x)] - \frac{d}{dx}f_z[y(x)]\right)}_{\text{EL equation}} v(x)dx + f_z[y(b)]v(b) \end{aligned}$$

Recall that for y to be a stationary point, we require $\delta J(y; v) = 0, \forall v \in \mathcal{A}$. Where

$$\mathcal{A} = \{v \in C^1[a,b] | v(a) = 0\} \supset \{v \in C^1[a,b], v(a) = v(b) = 0\}$$

Thus y must satisfy the Euler-Lagrange equation. If so, then $\delta J(y; v) = f_z[y(b)]v(b)$. Thus in order for y to be a stationary point, we require that $f_z[y(b)] = 0$. Notice that by allowing one endpoint to be free, we "lost" a boundary condition to acquire a unique solution. However, in our analysis we have that

$$f_z[y(b)] = 0 \tag{12}$$

which is a new, natural boundary condition.

Theorem 1.27. *Let $\mathcal{D} \subset \mathbb{R}^2$. Assume that f is [strongly] pointwise on $[a,b] \times \mathcal{D}$. Each solution $y_0 \in D_1 = \left\{ y \in C^1[a,b] \,\middle|\, y(a) = y_a \right\}$ that satisfies*

$$f_y[y(x)] - \frac{d}{dx} f_z[y(x)] = 0$$

[uniquely] minimizes J on $\mathcal{D}$ if $f_z[y_0(b)] = 0$.

Example: Steady State Temperature in a Bar

Consider a isotropic, uniform density bar of length L subject to a thermal condition imposed at the endpoint as $y(0) = 100$. The steady state solution to this problem will be the one that minimizes the potential energy over the entire bar. We can write this as

$$U(y) = k \int_0^L y'(x)^2 dx, \qquad k > 0$$

Since $f(z) = kz^2$ is strongly pointwise convex, J is strictly convex. Hence the Euler-Lagrange equation yields

$$\begin{aligned} f_z[y(x)] &= c_1 \\ 2ky'(x) &= c_1 \\ y(x) &= cx + d \end{aligned}$$

The boundary conditions (both imposed and natural) are

$$y(0) = 100, \qquad f_z[y(L)] = 0 \Rightarrow 2ky'(L) = 0$$

Thus $y(x) = 100$ is the unique minimizing function.

Exercise: Define $J(y) = \displaystyle\int_1^2 \frac{(y'(x) - 5)^2}{2x} dx$. Minimize J subject to the endpoint condition $y(1) = 1$.

Now that we've considered the case wherein one endpoint is left free, what happens when we allow both endpoints to be free? This leads us to the following theorem:

Theorem 1.28. *Let $\mathcal{D} \subseteq \mathbb{R}^2$ and assume that f is [strongly] pointwise convex on $[a,b] \times \mathcal{D}$. Each solution $y_0 \in C^1[a,b]$ of the Euler-Lagrange equation*

$$f_y[y] - \frac{d}{dx} f_z[y] = 0$$

minimizes J [uniquely to within an additive constant] on $C^1[a,b]$ if $f_z[y(b)] = 0$ and $f_z[y(a)] = 0$

Proof. This result is easily seen from the derivation for Theorem 1.27. Note that if y is a stationary function, then it must satisfy

$$f_z[y(b)]v(b) - f_z[y(a)]v(a) = 0$$

Note that when f is strongly pointwise convex, $J(y+v) - J(y) - \delta J(y; v) \geq 0$, with equality if and only if $v(x) =$ constant by precisely the same analysis as before. Thus if y_0 minimizes J, so does $y_0 +$ constant. $\square$

Example: Consider the heat equation, wherein we want to minimize the potential energy

$$U(y) = k\int_0^L y'(x)^2 dx$$

Then the Euler-Lagrange equation yields $y(x) = cx + d$ and the natural boundary conditions imply that $y'(0) = y'(L) = 0$ so that $y(x) \equiv 0$.

1.13 Variable Endpoint Problems

Let us now consider the problem wherein we allow our endpoint(s) to vary according to some prescribed curves. Notice that this is a compromise between the free-endpoint problem and fixed boundary points - the minimizing functions cannot be arbitrary, but are not strictly constrained. Our problem set up is as follows:

$$J(y) = \int_a^b f[y(x)]dx, \qquad y(a) = y_a, y(b) \in \{\varphi(x)\}$$

where $C(x)$ is a restraining curve on the terminal endpoint. We want to choose $\bar{b}$ to be an upper bound on the endpoint. Thus we consider $J : C^1[a, \bar{b}] \to \mathbb{R}$. In the event that $b < \bar{b}$ we can extend $C^1[a, b]$ linearly to $C^1[a, \bar{b}]$. Our next goal will be to find the Gateaux derivative of J, from which discovering conditions on the integrand $f[y]$ should immediately follow. Consider $y \in [a, b], y + \epsilon v \in [a, b + \epsilon\Delta]$ where y extends to $[a, b + \epsilon\Delta]$ if necessary. For the following calculation we will need to make use of the Leibniz Integral Rule described as follows:

$$\frac{d}{d\epsilon}\int_a^b f(x, \epsilon)dx = \int_a^b f(x, \epsilon)dx + f(b, \epsilon)\frac{db}{d\epsilon} - f(a, \epsilon)\frac{da}{d\epsilon}$$

The variation is (v, Δ), so assuming that f is differentiable with continuous partial derivatives yields

$$\begin{aligned}
J(y + \epsilon v, b + \epsilon\Delta) &= \int_a^{b+\epsilon\Delta} f(x, y + \epsilon v, y' + \epsilon v')dx \\
\frac{\partial J}{\partial \epsilon} &= \Delta f(b + \epsilon\Delta, (y + \epsilon v)(b + \epsilon\Delta), (y' + \epsilon v')(b + \epsilon\Delta)) \\
&\quad + \int_a^{b+\epsilon\Delta} f_y(x, y + \epsilon v, y' + \epsilon v')v + f_z(x, y + \epsilon v, y' + \epsilon v')v'dx \\
\left.\frac{\partial J}{\partial \epsilon}\right|_{\epsilon=0} &= \Delta f[y(b)] + \int_a^b f_y[y(x)]v(x) + f_z[y(x)]v'(x)dx \\
&= \Delta f[y(b)] + f_z[y(b)]v(b) + \int_a^b \left(f_y[y(x)] - \frac{d}{dx}f_z[y(x)]\right)v(x)dx \\
&= \delta J(y; (v, \Delta)) = 0
\end{aligned}$$

For y to be a stationary point of J, we must have that $\delta J(y; (v, \Delta)) = 0 \forall v, \delta$. In particular, this must be satisfied $\forall v$ such that $v(b) = 0$ and $\Delta = 0$. It is easy to see then that for the integrand to be zero, y must satisfy the Euler-Lagrange equation. Furthermore, other variations would imply that

$$f_z[y(b)]v(b) + \Delta f[y(b)] = 0 \tag{13}$$

for all admissable v, Δ.

Special Cases:

1. Let $y(b) = y_b$ for b fixed. Then $v(b) = 0, \Delta = 0$ so (13) will then always be 0. This is what we expect when our terminal point is explicitly stated.

2. If $x = b$ at the end point, $y(b)$ is free, then $v(b)$ is free and $\Delta = 0$. Thus (13) implies that

$$f_z[y(b)] = 0$$

which is precisely what we expect in the event we allow $y(b)$ to be free.

More generally, our endpont lies on the curve $\varphi(x)$ in which case $\Delta, v(b)$ are not necessarily zero, but also not independent. Assume that φ is C^1. Then

$$\begin{aligned} \tilde{y} &= y + \epsilon v \\ \tilde{y}(b + \epsilon\Delta) &= \varphi(b + \epsilon\Delta) \\ &= \phi(b) + \phi'(b)\epsilon\Delta + O(\epsilon^2\Delta^2) \end{aligned}$$

Extend y to $b + \epsilon\Delta$ so that

$$\begin{aligned} y(b + \epsilon\Delta) &= y(b) + y'(b)\epsilon\Delta \\ &= \varphi(b) + y'(b)\epsilon\Delta \end{aligned}$$

Then since $\epsilon v = \tilde{y} - y$ we get that

$$\epsilon v(b + \epsilon\Delta) = \tilde{y}(b + \epsilon\Delta) - y(b + \epsilon\Delta) \tag{14}$$

$$= \Big(\varphi'(b) - y'(b)\Big)\,\epsilon\Delta + O(\epsilon^2\Delta^2) \tag{15}$$

$$v(b + \epsilon\Delta) = \Big(\phi'(b) - y'(b)\Big)\,\Delta + O(\epsilon^2\Delta^2) \tag{16}$$

$$v(b) = \Big(\phi'(b) - y'(b)\Big)\,\Delta \tag{17}$$

Substituting this last expression into (13) we get that

$$\Bigg(\underbrace{f_z[y(b)]\,\Big(\phi'(b) - y'(b)\Big) + f[y(b)]}_{=0 \text{ since } \Delta \text{ arbitrary}}\Bigg)\Delta = 0$$

We can rewrite this in a way that is easier to remember. Begin by defining $p(x) = f_z[y(x)]$. Then

$$H(x) = -f[y(x)] + y'(x) f_z[y(x)]$$

The condition given by (17) can thus be written as

$$-H(b) + p(b)\varphi'(b) = 0$$

and is known as the transversality condition.

In summary, note the following set of equations required to solve our minimization problem. Specifically, the minimizing solution will be y, b such that

$$\begin{gathered} f_y[y(x)] - \tfrac{d}{dx} f_z[y(x)] = 0 \\ y(a) = y_a \\ -H(b) + p(b)\phi'(b) = 0 \\ y(b) = \phi(b) \end{gathered}$$

Example: Brachistochrone

Recall that the Brachistochrone problem has minimizing functional given by

$$T(y) = \int_0^b \frac{1}{\sqrt{2gx}} \frac{\sqrt{1+y'^2}}{\sqrt{1+y'^2}} dx, \qquad y(0) = 0$$

The minimizing function solves the Euler-Lagrange equation. In this case the solution is given by the cycloid with $y(0) = 0$. At the terminal point, we have

$$\begin{aligned} p = f_z &= \frac{y'}{\sqrt{2gx}\sqrt{1+y'^2}} \\ H &= -f + y' f_z \end{aligned}$$

So the transversality condition is $p(b)\varphi'(b) - H(b) = 0$ so that if $\varphi'(b) =\neq 0$ then

$$\frac{1}{\sqrt{2gx}\sqrt{1+y'^2}} \left(y'(b)\varphi(b) + \right) = 0$$

So that $y'(b) = -\varphi''(b)^{-1}$. This is generally true for problems of the form

$$\int_a^b h(x,y)\sqrt{1+y'^2} dx$$

1.14 General Conditions for Broken Extremals

Example: Define $J(y) = \int_{-1}^1 y^2(x)(1-y'(x))^2 dx$ and minimize this functional over

$$\mathcal{D} = \left\{ y \in C^1[-1,1] \;\middle|\; y(-1) = 0, y(1) = 1 \right\}$$

We can do this by solving the Euler-Lagrange equation. Since there is no dependence on x we need only solve

$$\begin{aligned} \text{constant} = H &= -f + f_z y' \\ &= -y^2(1-y'(x))^2 - 2y^2 y'(1-y') \\ &= y^2(x)(1-y'(x)^2) \end{aligned}$$

Since $y(1) = 1$, our solution cannot be $y(x) \equiv 0$. Thus we consider the case when $y'(x) \equiv \pm 1$. However,

$$y'(x) = -1 \Rightarrow y(x) = -1 - x$$
$$y'(x) = 1 \Rightarrow y(x) = x + 1$$

These both contradict $y(1) = 1$, so there are no stationary functions in $\mathcal{D}$. Notice the function

$$y(x) = \begin{cases} 0 & \text{on } [-1, 0] \\ x & \text{on } [0, 1] \end{cases}$$

satisfies the end points, and but is not in $\mathcal{D}$ since it is not continuously differentiable, only continuous.

With this example in mind, define

$$\hat{C}^1[-1, 1] = \left\{ y \in C[-1, 1] \,\middle|\, y' \text{ exists and is discontinuous at finitely many points} \right\}$$

In general, consider $J(y) = \int_a^b f(x, y(x), y'(x))dx$ for $y(a) = y_a, y(b) = y_b$. Let us assume that y may have a corner point at $c \in [a, b]$. We can then split up our interval as

$$\begin{aligned} J(y) &= \int_a^c f(x, y(x), y'(x))dx + \int_c^b f(x, y(x), y'(x))dx \\ \delta J(y; v, \Delta) &= \int_a^c f_y[y(x)]v(x) + f_z[y(x)]v'(x)dx + f[y(c_-)]\Delta \\ &\quad + \int_c^b f_y[y(x)]v(x) + f_z[y(x)]v'(x)dx - f[y(c_+)]\Delta \\ &= \int_a^c \left(f_y[y(x)] - \frac{d}{dx} f_z[y(x)] \right) v(x)dx + f_z[y(c_-)]v(c_-) + f[y(c_-)]\Delta \\ &\quad + \int_c^b \left(f_y[y(x)] - \frac{d}{dx} f_z[y(x)] \right) v(x)dx - f_z[y(c_+)]v(c_+) - f[y(c_+)]\Delta \end{aligned}$$

In order to minimize $J(y)$, we must have that $\delta J(y; v, \Delta) = 0, \forall v \in \mathcal{D}$ the set of admissable variations. In particular, we want to consider when $v(c_+) = 0, \Delta = 0$. Thus the Euler-Lagrange equation must be satisfied on each interval $[a, c], [c, b]$. We can think of c as being an open endpoint, lying on a curve φ so that $v(c_-) = (\varphi'(c_-) - y'(c_-))\Delta$. This leads to

$$\Big(\underbrace{f_z[y(c_-)]}_{P} - f_z[y(c_+)] \Big) \varphi'(c) + \Big(H[y(c_-)] - H[y(c_+)] \Big) \Delta = 0$$

Since φ and Δ are arbitrary,

$$P[y(c_-)] = P[y(c_+)]$$
$$H[y(c_-)] = H[y(c_+)]$$

This is because P, H are continuous for a stationary curve, even at points of discontinuity. These are the *Weierstrass - Erdmann* corner conditions.

We want to find a condition that indicates whenever a corner is present. Define $g(z) = f_z(c, y(c), z)$. If $y'(c_+) \neq y'(c_-)$ but y is a stationary function so that

$$\begin{aligned} g(\underbrace{y'(c_+)}_{z_1}) &= g(\underbrace{y'(c_-)}_{z_2}) \\ g(z_1) &= g(z_2), \quad z_1 \neq z_2 \end{aligned}$$

Then if g' is define, the mean value theorem implies that $g'(z) = 0$. However, $g'(z) = f_{zz}(c, y(c), z)$. Thus if f_{zz} is defined at a corner c, $\exists z$ such that

$$f_{zz}(c, y'(c), z) = 0$$

Let us now go back and consider the example at the beginning of this section.

Example: (Continued) Recall that we have $f(x, y, z) = y^2(1 - z)^2$, so that

$$\begin{aligned} p &= f_z = -2y^2(1 - z) \\ f_{zz} &= 2y^2 \end{aligned}$$

Hence we can conclude that corners are *possible* whenever $y = 0$. Stationary functions will then solve the Euler-Lagrange equation on each sub-interval. Then

$$\begin{aligned} H &= \text{ constant} \\ &= -y^2(1 - y'^2) \end{aligned}$$

Since $y(-1) = 0$ then $c = 0$, and we have that $H = -y^2(1 - y'^2) = 0$. Now for all x we have the possible solutions $y = 0, y' = 1, y' = -1$. Now we want both H and $P = -2y^2(1 - y')$ to be continuous at our given points. Notice that if $y = 0$ then $y' = \pm 1$ at the corner satisfies the corner conditions. However, having $y' = -1, y' = 1$ leads to P being discontinuous. We reject $y' = -1$ since $y(1) = 1$ and hence our minimizing solution is precisely

$$y(x) = \begin{cases} 0 & x \in [-1, 0] \\ x & x \in [0, 1] \end{cases}$$

1.15 Minimum Surface of Revolution

OUr goal is to find a curve joining two points so that the area of the surface formed by rotating this curve about the x-axis is minimal. We can normalize this problem so that one point has coordinate $(0, 1)$. Let us denote the second point by (x_1, y_1). The surface area of the given surface is given by

$$J(y) = 2\pi \int_0^{x_1} y(x)\sqrt{1 + y'(x)^2}dx$$

and the set of admissable variations is given by

$$\mathcal{D} = \left\{ y \in C^1[0, x_1] \,\middle|\, y(0) = 1, y(x_1) = y_1 \right\}$$

Now the curve $f(y,z) = y\sqrt{1+z^2}$ is not convex, but we will attempt to find stationary functions regardless. Since the Lagrangian does not depend on x, we see that the Euler-Lagrange equation reduces as follows

$$\begin{aligned}\frac{d}{dx}\Big(f(x) - y'(x)f_z(x)\Big) &= f_y y' + f_z y'' - y''f_z - y'\frac{d}{dx}f_z \\ &= y'\Big[\underbrace{\quad f_y - \frac{d}{dx}f_z \quad}_{Euler-Lagrange}\Big]\end{aligned}$$

Thus we must have that $f - y'f_z = c$, which gives us that

$$\begin{aligned} f - y'f_z &= c \\ y\sqrt{1+y'^2} - \frac{y'2y}{\sqrt{1+y'^2}} &= c \\ \frac{y(1+y'^2) - y'^2 y}{\sqrt{1+y'^2}} &= c \\ \frac{y}{\sqrt{1+y'^2}} &= c \end{aligned}$$

If $c = 0$ then $y = 0$, but $y(0) = 1$ so we conclude that $c \neq 0$. Rearranging we get that

$$\begin{aligned} \frac{dy}{dx} &= \sqrt{\frac{y^2-c^2}{c^2}} \\ c\log\left(\frac{y+\sqrt{y^2-c^2}}{c}\right) &= x + d \\ e^{\frac{x+d}{c}} &= \frac{y+\sqrt{y^2-c^2}}{c} \\ \frac{1}{c}\left(y+\sqrt{y^2-c^2}\right)\frac{1}{c}\left(y-\sqrt{y^2-c^2}\right) &= \frac{1}{c^2}\left(y^2-(y^2-c^2)\right) = 1 \\ e^{-\frac{(x+d)}{c}} &= \frac{1}{c}\left(y-\sqrt{y^2-c^2}\right) \\ e^{\frac{x+d}{c}} + e^{-\frac{(x+d)}{c}} &= \frac{2y}{c} \\ \cosh\left(\frac{x+d}{c}\right) &= \frac{y}{c} \\ y &= c\cosh\left(\frac{x+d}{c}\right) \end{aligned}$$

Imposing the boundary conditions, $y(0) = 1 \Rightarrow c\cosh\left(\frac{d}{c}\right) = 1$. Define $\hat{d} = \frac{d}{c}$ so that $c = \frac{1}{\cosh(\hat{d})}$. Then our solution is

$$y(x) = \frac{\cosh(x\cosh(\hat{d}) + \hat{d})}{\cosh(\hat{d})}$$

The other boundary condition implies that we want to choose d such that $y(x_1) = y_1$, so we consider $y(x_1)$ as a function of d as d varies.

2 Hamilton's Principle

2.1 Introduction

Throughout the course thus far we have considered the problem of minimizing a functional through variational methods, making extensive use of the Euler-Lagrange equation. This is a successful technique in solving many problems such as the Brachistochrone, Bernoulli's Principle of Minimal Energy, and Snell's law. Foundations in these and the following techniques can be attributed to Lagrange, Euler, Poisson, and Hamilton. Let us begin by defining the action integral

$$A(y) = \int_a^b L(t, y, \dot{y})dt, \qquad L = T - U$$

where T is the kinetic energy and U is the potential energy.

Definition 2.1. A **potential** is a scalar quantity whose negative gradient gives a force acting on a body.

Let $\mathcal{D}$ be the set of all admissable trajectories for a system which have the prescribed end values.

Theorem 2.2 (Hamilton's Principle)**.** *Between fixed times a, b a system moves along the trajectory that makes stationary the action integral over all admissable trajectories.*

Example: Consider a cart of mass m subject to some potential energy $U(y)$ with a force $F = -U_y$. The kinetic energy is given by $\frac{1}{2}m\dot{y}^2$. Thus the Lagrangian and hence the action is given by

$$\begin{aligned} L(t, y, \dot{y}) &= \frac{1}{2}m\dot{y}^2 - U \\ A(y) &= \int_a^b \frac{1}{2}m\dot{y}^2 - U(y)dt \end{aligned}$$

Hamilton's Principle states that $\delta A(y; v) = 0$ for all admissable trajectories $y, y + v$ such that the Euler-Lagrange equation is satisfied.

$$\begin{aligned} L_y - \frac{d}{dt}L_{\dot{y}} &= 0 \\ -U_y - \frac{d}{dt}m\dot{y} &= 0 \\ m\ddot{y} &= -Uy = F \end{aligned}$$

Example: Consider the Spring-Mass problem in which $T = \frac{1}{2}m\dot{y}^2, U = \frac{1}{2}ky^2$. The action integral is

then given by $A(y) = \int_a^b \frac{1}{2}m\dot{y}^2 - \frac{1}{2}ky^2 dt$, and the Euler-Lagrange equation becomes

$$\begin{aligned}
\frac{d}{dt}L_{\dot{y}} - L_y &= 0 \\
\frac{d}{dt}m\dot{y} + ky &= 0 \\
m\ddot{y} + ky &= 0 \\
\ddot{y} &= -\frac{k}{m}y \\
y(t) &= A\sin\left(\sqrt{\frac{k}{m}}t\right) + B\cos\left(\sqrt{\frac{k}{m}}t\right)
\end{aligned}$$

However, since physically most instances of problems appear in more that one dimension, it will be of great use to review some notions for vector-valued functions. Consider

$$J(Y) = \int_a^b F(x, Y, Y')dx, \qquad J : (C^1[a,b])^n \to \mathbb{R}, \qquad Y = (y_1, \ldots, y_n)$$

Then we want to find an expression for $\delta J(Y, v)$ so that we can find stationary functions. Assume that f has continuous partial derivatives such that

$$\begin{aligned}
J(Y + \epsilon V) &= \int_a^b f(x, y_1 + \epsilon v_1, \ldots, y_n + \epsilon v_n, y_1' + \epsilon v_n', \ldots, y_n' + \epsilon v_n')dx \\
\frac{\partial}{\partial \epsilon}J(Y + \epsilon V) &= \int_a^b f_{y_1}[Y + \epsilon V]v_1 + \ldots + f_{y_n}[Y + \epsilon V]v_n + f_{y_1'}[Y + \epsilon V]v_1' + \ldots + f_{y_n'}[Y + \epsilon V]v_n' dx \\
\delta J(Y; V) &= \frac{\partial J}{\partial \epsilon}(Y + \epsilon V)\Big|_{\epsilon=0} = \int_a^b f_{y_1}[Y]v_1 + \ldots + f_{y_n}[Y]v_n + f_{y_1'}[Y]v_1' + \ldots + f_{y_n'}[Y]v_n' dx \\
&= \int_a^b \left(f_{y_1}[Y] - \frac{d}{dx}f_{y_1'}[Y]\right)v_1 + \ldots + \left(f_{y_n}[Y] - \frac{d}{dx}f_{y_n'}[Y]\right)v_n dx \\
&\quad + f_{y_1}[Y]v_1\Big|_a^b + \ldots + f_{y_n'}[Y]v_n\Big|_a^b
\end{aligned}$$

For $\delta J(Y; V) = 0$ when $V = (v_1, 0, \ldots 0)$ we must have that $f_{y_1} - \frac{d}{dx}f_{y_1'} = 0$. A similar result holds, so in general we have

$$f_{y_i} - \frac{d}{dx}f_{y_i'} = 0, \qquad i = 1, \ldots n$$

Similarily, define $H = -L + \sum_{i=1}^{n} \dot{y}_i L_{\dot{y}_i}$. If L does not depend on t then

$$\begin{aligned}
\frac{dH}{dt} &= -\sum_{i=1}^{n} L_{y_i}\dot{y}_i - \sum_{i=1}^{n} L_{\dot{y}_i}\ddot{y}_i + \sum_{i=1}^{n} \ddot{y}_i L_{\dot{y}_i} + \sum_{i=1}^{n} \dot{y}_i \frac{d}{dt} L_{\dot{y}} \\
&= \sum_{i=1}^{n} \dot{y}_i \Bigg(\underbrace{L_{y_i} - \frac{d}{dt} L_{\dot{y}_i}}_{\text{EL equation}} \Bigg) \\
&= 0
\end{aligned}$$

Thus in the multivariate case if the Lagrangian is independent of t we get that H is a constant.

Example: Consider a particle of mass m, moving in $\mathbb{R}^3$ with rectangular coordinates $(y_1(t), y_2(t), y_3(t))$. Assume that the particle has potential energy U. That is, there is $U(y_1, y_2, y_3)$ such that there is a force acting on the particle with components

$$F_1 = -U_{y_1}, \quad F_2 = -U_{y_2}, \quad F_3 = -U_{y_3}$$

The associated kinetic energy of the system is given by

$$T = \frac{1}{2} m \left(\dot{y}_1^2 + \dot{y}_2^2 + \dot{y}_3^2 \right)$$

Defining $L = T - V$, we can use Hamilton's Principle to find the equations of motion. To do this, we set $A(y) = \int_a^b L(y, \dot{y}) dt$ and find y such that $\delta A(y; v) = 0$ for all admissable variations. This will occur when the vector-valued Euler-Lagrange equations are satisfied.

$$\begin{aligned}
L_{y_1} - \frac{d}{dt} L_{\dot{y}_1} &= 0 \\
L_{y_2} - \frac{d}{dt} L_{\dot{y}_2} &= 0 \\
L_{y_3} - \frac{d}{dt} L_{\dot{y}_3} &= 0
\end{aligned}$$

Via the symmetry of our setup, we see that these equations reduce to

$$\begin{aligned}
-U_{y_i} - \frac{d}{dt}(m\dot{y}_i) = 0, & \qquad i = 1, 2, 3 \\
m\ddot{y}_i = -U_{y_i}, & \qquad i = 1, 2, 3
\end{aligned}$$

Example: Two Body Problem - Planet Around a Sun

Note that the potential energy of our system is given by $U = -\frac{k}{r}$ where r is the distance between the centre of mass of both bodies and k is a constant. The kinetic energy is

$$\begin{aligned}
T &= \frac{1}{2} m(\dot{x}^2 + \dot{y}^2) \\
&= \frac{1}{2} m(\dot{r}^2 + r^2 \dot{y}^2) \qquad \text{since } x = r\cos\theta, y = r\sin\theta
\end{aligned}$$

The Lagrangian is the difference between the kinetic and potential energy, $L(r,\theta) = \frac{1}{2}m(\dot{r}^2 - r^2\dot{y}^2) + \frac{k}{r}$. The motion of the planet is then a stationary function of the action

$$\begin{aligned} L_\theta - \frac{d}{dt}L_{\dot{\theta}} &= 0 \\ -\frac{d}{dt}(mr^2\dot{\theta}) &= 0 \\ L_r - \frac{d}{dt}L_{\dot{r}} &= 0 \\ mr\dot{\theta}^2 - \frac{k}{r^2} - m\ddot{r} &= 0 \end{aligned}$$

Note that from these equations we can derive all of Kepler's Laws as follows:

1. The angular momentum of the system is constant. That is $mr\dot{\theta} = \text{constant}$. If $r(\theta)$ is the path of a planet as a function of θ then

$$\begin{aligned} r^2(\theta)\dot{\theta} &= \frac{P}{m} \\ \int_{\theta_0}^{\theta_1} r^2 d\theta - \frac{P}{m}(t_1 - t_0) \end{aligned}$$

   We note that equal areas are swept out in equal times, since the arclength of the curve is $rd\theta$.

2. Since L does not directly depend on t we know that H is a constant, say E. Then

$$\begin{aligned} H &= E \\ &= -L + \dot{r}L_{\dot{r}} + \dot{\theta}L_{\dot{\theta}} \\ &= -\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{k}{r} + m\dot{r}^2 + mr^2\dot{\theta} \\ p &= mr^2\dot{\theta} \\ E &= \frac{1}{2}m\dot{r}^2 + \frac{1}{2}\frac{p^2}{mr^2} - \frac{k}{r} \\ \dot{r}^2 &= \frac{2}{m}\left(E - \frac{1}{2}\frac{p^2}{mr^2} + \frac{k}{r}\right) \end{aligned}$$

   Which is a separable first order equation. We can solve to find

$$r(\theta) = \frac{c}{1 + \epsilon\cos(\theta - \theta_0)}, \qquad \epsilon = \sqrt{1 + \frac{2Ep^2}{mk^2}}, \qquad c = \frac{p^2}{mk}$$

   One might recognize this as the equation for a conic section in polar coordinates.

$\epsilon = 0$	Circle
$0 < \epsilon < 1$	Ellipse
$\epsilon = 1$	Parabola
$\epsilon > 1$	Hyperbola

   For a closed orbit, $\epsilon < 1$. Kepler's first law is that planets have elliptical orbits.

3. Note that the area of an ellipse is πab where $a = \frac{c}{\sqrt{1-\epsilon^2}}, b = a\sqrt{1-\epsilon^2}$. So the area swept out by the orbit is

$$\begin{aligned} \frac{dA}{dt} &= \frac{1}{2}r^2\dot{\theta} = \frac{p}{2m} \\ \int_0^\tau \frac{dA}{dt}dt &= \frac{p}{2m}\tau \\ \pi ab &= \frac{p}{2m}\tau \\ a^3 &= \frac{p^2}{4m^2\pi^2 c}\tau^2 \end{aligned}$$

That is, the square of the period is proportional to the cube of the semi-major axis.

Example: Pendulum

Consider an ideal pendulum whose energy can be described as

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2), \qquad U = -mgy + \text{constant} = -mg(y - \ell)$$

The set of admissable trajectories are those starting at (x_0, y_0) and ending at (x_1, y_1), satisfying the constraint $x^2 + y^2 = \ell^2$. If we convert to polar coordinates, we get $r = \ell = \text{ constant }$ so that there is only one unconstrained variable on the system, θ. The energy of the system transforms as

$$T = \frac{1}{2}m\ell^2\dot{\theta}^2, \qquad U = mg\ell(1 - \cos\theta)$$

So that the associated Euler-Lagrange equation is given by

$$\begin{aligned} \frac{d}{dt}L_{\dot{\theta}} &= L_\theta \\ \frac{d}{dt}(m\ell^2\dot{\theta}) &= -mg\ell\sin\theta \\ \ell\ddot{\theta} &= -g\sin\theta \end{aligned}$$

The advantage of Hamilton's principle is that it allows us to easily model more complex system, with our knowledge based purely on the system energy. Keeping with what we did in the case of the pendulum example, we see that some choices of coordinate systems can result in complicated restraints on the system that we would like to ignore. Our goal will be instead to find a more generalized coordinate system that describes our process using n <u>independent</u> variables $q_1, \ldots, q_n$ which constitute a position vector $Q \in \mathbb{R}^n$.

Example: Double Pendulum on a Cart

Consider a double pendulum attached to the dorsal side of a car free to move on a one-dimensional track. We can completely specify the nature of the cart-pendulum system by considering the generalized coordinates consisting of two angles (corresponding to the cart-pendulum, pendulum-pendulum joints), and the position of the cart with respect to the track. Notice that all of three of these variables are mutually independent, and so the system does not have any interdependent constraints.

With this notion of generalized coordinates in mind, let us rephrase Hamilton's Principle to suit our new paradigm:

Hamitlon's Principle (Revised): Between fixed times $a \leq t \leq b$, a physical system should move along those trajectories represented by generalized coordinates $Q \in (C^1[a,b])^n$ with given values at $t = a$, and $t = b$, which make stationary the action integral

$$A(Q) = \int_a^b L(t, Q, \dot{Q})dt, \qquad L = T - U$$

Example:Pendulum on a Spring

Consider a mass m attached to a spring with spring-constant k allowed to hang freely in a pendulum-like manner. Allow the equilibrium position of the spring to be given by r_0. In this case, the generalized coordinates are given by the angle of the spring from the equilibrium, and the contraction/expansion of the spring from equilibrium. The system energy is given by

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2), \qquad U = \frac{1}{2}k(r - r_0)^2 + -mgy + \text{constant}$$

Allowing

$$\dot{x} = \frac{d}{dt}(r\sin\theta) = \dot{r}\sin\theta + r\dot{\theta}\cos\theta, \qquad \dot{y} = \frac{d}{dt}(r\cos\theta) = \dot{r}\cos\theta - r\dot{\theta}\sin\theta$$

we find that our transformed energy is given by

$$T = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2), \qquad U = \frac{1}{2}k(r - r_0)^2 + mg(r_0 - r\cos\theta)$$

Then the system Lagrangian is given by

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{1}{2}k(r - r_0)^2 - mg(r_0 - r\cos\theta)$$

The equations of motion are given by the Euler-Lagrange equation(s)

$$\begin{aligned}
L_r - \frac{d}{dt}L_{\dot{r}} &= 0 \\
m\dot{\theta}^2 r - k(r - r_0) + mg\cos\theta - \frac{d}{dt}(m\dot{r}) &= 0 \\
\ddot{r} &= \dot{\theta}^2 r - \frac{k}{m}(r - r_0) + g\cos\theta \\
L_\theta - \frac{d}{dt}L_{\dot{\theta}} &= 0 \\
-mgr\sin\theta - \frac{d}{dt}(mr^2\dot{\theta}) &= 0 \\
-mgr\sin\theta - 2mr\dot{r}\dot{\theta} - mr^2\ddot{\theta} &= 0 \\
2\dot{r}\dot{\theta} + r\ddot{\theta} &= -g\sin\theta
\end{aligned}$$

These are the equations of motion for the spring-pendulum system.

Example: Consider a mass m at (x, y) attached to a strictly horizontal string with spring constant k. Attached to the mass m is a pendulum with bob of mass M at (x_1, y_1). Our choice of generalized coordinates will be the deviation from the equilibrium position of the spring, x; as well as the angle θ of the pendulum with respect to the vertical. The system energy is given by

$$U(x,\theta) = \int_0^x k(s)ds + Mg\ell(1 - \cos\theta), \qquad T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{1}{2}M(\dot{x}_1^2 + \dot{y}_1^2)$$

2.2 Canonical Equations

We have seen that the Euler-Lagrange equation often yields a system of second-order, coupled differential equations. With very few exceptions, the corresponding equations will be difficult, if not impossible to solve analytically. For this reason we often resort to numerical methods. O common technique is to convert the second order equation into a system of first order differential equations. This can be done by defining $z_i = \dot{q}_i$ which creates an integration chain of first order DEs. However, there is another approach inspired by Hamilton's Principle. Define a new variable $p_i = L_{q_i}$, often referred to as the canonical (conjugate) momenta. This should be a co-oridinate transformation, and can be solved for $\dot{q}_i$ in terms of t, q_i, p_i. That is, the Jacobian of $F : \dot{q} \to p$ must be non-zero.

Case 1: Single Variate

Consider the case where the Lagrangian is given by

$$L(q) = a(q)\dot{q}^2 - u(T, Q)$$

It is easy to see that $p = L_{\dot{q}} = 2a(q)\dot{q}$ so that $\dot{q} = \frac{p}{2a(q)}$ This transformation is well defined so long as $a(q) \neq 0$. The Euler-Lagrange equation then becomes

$$\begin{aligned} L_q - \frac{d}{dt} L_{\dot{q}} &= 0 \\ \dot{p} &= L_q \end{aligned}$$

Thus we have the canonical equations of motion given by two partial differential equations

$$\dot{p} = L_q, \qquad \dot{q} = \frac{p}{2a(q)}$$

Case 2: Multivariate

The appropriate representation of the Lagrangian is given by

$$\begin{aligned} L &= \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} q \dot{q}_i \dot{q}_j}_{T} - \underbrace{U(t,q)}_{U} \\ &= \dot{q}^T A \dot{q} - U(t,q) \qquad \text{where } [A]_{ij} = a_{ij}(q) \\ p_j = L_{\dot{q}_j} &= \sum_{i=1}^{n} a_{ij}(q)\dot{q}_i \\ &= A^T \dot{q} \end{aligned}$$

In general, A is a positive definite matrix since $T = \dot{q}^T A \dot{q}$ and as such is invertible. We can then recover $\dot{q}$ from p. Let us denote this transformationn by G with components g_i. Then

$$\dot{q}_i = g_i(t, q, p), \quad i = 1, \ldots, n$$

The Euler-Lagrange equation can be written in terms of p as follows:

$$\begin{aligned} L_{q_i} - \frac{d}{dt} L_{\dot{q}_i} &= 0 \\ \dot{p}_i &= L_{q_i}(t, q, G(t, q, p)) \end{aligned}$$

In this scenario, we have the canonical equations of motion are given by

$$\dot{p}_i = L_{q_i}(t, q, G(t, q, p)), \qquad \dot{q}_i = g_i(t, q, p)$$

However, there is a much simpler way of writing these equations if we use the system Hamiltonian. Recall that

$$\begin{aligned}
H &= \sum_{i=1}^{n} L_{\dot{q}_j}\dot{q}_j - L \\
&= \sum_{j=1}^{n} p_j g_j(t, q, p) - L(t, q, G(t, q, p)) \\
\frac{\partial H}{\partial p_j} &= g_i(t, q, p) + \sum_{j=1}^{n} p_j \frac{\partial g_j}{\partial p_i} - \sum_{j=1}^{n} \frac{\partial L}{\partial \dot{q}_j}\frac{\partial g_j}{\partial p_i} \\
&= g_i(t, q, p) + \frac{\partial g_j}{\partial p_i}\sum_{j=1}^{n}\left(p_j - \cancelto{0}{\frac{\partial L}{\partial \dot{q}_j}}\right)\frac{\partial g_i}{\partial p_i} \\
&= \dot{q}_i
\end{aligned}$$

We can perform a similar calculation

$$\begin{aligned}
\frac{\partial H}{\partial q_i} &= \sum_{j=1}^{n} p_j \frac{\partial g_j}{\partial q_i} - \frac{\partial L}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial L}{\partial \dot{q}_j}\frac{\partial g_j}{\partial q_i} \\
&= -\frac{\partial L}{\partial q_i} + \sum_{j=1}^{n}\left(p_j - \cancelto{0}{\frac{\partial L}{\partial \dot{q}_j}}\right)\frac{\partial g_j}{\partial q_i} \\
\dot{p}_i &= -\frac{\partial H}{\partial q_i}
\end{aligned}$$

To summarize, express the system Hamiltonian in terms of p, q using the canonical momenta, and then find the equations of motion given by

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}$$

Example: Pendulum We recall that the system energy is given by

$$T = \frac{1}{2}m\ell^2\dot{\theta}^2, \qquad U = mg\ell(1 - \cos\theta) \qquad L = T - U$$

Defining $p = L_{\dot{\theta}} = m\ell^2\dot{\theta}$ we can solve for $\dot{\theta}$ to get that $\dot{\theta} = \frac{p}{m\ell^2}$. The Hamiltonian is given by

$$\begin{aligned}
H &= -L + L_{\dot{\theta}}\dot{\theta} \\
&= -T + U + p\left(\frac{p}{m\ell^2}\right) \\
&= \frac{1}{2}m\ell^2\left(\frac{p^2}{m^2\ell^4}\right) + mg\ell(1 - \cos\theta) + \frac{p^2}{m\ell^2} \\
&= \frac{p^2}{2m\ell^2} + mg\ell(1 - \cos\theta)
\end{aligned}$$

The canonical equations then follow easily

$$\dot{\theta} = \frac{\partial H}{\partial p} = \frac{p}{m\ell^2}$$
$$\dot{p} = -\frac{\partial H}{\partial \theta} = -\left(mg\ell \sin\theta\right)$$

Example: Orbital Motion

We found previously that the system energy is given by

$$T = \frac{1}{2}mv^2, \qquad U = -\frac{k}{r}$$

The Euler-Lagrange equations yield

$$mr\dot{\theta}^2 - m\ddot{r} = \frac{k}{r^2}$$
$$\frac{d}{dt}\left(mr^2\dot{\theta}\right) = 0$$

Converting to the canonical equation formalism we have the canonical momenta described as $p_1 = L_{\dot{r}} = m\dot{r}, p_2 = L_{\dot{\theta}} = mr^2\dot{\theta}$ so that $\dot{r} = \frac{p_1}{m}, \dot{\theta} = \frac{p_2}{mr^2}$. The Hamiltonian can be described as

$$\begin{aligned} H &= -L + \dot{r}p_1 + \dot{\theta}p_2 \\ &= -T + U + \dot{r}p_1 + \dot{\theta}p_2 \\ &= \frac{p_1^2}{2m} + \frac{p_2^2}{2mr^2} - \frac{k}{r} \\ \dot{r} &= \frac{\partial H}{\partial p_1} = \frac{p_1}{m} \\ \dot{\theta} &= \frac{\partial H}{\partial p_2} = \frac{p_2}{mr^2} \\ \dot{p}_1 &= -\frac{\partial H}{\partial r} = \frac{p_2^2}{mr^3} - \frac{k}{r^2} \\ \dot{p}_2 &= -\frac{\partial H}{\partial \theta} = 0 \end{aligned}$$

Example: Consider a mass m attached to a non-linear horizontal spring with spring constant $k(x)$. Also, let there be a bob of mass M attached to the mass m in a pendulum form. Define the generalized coordinate system to be x, the displacement of the mass from the equilibrium position; and θ, the angle of the pendulum from the horizontal position. Before converting to generalized coordinates, represent the Cartesian coordinates of the pendulum bob as (x_p, y_p). The kinetic and potential energy are given by

$$T = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}M(\dot{x}_p^2 + \dot{y}_p^2), U = Mg(L - y) + \int_0^x k(s)ds$$

Using the coordinate translation $x_p = L\sin\theta + x, y_p = L\cos\theta$ we get

$$T = \frac{1}{2}(m + M)\dot{x}^2 + \dot{x}ML\dot{\theta}\cos\theta + \frac{1}{2}ML^2\dot{\theta}^2, U = MgL(1 - \cos\theta) + \int_0^x k(s)ds$$

Now we find the canonical momenta,

$$\begin{aligned} p_x &= L_{\dot{x}} = (m+M)\dot{x} + ML\dot{\theta}\cos\theta \\ p_\theta &= L_{\dot{\theta}} = \dot{x}ML\cos\theta + ML^2\dot{\theta} \end{aligned}$$

we need to solve for $\dot{x}, \dot{\theta}$ in terms of p_x, p_θ, to find that

$$\begin{aligned} \begin{pmatrix} p_x \\ p_\theta \end{pmatrix} &= \underbrace{\begin{pmatrix} m+M & ML\cos\theta \\ ML\cos\theta & ML^2 \end{pmatrix}}_{A} \begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} \\ \det A &= (m+M)ML^2 - (ML\cos\theta)^2 = mML^2 + M^2L^2(1-\cos^2\theta) \\ &= mML^2 + M^2L^2\sin^2\theta > 0 \\ \begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} &= \frac{1}{\det A}\begin{pmatrix} ML^2 & -ML\cos\theta \\ -ML\cos\theta & m+M \end{pmatrix}\begin{pmatrix} p_x \\ p_\theta \end{pmatrix} \\ &= \frac{1}{\det A}\begin{pmatrix} ML^2p_x - ML\cos\theta\ p_\theta \\ -ML\cos\theta\ p_x + (m+M)p_\theta \end{pmatrix} \end{aligned}$$

Then the system Hamiltonian is given by

$$\begin{aligned} \mathcal{H} &= -L + L_{\dot{\theta}}\dot{\theta} + L_{\dot{x}}\dot{x} \\ &= -T + U + p_\theta\dot{\theta} + p_x\dot{x} \\ &= -\frac{1}{2}(m+M)\dot{x}^2 - \frac{1}{2}\dot{x}ML\dot{\theta}\cos\theta - \frac{1}{2}\dot{x}\dot{\theta}ML\cos\theta - \frac{1}{2}ML^2\dot{\theta}^2 \\ &\quad + \int_0^x k(s)dx + MgL(1-\cos\theta) + \dot{x}p_x + \dot{\theta}p_\theta \\ &= -\frac{1}{2}\left((m+M)\dot{x} + ML\dot{\theta}\cos\theta\right)\dot{x} - \frac{1}{2}\left(\dot{x}ML\cos\theta + ML^2\dot{\theta}\right)\dot{\theta} \\ &\quad + \int_0^x k(s)ds + MgL(1-\cos\theta) + \dot{x}p_x + \dot{\theta}p_\theta \\ &= -\frac{1}{2}p_x\dot{x} - \frac{1}{2}p_\theta\dot{\theta} + \int_0^x k(s)ds + MgL(1-\cos\theta) + \dot{x}p_x + \dot{\theta}p_\theta \\ &= \frac{1}{2}p_x\dot{x} + \frac{1}{2}p_\theta\dot{\theta} + \int_0^x k(s)ds + MgL(1-\cos\theta) \end{aligned}$$

The canonical equations of motion are then

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p_x} = \frac{1}{\det A}ML^2p_x - ML\cos\theta p_\theta \\ \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = -\frac{1}{\det A}ML\cos(\theta)p_x + (m+M)p_\theta \\ \dot{p}_x &= -\frac{\partial H}{\partial x} = -k(x) \\ \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = \text{ exercise} \end{aligned}$$

2.3 Spatially Distributed Problems

1. **Equilibrium Prolbems**

Let us now move to considering problems with multiple spatial dimensions. Consider the cost integral defined as

$$J(y) = \int_D f\left(\vec{x}, y(\vec{x}), \nabla y(\vec{x})\right) d\vec{x}, \qquad D \subseteq \mathbb{R}^n, y(x) = \gamma, x = \partial D, \gamma(x) \in C(\partial D)$$

Assume that $f \in C^1(\bar{D} \times \mathbb{R} \times \mathbb{R}^n)$. We want to find conditions for y to be a stationary function of J on the set

$$\mathcal{D} = \left\{y \in C^1(D), y(x) = \gamma(x), x \in \partial D\right\}$$

The set of admissable variations are

$$\mathcal{A} = \left\{v \in C^1(D) \;\middle|\; v(x) = 0, \forall x \in \partial D\right\}$$

In an effort to find the associated Gateaux derivative, we have

$$J(y + \epsilon v) = \int_D f\left(\vec{x}, y(\vec{x}) + \epsilon v(\vec{x}), \nabla y(\vec{x}) + \epsilon \nabla v(\vec{x})\right) d\vec{x}$$

Since f has continuous partial derivative, this functional is continuous in a neighbourhood of $\epsilon = 0$.

$$\delta J(y; v) = \int_D f_y[y(\vec{x})]v(\vec{x}) + f_{\nabla y}[y(\vec{x}) \cdot \nabla v(\vec{x}) d\vec{x}$$

where $f_{\nabla y}[y(\vec{x})] = \left(\dfrac{\partial f}{\partial z_1}, \dfrac{\partial f}{\partial z_2}, \dfrac{\partial f}{\partial z_3}\right)$. The multidimensional analogy to integration by parts is Green's Theorem, so we assum ethat D is a boundend Green's domain so that

$$\begin{aligned}
\delta J(y; v) &= \int_D f_y[y(\vec{x})]v(\vec{x})d\vec{x} + \int_{\partial D} v(\vec{x}) f_{\nabla y}[y(\vec{x})] \cdot \eta(\vec{x}) dx - \int_D v(x) \nabla \cdot f_{\nabla y}[y(\vec{x})] d\vec{x} \\
&= \int_D \left[f_y[y(\vec{x})] - \nabla \cdot f_{\nabla y}[y(\vec{x})]\right] v(\vec{x}) d\vec{x} - \int_{\partial D} \cancelto{0}{v(x)} f_{\nabla y}[y(\vec{x})] \cdot \eta(x) d\vec{x}
\end{aligned}$$

Since for y to be a stationary function of $J(y)$ we expect $\delta J(y; v) = 0$. Since $v(x)$ is arbitrary, it must follow that

$$f_y[y(\vec{x})] - \nabla \cdot f_{\nabla y}[y(\vec{x})] = 0 \tag{18}$$

This is a multi-dimensional partial differential equation that is the generalization of the Euler-Lagrange equation.

Example:

Consider general three-dimensional space. The potential energy can often be given as

$$U(y) = \int_D \frac{1}{2}\left(y_{x_1}^2 + y_{x_2}^2 + y_{x_3}^2\right) d\vec{x}, \qquad D \subseteq \mathbb{R}^3$$

In the case of heat diffusion, we get that $y(\vec{x})\Big|_{\partial D} = \gamma(x)$. To evaluate the PDE in (18), we can see that $f_y = 0$ and $f_{\nabla y} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial y}{\partial x_3}\right)$ to yield

$$\begin{aligned} -\nabla \cdot \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial y}{\partial x_3}\right) &= 0 \\ \frac{\partial^2 y}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2} + \frac{\partial^2 y}{\partial x_3^2} &= 0 \\ \nabla^2 y(\vec{x}) &= 0 \qquad \text{Laplace's Equation} \end{aligned}$$

It is relatively simple to extend the theoretical results of our previous derivations to multi-dimensional space. For example, in the event that $\gamma(x)$ is only specified on $S \subset \partial D$ then we impose the transversality condition that

$$f_{\nabla y}[y(\vec{x})] = 0, \qquad \forall x \in \partial D \backslash_S$$

2. **Non-Equilbrium Problems**

Some systems have motion that varies with a spatial co-ordinate as well as time. For example, vibrations through a medium will vary with both space and time. Consider a fixed length $x \in [0, L]$. Hamilton's Principle is very versatile, and should still apply in this situation as well. If the system energies are given by

$$T = \int_0^L \frac{1}{2}\rho y_t(x,t)^2 dx, \qquad U(x, y, y_x) = \int_0^L u(x, y, y_x) dx$$

Then the action integral is

$$\begin{aligned} A(y) &= \int_a^b T(y) - U(y) dt \\ &= \int_a^b \int_0^L \frac{1}{2}\rho y_t(x,t)^2 - U dx dt \end{aligned}$$

The motion will make $A(y)$ stationary over all admissable trajectories. Suppose the ends are fixed so that

$$y(0,t) = 0, y(L,t) = 0, \qquad b \le t \le b$$

Then our goal will be to minimize $A(y)$ over

$$\mathcal{D} = \left\{ y \in C^1\left([0,L] \times [a,b]\right) \;\middle|\; \begin{array}{ll} y(0,t) = 0, & y(L,t) = 0 \\ y(x,a) = y_a, & y(x,b) = y_b \end{array} \; 0 \le x \le L \right\}$$

In the special case where

$$J(y) = \int_a^b \int_0^L f(x, t, y, y_x, y_t) dx dt$$

and the Lagragian has continuous partial derivative, we get that

$$\begin{aligned} \delta J(y; v) &= \frac{\partial J}{\partial \epsilon}(y + \epsilon v)\Big|_{\epsilon=0} \\ &= \int_a^b \int_0^L f_y[y]v + f_{y_x}[y]v_x + f_{y_t}[y]v_t dx dt \end{aligned}$$

Because we're integrating with respect to separate variables, we can interchange the orders of integration, simplifying the internal integrands as follows

$$\int_a^b \int_0^L f_{y_x}[y]v_x \; dx \; dt = \int_a^b \left(f_{y_x}[y]v \; \Big|_{x=0}^{L} - \int_0^L \frac{d}{dx} f_{y_x}[y]v \; dx \right) dt$$
$$\int_0^L \int_a^b f_{y_t}[y]v_t \; dt \; dx = \int_0^L \left(f_{y_t}[y]v \; \Big|_{t=a}^{b} - \int_a^b \frac{d}{dt} f_{y_t}[y]v \; dt \right) dx$$

Substituting these back into our original expression for $\delta J(y; v)$ we get

$$\delta J(y; v) = \int_a^b \int_0^L \left[f_y[y] - \frac{d}{dx} f_{y_x}[y] - \frac{d}{dt} f_{y_t}[y] \right] v(x) \; dx \; dt$$
$$+ \int_a^b f_{y_x}[y]v(x,t) \; \Big|_{x=0}^{L} dt + \int_0^L f_{y_t}[y]v(x,t) \; \Big|_{t=a}^{b} dx$$

The boundary conditions over the set of admissable functions implies that

$$v(x,a) = 0, \quad v(x,b) = 0, \quad \forall 0 \le x \le L$$
$$v(0,t) = 0, \quad v(L,t) = 0, \quad \forall a \le t \le b$$

A stationary function must then satisfy

$$f_y[y] - \frac{d}{dx} f_{y_x}[y] - \frac{d}{dt} f_{y_t}[y] = 0 \tag{19}$$

Example: Vibrating String

The system energies for the vibrating string are given by

$$T = \int_0^L \frac{1}{2} \rho y_t(x,t)^2 \; dx$$
$$U = \int_0^L \tau(x,t) \left(\sqrt{1 + y_x^2} - 1 \right) \; dt$$
$$f = \frac{1}{2} \rho y_t(x,t)^2 - \tau(x,t) \left(\sqrt{1 + y_x^2} - 1 \right)$$

3 Optimal Control

3.1 Introduction

Control theory is the study of dynamical systems wherein we are given limited control of the system mechanics through input variables. Our goal is analyze how we can affect the system to manipulate it into achieving some desired state. For our purposes, we shall be considering the field of optimal control theory, which we will see is a natural extension of variational calculus. In general, we will want to consider a problem of the following form:

General Problem Statement:

Let $f : \mathbb{R}^3 \to \mathbb{R}, g : \mathbb{R}^3 \to \mathbb{R}^n$ and define the cost functional

$$J(y,u) = \int_{t_0}^{t_f} f(t,y,u)\ dt \qquad \begin{matrix} y \in C^1\left([t_0,t_f];\mathbb{R}^n\right) \\ u \in C\left([t_0,t_f];\mathbb{R}^m\right) \end{matrix} \quad m \leq n \tag{20}$$

We want to minimize $J(y,u)$ whose value is subject to dynamics governed by a system of differential equations of the form

$$\begin{aligned} \dot{y}(t) &= g(t,y,u) \\ y(0) &= y_0 \end{aligned} \tag{21}$$

The values t_0, t_f are the initial and final times respectively. In a system that is conservative in time, we can set $t_0 = 0$ without loss of generality. Notice that $g(t,y,u)$ is a function of $u(t)$, a parameter that we are given the freedom to control (though with possible restriction as we shall see later). Because of it's nature, we shall refer to $u(t)$ as the control variable which may itself be a vector. Thus our goal in general will be to find a $u(t)$ that will minimize our cost functional given in (20).

Examples:

1. Consider a two dimensional dynamical system, wherein we are tasked with the minimization of $J(y,u) = \int_0^T u(t)^2\ dt$ subject to the differential equation given by

   $$\begin{aligned} \dot{y}_1(t) &= y_2(t) \\ \dot{y}_2(t) &= u(t) \end{aligned}$$

   Note that from our general problem statement, we have that

   $$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \qquad f(t,y,u) = u(t)^2, \qquad g(t,y,u) = \begin{pmatrix} y_2(t) \\ u(t) \end{pmatrix}$$

   This cost functional can be interpreted as minimizing the work done in controlling the system. Notice that we must consider $f(t,y,u) = u(t)^2$ so that the integrand is convex and non-negative.

2. Minimize the functional $J(y,u) = \int_0^T dt$ subject to the differential system

   $$\dot{y}(t) = g(t,y,u), \qquad y(0) = 0, y(T) = y_f$$

   In this case, $J(y,u)$ has a very simple form, and can be evaluated to yield the total time. This is an example of *time-optimal control*, wherein we want to drive the system to a desired state in the least amount of time possible.

3. In the future we will consider the dynamics of an inverted pendulum. Let Q be a positive semi-definite matrix. The cost functional that we will want to consider is a special case of an *Infinite Horizon Linear Quadratic Regulator* and has the form

   $$J(y,u) = \int_0^\infty y^\dagger(t)Qy(t) + u(t)^2\ dt$$

   where $\dagger$ represents the adjoint operator (in finite dimensions, this is the Hermitian transpose).

4. Control theory can also play a role in systems biology. Certain insects, such as wasps, can be classified according to a caste system. For our purposes, let there be two groups consisting of workers (w) and reprodutives (r). The reproductives are responsible for perpetuating the existence of a colony through reproduction, and the workers are responsible for tasks such as food gathering. Furthermore, assume that at the end of each season, the worker class dies off, leaving only the reproductives. These two groups are dependent on one another for survival and we can model this behaviour as follows:

$$\begin{aligned} \dot{w}(t) &= a_1 u(t) w(t) - b_1 w(t) \\ \dot{r}(t) &= a_2 (1 - u(t)) w(t) - b_2 r(t) \end{aligned}$$

To simulate this natural process, we want to maximize $r(T)$, where the terminal point T represents the time at which the season ends. Our hypothesis is that through adaptation, the population has a breeding strategy that maximizes $r(T)$.

Even though in our general problem statement we did not specify the terminal point $y(t_f)$, we have already analyzed the necessary framework for finding the transversality conditions. Our next obvious topic of discussion will be how to solve such optimal control problems. A typical mathematical strategy is to reduce the problem to one that we have already solved.

Idea 1 Solve the differential equation for y and substitute this into the cost functional to find a suitable condition on u. The problem with this strategy is that y doesn't always appear directly in the cost functional (such as in example 1 and 2). Furthermore, we generally can't find analytical solutions to the differential equation.

Idea 2 Substitute an expresion for $u(t)$ given from the differential equations in hopes of finding constraints on $u(t)$ from $\dot{y} = g(t, y, u)$. We face a problem in that not all differential equations may be solvable in terms of $u(t)$. For example, let $y(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ for $m < n$. Even in the special case of linear dynamics

$$\dot{y} = Ay + Bu, \qquad A \in M_n(\mathbb{C}), B \in M_{n \times m}(\mathbb{C})$$

the matrix B is not square and hence not invertible.

While this may not be ideal in the general case, let us consider an example where we are able to apply this technique. This will give us motivation towards the more general schema developed later.

Minimize $\displaystyle\int_0^T u(t)^2 dt$ subject to the system dynamics $\ddot{y}(t) = u(t)$ with initial conditions given by

$$\begin{array}{ll} y(0) = y_{10} \neq 0 & \dot{y}(0) = 0 \\ y(T) = 0 & \dot{y}(T) = 0 \end{array} \tag{22}$$

Substitute $\ddot{y} = u$ into the cost functional J to find that $J(y) = \displaystyle\int_0^t \ddot{y}(t)^2 \, dt$. We know how to handle

this situation, and revert to using the Euler-Lagrange equation

$$\begin{aligned} f_y - \frac{d}{dt} f_{\dot{y}} + \frac{d^2}{dt^2} f_{\ddot{y}} &= 0 \\ \frac{d^2}{dt^2} 2\ddot{y}(t) &= 0 \\ \ddot{y}(t) &= at + b \\ y(t) &= \frac{a}{6}t^3 + \frac{b}{2}t^2 + ct + d \end{aligned}$$

It is then a simply matter of applying the boundary conditions from (22) to find that

$$a = \frac{12y_{10}}{T^3}, \qquad b = -\frac{6y_{10}}{T^2}$$

However, our goal was to find the optimal ontrol $u(t)$ which is given by $\ddot{y(t)} = u(t)$, so $u(t) = at + b$.

General Approach

Recall that we want to minimize the functional

$$J = \int_0^T f(t, y, u)\ dt$$

subject to the differential equation

$$\begin{aligned} \dot{y}(t) &= g(t, y, u) \\ y(0) &= y_0 \end{aligned}$$

By analogy with a constraint on a minimization problem in $\mathbb{R}^n$, we will try an approach similar to that of Lagrange multipliers. Define the augmented problem as minimization of

$$\tilde{J}(y, u, p) = \int_0^T f(t, y, u) + p(t)^* \left(g(t, y, u) - \dot{y}(t) \right)\ dt, \qquad y(0) = y_0$$

Where p is the continuous analog of a Lagrange multiplier. The set of admissable functions will be

$$y, p \in C^1\left([0, T]; \mathbb{R}^n\right) \qquad u \in C\left([0, T]; \mathbb{R}^m\right)$$

Since f is scalar valued, we can define

$$f_y = \left[\frac{\partial f}{\partial y_1}, \cdots, \frac{\partial f}{\partial y_n} \right]$$

However, g is a vector itself and as such have several different notions of derivatives with respect to g. These include

$$(g_i)_y = \left[\frac{\partial g_i}{\partial y_1}, \cdots, \frac{\partial g_i}{\partial y_n} \right], \qquad g_{y_i} = \begin{bmatrix} \frac{\partial g_1}{\partial y_i} \\ \vdots \\ \frac{\partial g_n}{\partial y_i} \end{bmatrix}$$

The most general form is a combination of the previous two respresentations, and is easily recognizable as the Jacobian of g.

$$g_y = \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} & \cdots & \frac{\partial g_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial y_1} & \frac{\partial g_n}{\partial y_2} & \cdots & \frac{\partial g_n}{\partial y_n} \end{pmatrix}$$

Or more succinctly, $(g_y)_{ij} = \dfrac{\partial g_i}{\partial y_j}$. With these definitions in mind, we can state our first result concerning stationary functions of the augmented cost functional.

Theorem 3.1. *Consider the augmented cost functional given by*

$$\tilde{J}(y,u,p) = \int_0^T f(t,y,u) + p(t)^* \Big(g(t,y,u) - \dot{y}(t) \Big) \; dt$$

Then the triple $(\hat{y}, \hat{u}, \hat{p})$ that satisfies the following system of differential equations

$$\begin{aligned} \dot{p} &= -f_y^* - g_y^* p \\ \dot{y} &= g(t,y,u) \\ f_u &= p^* g_y \end{aligned} \tag{23}$$

is a stationary point for $\tilde{J}(y,u,p)$ with boundary conditions $y(0) = y_0, y(T) = y_f$. In the event that $y(T)$ is unspeficied, we further demand the transversality condition that $p(T) = 0$.

Proof. For brevity sake, define

$$\tilde{f}(t,y,u) = f(t,y,u) + p(t)^* \Big(g(t,y,u) - \dot{y}(t) \Big)$$

We need to consider three variations on the arguments of J, say v_1, v_2, v_3. Then we can write the Gateaux derivative as

$$\begin{aligned} \delta \tilde{J}((y,u,p);\vec{v}) &= \frac{\partial \tilde{J}}{\partial \epsilon}((y,u,p) + \epsilon \vec{v}) \Big|_{\epsilon=0} \\ &= \int_0^T \sum_{i=1}^n \tilde{f}_{y_i} v_{1,i} + \sum_{i=1}^n \tilde{f}_{\dot{y}_i} \dot{v}_{1,i} + \sum_{i=1}^m \tilde{f}_u v_{2,i} + \sum_{i=1}^n \tilde{f}_p v_{3,i} \; dt \\ &= \int_0^T \sum_{i=1}^n \left(\tilde{f}_{y_i} - \frac{d}{dt} \tilde{f}_{\dot{y}_i} \right) v_{1,i} + \sum_{j=1}^m \tilde{f}_{u_j} v_{2,j} + \sum_{i=1}^n \tilde{f}_{p_i} v_{3,i} \; dt + \sum_{i=1}^n \tilde{f}_{\dot{y}_i} v_{1,i} \Big|_{t=0}^T \end{aligned}$$

Thus for the triple (y,u,p) to be a stationary function, we demand that the Gateaux varation disappear for all admissable variations satisfying the boundary conditions $v_i(0) = v_i(T) = 0$. This yields the follow set of equations

$$\begin{aligned} \tilde{f}_{y_i} - \tfrac{d}{dt} \tilde{f}_{\dot{y}_i} &= 0 \qquad & i = 1, \dots n \\ \tilde{f}_{p_i} &= 0 \qquad & i = 1, \dots, n \\ \tilde{f}_{u_j} &= 0 \qquad & j = 1, \dots, m \end{aligned}$$

Now using our definition of $\tilde{f}$ we get rewrite these as

$$\begin{aligned} f_{y_i} + p_i g_{y_i} + \dot{p}_i &= 0 \qquad & i = 1, \dots, n \\ g_i(t, y, u) - \dot{y}_i &= 0 \qquad & i = 1, \dots, n \\ f_{u_j} + p^\dagger g_{u_j} &= 0 \qquad & j = 1, \dots, m \end{aligned}$$

These equations correspond directly to (23). In the event that $y(T)$ is not specified, then

$$\tilde{f}_{\dot{y}_i}[y(T), u(T), p(T)] = 0$$

which corresponds to $p(T) = 0$. □

Proposition 3.2. *If $(\hat{y}, \hat{u}, \hat{p})$ minimizes $\tilde{J}$ subject to initial condition $y(0) = y_0$, then $(\hat{y}, \hat{u})$ minimize J subject to $y(0) = y_0, \dot{y} = g(t, y, u)$.*

Proof. Suppose that $(\hat{y}, \hat{u}, \hat{p})$ minimizes $\tilde{J}$. Then the differential equation is satisfied and $J(\hat{y}, \hat{u}) = \tilde{J}(\hat{y}, \hat{u}, \hat{p})$. For any other $\hat{y} + v_1, \hat{u} + v_2$ that satisfies the differential equatoin and boundary condition, we have

$$J(\hat{y} + v_1, \hat{u} + v_2) - J(\hat{y}, \hat{u}) = \tilde{J}(\hat{y} + v_1, \hat{u} + v_2, \hat{p}) - J(\hat{y}, \hat{u}, \hat{p}) \geq 0$$

□

Example: Minimize $J(y, u) = \displaystyle\int_0^T u(t)^2 \, dt$ subject to

$$\ddot{y}(t) = u(t), \qquad \begin{matrix} y(0) = y_0 & \dot{y}(0) = 0 \\ y(T) = 0 & \dot{y}(T) = 0 \end{matrix} \tag{24}$$

This is a second order differential equation, so in order to apply our formalism we must first convert this into a system of first order equations. This can be done as follows

$$\begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{pmatrix} = \begin{pmatrix} y_2(t) \\ u(t) \end{pmatrix}$$

Now $f(t, y, u) = u^2$ and $y \in \mathbb{R}^3$ so we can calculate our derivatives as follows

$$f_y = \begin{pmatrix} 0 & 0 \end{pmatrix}, \qquad f_u = 2u, \qquad g_y = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad g_u = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

So the optimality equation for p is as follows

$$\begin{aligned} \dot{p} &= -f_y^\dagger - g_y^\dagger p \\ &= -\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} p \\ \begin{pmatrix} \dot{p}_1(t) \\ \dot{p}_2(t) \end{pmatrix} &= \begin{pmatrix} 0 \\ -p_1(t) \end{pmatrix} \end{aligned}$$

This is a very simply system of equations to solve. Since $\dot{p}_1(t) = 0$ then $p_1(t) = c_1$ for some constant $c_1 \in \mathbb{R}$. Then $\dot{p}_2(t) = c$ implies that $p_2(t) = c_1 t + c_2$.

The optimality equation for u is

$$\begin{aligned} 0 &= f_u + p^\dagger y_u \\ &= 2u + \begin{pmatrix} p_1 & p_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned}$$

This implies that $u(t) = at + b$ where $a = \frac{c_1}{2}, b = \frac{c_2}{2}$. Finally, the last optimality condition yields

$$\begin{aligned} \dot{y_1}(t) &= y_2(t) \\ \dot{y_2}(t) &= u(t) = at + b \\ \Rightarrow y_2(t) &= \frac{1}{2}at^2 + bt + c \\ \Rightarrow y_1(t) &= \frac{1}{6}at^3 + \frac{1}{2}bt^2 + ct + d \end{aligned}$$

Using the initial conditions from (24) we get

$$\begin{array}{ll} y_1(0) = d & y_1(T) = \frac{a}{6}T^3 + \frac{b}{2}T^2 + y_{10} \\ y_2(0) = c & y_2(T) = \frac{a}{2}T^2 + bT \end{array}$$

We can solve this to find

$$a = \frac{12y_0}{T^3}, \qquad b = -\frac{6y_0}{T^2}$$

Theorem 3.3. *Assume that $\tilde{f}$ is pointwise convex in y, u on $[0,T] \times D \times U$ where $D \subseteq \mathbb{R}^N$ is open. Then each solution $y_0 \in D, u_0 \in U, p_0$ of (23) minimizes J subject to (21) on*

1.

$$\mathcal{D}_T = \left\{ (y,u) \in D \times U, \,\middle|\, y \in C^1\left([0,T];\mathbb{R}^n\right), u \in C\left([0,T];\mathbb{R}^m\right) \right\}$$

in the case when $p(T) = 0, (y_0, u_0) \in \mathcal{D}_T$.

2.

$$\mathcal{D} = \left\{ (y,u) \in D \times U \,\middle|\, y \in C^1\left([0,T];\mathbb{R}^n\right), u \in C\left([0,T];\mathbb{R}^m\right), y(0) = y_0, y(T) = y_T \right\}$$

if $(y_0, u_0) \in \mathcal{D}$.

Proposition 3.4. *If $f(t,y,u)$ is pointwise convex in y, u and $g(t,y,u)$ is linear in y, u then $\tilde{f}$ is pointwise convex in y, u.*

Example: Minimize the cost functional $\displaystyle\int_0^T u(t)^2 dt$ subject to $\ddot{w}(t) = u(t)$. It is easy to see that the integrand of the cost functional is strictly pointwise convex in u, and that the differential equation is linear. Hence the augmented cost functional will also be pointwise convex.

3.2 Finite-Horizon Linear Quadratic Control

In this section we will consider the case of finding optimal controls for a very specific form of the cost functional. We can write the corresonding cost functional as

$$\begin{aligned} J(u) &= \int_0^T \|Cy(t)\|^2 + \|Eu(t)\|^2 \ dt \\ &= \int_0^T \langle Cy(t), Cy(t)\rangle + \langle Eu(t), Eu(t)\rangle \ dt \\ &= \int_0^T y(t)^\dagger \underbrace{C^\dagger C}_{Q} y(t) + u(t)^\dagger \underbrace{E^\dagger E}_{R} u(t) \ dt \end{aligned}$$

Note that Q is symmetric and positive semidefinite since $\langle Qy, y\rangle = \langle Cy, Cy\rangle = \|Cy\|^2 \geq 0$. R is assumed to be symmetric and postive definite so that $u^\dagger Ru > 0, \forall u \in U\backslash_{\{0\}}$.

Furthemore, consider the linear system of differential equations given by

$$\begin{aligned} \dot{y}(t) &= Ay(t) + Bu(t) \qquad A \in M_n(\mathbb{R}), B \in M_{n\times n}(\mathbb{R}) \\ y(0) &= y_0 \end{aligned}$$

Then

$$\begin{aligned} f_y &= 2y^\dagger Q \quad f_u = 2u^\dagger R \\ g_y &= A \qquad\quad g_u = B \end{aligned}$$

The optimality conditions from (23) are transformed as follows

$$\begin{aligned} \dot{p} &= -f_y^\dagger - g_y^\dagger p \\ &= -2Qy(t) - A^\dagger p(t) \\ \dot{y}(t) &= Ay(t) + Bu(t) \\ 0 &= f_u + p^\dagger g_u \\ &= 2u^\dagger R + p^\dagger B \\ &= 2Ru + B^\dagger p \\ u(t) &= -R^{-1}B^\dagger \frac{1}{2} p(t) \end{aligned}$$

If we define $z(t) = \frac{1}{2}p(t)$ then the non-trivial optimality equations become

$$\begin{aligned} \dot{z}(t) &= -A^\dagger z(t) - Qy(t) \\ u(t) &= R^{-1}B^\dagger z(t) \end{aligned}$$

Thus we can limit ourselves to considering the transformed system given by

$$\begin{aligned} \dot{z}(t) &= -A^\dagger z(t) - Qy(t) \\ \dot{y}(t) &= Ay(t) - BR^{-1}B^\dagger z(t) \\ y(0) &= y_0 \qquad z(T) = 0 \end{aligned}$$

Let $y_u(t)$ indicate the state of the system with control u, and $\exp(At)$ the matrix exponential of At. Then we can solve this system as

$$y_n(t) = \exp(At)y_0 + \int_0^t \exp(A(t-s))Bu(s)\ ds$$
$$z(t) = \int_t^T \exp(A^\dagger(s-t))Qy_u(s)\ ds$$

and so the optimal control is given by

$$\hat{u}(t) = -R^{-1}B\int_t^T \exp(A^\dagger(s-t))Qy_u(s)\ ds$$

We notice however that this equation is anti-causal in that is depends on future values of $y_u(t)$.

Theorem 3.5. *The minimum cost for the quadratic cost problem (20) subject to a linear differential equation (21) is* $J(\hat{u}, T) = y_0^\dagger P(0,T)y_0$ *where* $P(t,T)$ *is the solution to the* Differential Riccati Equation *given by*

$$\dot{P}(t,T) + A^\dagger P(t,T) + P(t,T)A - P(t,T)BR^{-1}B^\dagger P(t,T)$$
$$P(T,T) = \theta$$

where θ *is the matrix of all zeroes. The corresponding optimal control is then given by*

$$\hat{u}(t) = -\underbrace{R^{-1}B^\dagger P(t,T)}_{K(t)}y_u(t)$$

The augmented cost functional for this finite-time linear quadratic control problem is

$$\int_0^T \underbrace{y(t)^\dagger Qy(t) + u(t)^\dagger Ru(t) + p(t)^\dagger(Ay(t) + Bu(t) - \dot{y}(t))}_{\tilde{f}(y,u,p)}\ dt$$

This yields three sets of Euler-Lagrange equations

$$\tilde{f}_y - \frac{d}{dt}\tilde{f}_{\dot{y}} = 0$$
$$\tilde{f}_p - \frac{d}{dt}\tilde{f}_{\dot{p}} = 0$$
$$\tilde{f}_u - \frac{d}{dt}\tilde{f}_{\dot{u}} = 0$$

Evaluating using the augmented functional we find that

$$\begin{aligned}
0 &= \tilde{f}_y - \frac{d}{dt}\tilde{f}_{\dot{y}} \\
&= 2y(t)^\dagger Q + p(t)^\dagger A + \dot{p}(t)^\dagger \\
\dot{z}(t) &= -Qy(t) - A^\dagger z(t) && \text{where } z(t) = \frac{1}{2}p(t) \\
0 &= \tilde{f}_p - \frac{d}{dt}\tilde{f}_{\dot{p}} \\
&= Ay(t) + Bu(t) - \dot{y}(t) \\
\dot{y}(t) &= Ay(t) + Bu(t) \\
0 &= \tilde{f}_u - \frac{d}{dt}\tilde{f}_{\dot{u}} \\
&= 2Ru(t) + B^\dagger p(t) \\
u(t) &= -\frac{1}{2}R^{-1}B^\dagger p(t) = -R^{-1}B^\dagger z(t)
\end{aligned}$$

From these simplifications, we get the following linear system of equations

$$\begin{aligned}
\dot{z}(t) &= -Qy(t) - A^\dagger z(t) && z(T) = 0 \\
\dot{y}(t) &= Ay(t) - BR^{-1}B^\dagger z(t) && y(0) = y_0
\end{aligned}$$

By solving this system of equations, we can find all of the necessary components to solve for $u(t)$, our optimal control. Exploiting the matrix exponential, we then have

$$\begin{aligned}
y_u(t) &= \exp(At)y_0 + \int_0^t \exp(A(t-s))Bu(s)\ ds \\
z(t) &= \int_t^T \exp(A^\dagger(s-t))Qyu(s)\ ds \\
\hat{u}(t) &= -R^{-1}B^\dagger \int_t^T \exp(A^\dagger(s-t))Q\hat{y}(s)\ ds
\end{aligned}$$

Example:

Minimize the cost functional $\displaystyle\int_0^T 3y(s)^2 + u(s)^2\ ds$ subject to the differential equation

$$\dot{y}(t) = y(t) + u(t), \qquad y(0) = y_0$$

This is a scalar problem, but we can nonetheless apply our previous techniques in order to solve. Setting $Q = 3, R = 1, A = 1, B = 1$ we have the the differential Ricatti equation is given by

$$\dot{p}(t) + 2p(t) - p(t)^2 + 3 = 0, \qquad p(T) = 0$$

While this is non-linear, it is solvable and has solution

$$p(t) = \frac{3(1 - e^{-4(T-t)})}{1 + 3e^{-4(T-t)}}$$

and so the optimal control is given by $\hat{u}(t) = -p(t)\hat{y}(t)$.

3.3 Linear Quadratic Regulators

In this section, we will limit our system dynamics to strictly linear systems so that they can be described by

$$\dot{y}(t) = Ay(t) + Bu(t)$$

In this situation however, we shall consider an infinite horizon functional

$$J(u) = \int_0^\infty y(t)^\dagger Q y(t) + u(t)^\dagger R u(t)\ dt$$

where as before, we assume that Q, R are symmetric and positive (semi-)definite. Because we are considering an infinite integral, we must ensure that our functions are in appropriate spaces to avoid general blow up.

Definition 3.6. The system (A, B) is **open loop stabilizable** if for every initial condition y_0 there is a control $u \in L_2\left((0,\infty);\mathbb{R}^m\right)$ such that $y \in L_2\left((0,\infty);\mathbb{R}^m\right)$.

Lemma 3.7. *Consider the finite time problem. Then the differential equation*

$$\begin{gathered}\dot{\Pi}(t) = A^\dagger \Pi(t) + \Pi(t)A - \Pi(t)BR^{-1}B^\dagger \Pi(t) + q \\ \Pi(0) = \theta\end{gathered} \tag{25}$$

has a unique, symmetric, non-negative solution. Furthermore,

$$y_0^\dagger \Pi(T) y_0 = \min_{u \in L_2\left((0,\infty);\mathbb{R}^m\right)} J(u;T)$$

Proof. Define $\Pi_T(t) = P(T-t,T)$ and notice that $\Pi_T(0) = P(T,T) = 0$ and $\Pi_T(t)$ satisfies (25). Thus we need to show that $\Pi_T(t)$ is independent of T. Choose any t_1, t_2 and without loss of generality, take $t_1 < t_2$. Then $\Pi_{t_1}(t), \Pi_{t_2}(t)$ are symmetric, continuous on $[0, t_1]$ and satisfy (25). Since the solution to (25) is unique, it must then follow that

$$\Pi_{t_1}(t) = \Pi_{t_2}(t) \qquad 0 \le t \le t_1$$

Since t_1, t_2 were chosen arbitrarily, we conclude that $\Pi_T(t)$ is independent of T and write $\Pi(t)$ for brevity. □

Corollary 3.8. *Given the problem set up described in Lemma 3.7, the function $y_0^\dagger \Pi(t) y_0$ is monotonic in t. That is,*

$$y_0^\dagger \Pi(t_1) y_0 \le y_0^\dagger \Pi(t_2) y_0, \qquad \forall y_0, 0 \le t_1 \le t_2$$

Proof. Notice that

$$\begin{aligned} y_0^\dagger \Pi(t_2) y_0 &= \min_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u;t_2) \\ &\ge \min_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u;t_1) \\ &= y_0^\dagger \Pi(t_1) y_0 \end{aligned}$$

which is precisely the result we wanted to show. □

Lemma 3.9. *Suppose that* Π *is a symmetric solutoin to the algebraic Riccati equation (ARE) given by*

$$A^\dagger \Pi + \Pi A - \Pi B R^{-1} B^\dagger \Pi + q = 0 \tag{26}$$

Then for every $u \in L_2\left((0,t_2);\mathbb{R}^m\right)$ *and* $t \geq 0$ *we have that*

$$J(u;t) = y_0^\dagger \Pi y_0 - y(t)^\dagger \Pi y(t) + \int_0^t \left[u(s) + R^{-1}B^\dagger \Pi y(s)\right]^\dagger R[u(s) + R^{-1}B^\dagger \Pi y(s)]\, ds$$

Theorem 3.10. *Assume that given a dynamical system, it is open loop stabilizable. Then the infinite time problem has a minimum for each* y_0*. Furthermore, there exists a symmetric, non-negative* Π *such that*

$$\min_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u) = J(\hat{u}) = y_0^\dagger \Pi y_0$$

where $\hat{u}(t) = -R^{-1}B^\dagger \Pi y(t)$*. If* $\Pi(t)$ *solves the differential equation given by (25), define*

$$\Pi = \lim_{t\to\infty} \Pi(t)$$

Equivalently, Π *is the minimal non-negative solution to the algebraic Ricatti equation (26).*

Proof. Since the problem is open-loop stabilizable, $\exists \tilde{u} \in L_2\left((0,t_2);\mathbb{R}^m\right)$ such that

$$y_0^\dagger \Pi(t) y_0 = \min_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u,t) \leq J(\tilde{u})$$

Thus $\Pi(t)$ is a bounded, monotonic sequence and has a limit point Π. Π is non-negative and symmetric. Since $\Pi(t)$ satisfies the differential Riccati equation (25) the right hand side of (25) converges to

$$A^\dagger \Pi + \Pi A - \Pi B R^{-1} B^\dagger \Pi + q$$

The left hand side is $\dot{\Pi}(t)$ which must also converge. Since Π is constant, $\lim_{t\to\infty} \dot{\Pi}(t) = 0$ which is precisely the algebraic Riccati equation (26). Then

$$\begin{aligned} \inf_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u) &\geq \inf_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u,t) \\ &= y_0^\dagger \Pi(t) y_0 \end{aligned}$$

Since $\Pi = \lim_{t\to\infty} \Pi(t)$ we can take limits on both sides to find

$$\inf_{u \in L_2\left((0,t_2);\mathbb{R}^m\right)} J(u) \geq y_0^\dagger \Pi y_0$$

From Lemma 3.9 we have

$$J(u;t) \leq y_0^\dagger \Pi y_0 + \int_0^t [u(s) + R^{-1}B^\dagger \Pi y(s)]^\dagger R[u(s) + R^{-1}B^\dagger \Pi y(s)]\, ds$$

We can chooose $u = \hat{u}(s) = -R^{-1}B^\dagger \Pi y(s)$ so that

$$J(u;t) \leq y_0^\dagger \Pi y_0$$
$$J(\hat{u}) \leq y_0^\dagger \Pi y_0$$

It remains to show that $\hat{u} \in L_2\Big((0,t_2);\mathbb{R}^m\Big)$. Let r be the smallest eigenvalue of R. Since R is positive definite, it must follow that $r > 0$ and that $\forall v \in \mathbb{R}^m$

$$v^\dagger v \leq \frac{1}{r} v^\dagger R v$$

Thus we have

$$\begin{aligned}\int_0^\infty \|\tilde{u}(t)\|^2 \; dt &\leq \frac{1}{r}\int_0^\infty \hat{u}(t)^\dagger R u(t) \; dt \\ &\leq \frac{1}{r} J(\hat{u}) \\ &= \frac{1}{r} y_0^\dagger \Pi y_0 < \infty\end{aligned}$$

Thus $\min\limits_{u \in L_2\Big((0,t_2);\mathbb{R}^m\Big)} J(u) = J(\hat{u}) = y_0^\dagger \Pi y_0$ □

3.4 Test for Open-Loop Stabilizability

Definition 3.11. A system (A, B) is **controllable** if for any initial condition y_0, terminal condition y_f, and time T, there is a piecewise continuous u such that $y(T) = y_f$.

Theorem 3.12. *If (A, B) is controllable then it's open-loop stabilizable.*

Proof. If a system is controllable, then $\forall T > 0$, there is a u so that $y(T) = 0$. Thus for all time, define

$$\hat{u}(t) = \begin{cases} u(t) & t < T \\ 0 & t > T \end{cases}$$

Then $\hat{u} \in L_2\Big((0,t_2);\mathbb{R}^m\Big)$. Since $y(t) = 0$ for $t \geq T$ then $y \in L_2\Big((0,t_2);\mathbb{R}^m\Big)$ also. □

Example: The converse of the Theorem 3.12 is false. Consider the differential system given by

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

For any initial conditoin $y(0)$ choose $u(t) = -2y_1(t)$ so that the equivalent scalar system becomes

$$\dot{y}_1(t) = -y_1(t)$$
$$\dot{y}_2(t) = -y_2(t)$$

which has a solution $y_i(t) = y_i(0)e^{-t}$. The cost functional will be finite, and so the system is open-loop stabilizable. However, this system is not controllable. In the event that $y_2(0) = 0$ then $y_2(t) = 0$ for all time $t \geq 0$ and we cannot steer the system to any point is $\Big\{(y_1, y_2) \;\Big|\; y_2 \neq 0\Big\}$.

Theorem 3.13. *The system (A, B) is controllable if and only if the* Kalman Controllability Matrix

$$C = \begin{pmatrix} B & AB & \cdots & A^{n-1}B \end{pmatrix}$$

has full row rank.

From this it is easy to see that the example above is not controllable.

Example: Test whether the system $\ddot{y}(t) = u(t)$ is controllable. If we break this into a scalar system of differential equations, we get

$$\begin{aligned} \dot{y}_1(t) &= y_2(t) \\ \dot{y}_2(t) &= u(t) \end{aligned}$$

Thus the matrix represenation of this linear system is

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and the corresponding Kalman controllability matrix is then

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which clearly has full row rank. Hence the system is controllable.

Example: Consider the problem of minimizing the vibrations of a flexible beam in moving between two points. Our control variable $u(t)$ shall represent the system torque. This system can be modelled by a partial differential equation, and while we haven't considered the problem of dealing with PDE's per se, we can attain arbitrary precision by truncating the Fourier series expansion, admitting a system of differential equations. The system's matrix representation is then given by

$$A = \begin{pmatrix} 0 & 1 & & & & & & & & \\ 0 & 0 & 0 & 1 & & & & & & \\ & & -\omega_1^2 & -2\xi\omega_1 & & & & & & \\ & & & & 0 & 1 & & & & \\ & & & & -\omega_2^2 & -2\xi\omega_2 & & & & \\ & & & & & & 0 & 1 & & \\ & & & & & & -\omega_3^2 & -2\xi\omega_3 & & \\ & & & & & & & & 0 & 1 \\ & & & & & & & & -\omega_4^2 & -2\xi\omega_4 \end{pmatrix} \qquad B = \frac{1}{0.0629} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2.8 \\ 0 \\ -2.3 \\ 0 \\ -0.91 \\ 0 \\ -4.5 \end{pmatrix}$$

Where $\omega = \begin{pmatrix} 55.89 & 131 & 313 & 603 \end{pmatrix}$ and $\xi = 0.002$. The Kalman controllability matrix is

$$C = \begin{pmatrix} B & AB & A^2B & \cdots & A^9B \end{pmatrix}$$

which has rank 10, therefore the system is controllable. Using matlab, we can then calculate the optimal control for various choices of Q, setting $R = 1$.

3.5 Reformulation of Optimality Conditions

Recall the augmented Lagrangian is given by

$$\tilde{f}(t,y,u) = f(t,y,u) + p(t)^\dagger \Big(g(t,y,u) - \dot{y}(t) \Big)$$

Then the optimality conditions were derived from the vector valued form of the Euler-Lagrange equations

$$\dot{p} = -f_y^\dagger - g_y^\dagger p \qquad \dot{y} = g \qquad f_u + p^\dagger g_u = 0$$

Define the **Pontryagin Hamiltonian** as $h = f + p^\dagger g$, and note that

$$\begin{aligned} h &= f + p^\dagger (g - \dot{y}) + p^\dagger \dot{y} \\ &= \tilde{f} + p^\dagger \dot{y} \\ &= \tilde{f} - \tilde{f_y} \dot{y} \end{aligned}$$

So $h = -H$ where H is typical system Hamiltonian. By taking derivatives of this Pontryagin Hamiltonian, we can find a connection with the aforementioned optimality conditions. Notice that the y derivative yields

$$h_y = f_y + p^\dagger g_y = (f_y^\dagger + g_y^\dagger p)^\dagger = -\dot{p}^\dagger$$

We can manipulate $h^\dagger$ to find that

$$\begin{aligned} h^\dagger &= f^\dagger + g^\dagger p \quad \Rightarrow (h^\dagger)_p = g^\dagger \\ h_p &= g = \dot{y} \end{aligned}$$

and finally, the control derivative is

$$h_u = f_u + p^\dagger g_u = 0$$

Since $h = -H$, we have that h is a continuous function of time. If f and g don't depend directly on time t then h is constant for stationary triples (y,u,p) that satisfy the optimality conditions. In summary, we can rewrite the optimality conditions in terms of the Pontryagin Hamiltonian as

$$-h_y = \dot{p}^\dagger, \qquad h_p = \dot{y}, \qquad h_u = 0$$

Example:

Consider the problem of minimizing the cost functiaonal $\displaystyle\int_0^T u(t)^2 \, dt$ subject to the system dynamics given by

$$\ddot{w}(t) = u(t) \quad \Leftrightarrow \quad \begin{matrix} \dot{y}_1(t) = y_2(t) & y_1(0) = y_{10} \\ \dot{y}_2(t) = u(t) & y_2(0) = y_{20} \end{matrix}$$

The Pontryagin Hamiltonian is then

$$h = f + p^\dagger g = u(t)^2 + p_1(t) y_2(t) + p_2(t) u(t)$$

Thus the optimality equation $\dot{p} = -h_y$ gives us

$$\begin{pmatrix} \dot{p}_1 \\ \dot{p}_2 \end{pmatrix} = - \begin{pmatrix} 0 \\ p_1 \end{pmatrix}$$

The second optimality condition $h_p = \dot{y}$ just gives us back our original system dynamics. And finally, the third optimality condition gives us that

$$0 = h_u = 2u + p_2 \tag{27}$$

Notice that neither f nor g are explicitly time dependent. We expect that the Pontryagin Hamiltonian will then be constant in time. We have that

$$\begin{aligned}
\dot{h} &= 2u\dot{u} + \dot{p}_1 y_2 + p_1 \dot{y}_2 + \dot{p}_2 u + p_2 \dot{u} \\
&= (2u + p_2)\dot{u} + p_1 u - p_1 u \\
&= 0 \qquad \text{by (27)}
\end{aligned}$$

Example: Consider the time-optimal control problem of driving the system with dynamics given by $\dot{y}(t) = u(t)$ from the point $y(0) = 1$ to $y(T) = 0$. Time-optimal control implies that we want to minimize the cost functional $\int_0^T dt$. The system Pontryagin Hamiltonian is then given by

$$h = f + p^\dagger g = 1 + pu$$

The optimality equations can then be easily calculatied to find

$$\dot{p} = 0, \qquad \dot{y} = u, \qquad p = 0$$

Thus there is no stationary value of h with respect to u (since h is linear in u). In fact $y(t) = 1 + \int_0^t u(s)\, ds$ so any control such that $\int_0^T u(s)\, ds = -1$ will lead to $y(T) = 0$. One choice is the constant function $u(t) = -\frac{1}{T}$. The infimum of the time required is 0, but this cannot be attained with a finite control $u(t)$, and certainly allowing $u(t)$ to take on infinite values is unreasonable. Hence most problems will impose limitations on the magnitude of u. For this problem, let us consider $|u(t)| \leq 2$. Then the minimum time is $\frac{1}{2}$ which is achieved by $u(t) = -2$. This problem is analogous to minimizing a linear function. Certainly no minimum exists over an infinite domain, but will occur at the endpoints of a closed domain.

The optimality condition $h_u = 0$ means that if h is convex, then h as a function of u is minimized at u^0. Define

$$h^0 = \min_{u \in \mathcal{U}} h(t, y^0(t), u(t))$$

where $\mathcal{U}$ is the set of admissable controls and y^0 is a stationary solution to the cost functional. This motivates the following theorem:

Theorem 3.14. *Suppose $\mathcal{U} \subseteq \mathbb{R}^m$, $J(y,u) = \int_0^T f(t,y,u)\, dt$ and define the Pontryagin Hamiltonian $h = f + p^\dagger g$ subject to the dynamical system $\dot{y} = g(t,y,u)$. Furthermore, let*

$$\mathcal{D}' = \left\{ (y,u) \;\middle|\; y \in \hat{C}^1\left([0,1]; \mathbb{R}^n\right), u \in \hat{C}\left([0,1]; \mathbb{R}^m\right), u(t) \in \mathcal{U}, y(0) = y_0 \right\}$$

and define

$$h^0(t,u) = \min_{u \in U} h(t,y,u)$$

If $h^0(t,u)$ is strongly pointwise convex, and y^0, u^0 are functions that satisfy the optimality conditions

$$\dot{p} = -h_y \qquad \dot{y} = h_p \qquad u = \operatorname*{argmin}_{u \in U} h(t, y^0, u)$$

on the set

$$\mathcal{D} = \begin{cases} \mathcal{D}' & \text{if } p(T) = 0 \\ \mathcal{D}' \cap \left\{ y \;\middle|\; y(T) = y_t \right\} & \text{otherwise} \end{cases}$$

then (y^0, u^0) *minimizes* $J(y, u)$.

Note: We can use the Weierstrass-Erdman Corner conditions for situations where y, u, p may not be smooth. We have that h is continuous and have assumed that $\tilde{f}_{\dot{y}}, \tilde{f}_{\dot{p}}, \tilde{f}_{\dot{u}}$ are continuous. Then $\tilde{f}_{\dot{p}} = 0, \tilde{f}_{\dot{u}} = 0$ but $\tilde{f}_{\dot{y}} = -p$. This implies that p is also continuous. Since $\dot{u}$ does not appear anywhere, we can also extend the theory in a straightforward manner to include the case where u is piecewise continuous. Since h is continuous, discontinuities of $u(t)$ are limited to those that satisfy

$$h(t, y^0(t), u^0(t_-)) = h(t, y^0(t), u^0(t_+))$$

In the literature, this is sometimes known as Pontryagin's *Maximum* Principle, but in this case $h = H$ and the problem becomes a maximization rather than a minimization.

Examples:

1. Minimize the cost functional $\int_0^1 \frac{1}{2}y(t)^2 + \frac{1}{2}u(t)^2 \, dt$ subject to the dynamics

$$\begin{aligned} \dot{y}(t) &= u(t) \\ y(0) &= 1 \end{aligned}$$

   with control constraint $|u(t)| \leq 4$.

   The first thing we want to do is find the form of the Pontryagin Hamiltonian.

$$h = f + pg = \frac{1}{2}y^2 + \frac{1}{2}u^2 + pu = \frac{1}{2}y^2 = u\left(\frac{u}{2} + p\right)$$

   Stationary functions must next satisfy the optimality conditions. It is easy to see that $h_u = u + p$, and so the argument minmum of this function will occur when $u^0 = -p$. However, our constraints on the magnitude of u means we must take care in so liberally assigning u^0 a value. In particular, we can define

$$u^0(t) = \begin{cases} -4 & \text{if } p(t) > 4 \\ -p(t) & \text{if } |p(t)| \leq 4 \\ 4 & \text{if } p(t) < -4 \end{cases}$$

   Due to the simplicity of this problem, we were easily able to find the form of $u^0(t)$. However, in the general case recall that we will have to resort to the differential Riccati equation to find $p(t) = P(t, 1)y(t)$.

2. Mnimize the cost function $\int_0^T dt$ subject to the system dynamics governed by

$$\begin{aligned} \dot{y}(t) &= u(t) \\ y(0) &= 1, y(T) = 0 \end{aligned}$$

   The form of the cost functional implies that this is a time-optimal control problem. Intuitively, we suspect that no optimal control will exist with an unconstrained control, since we will be able to

drive the system to the origin arbitrarily fast for a sufficient magnitude control. Hence assume that u is bounded; that is, $\exists M > 0$ such that $|u(t)| \leq M$. The Pontryagin Hamiltonian is given by $h = f + pg = 1 + pu$. To minize this Hamiltonian will use a control $u^0(t) = -\text{sgn}(p(t))M$. This is an example of a *bang-bang control*: the control switching between it's maximum and minimum values, and does not use any magnitude in between. Thus to completely characterize our solution, we must find the *switching time*, namely, the point(s) at which $p(t)$ changes sign.

Via our optimality equations, we know that $p(t)$ must satisfy $\dot{p}(t) = -h_y = 0$. This implies that p is a constant and that $u = \pm M, \forall t$. With $u = M$ we get $y(t) = 1 + Mt$ by solving the differential equatoin. Since $t \geq 0, y(T) = 0, u^0(t) = -M$ and the optimal time is $\frac{1}{M}$. In general, if $y(0) = y_0$ then $y(t) = y_0 \pm Mt$ and the optimal control will be given by $u^0(t) = -M\text{sgn}(y_0)$

3. Consider a problem from systems biology: a family of insects consists of two groups, workers (w) and reproductives (r). The workers harvest food, but die at the end of the season. The reproductives also harvest food, but survive at the end of the season. After each season, these roles change and the reproductives becomes workers. The relationship between these two groups can be modelled by

$$\begin{aligned} \dot{w}(t) &= a_1 u(t) w(t) - b_1 w(t) & w(0) &= 1 \\ \dot{r}(t) &= a_2(1 - u(t)) w(t) - b_w r(t) & r(0) &= 0 \end{aligned}$$

where $a_1 > b_1 > b_2 > 0$, and $u(t)$ is the effort to enlarging the working population, and $0 \leq u(t) \leq 1$. We hypothesize that evolution has maximized the number of reproductives at the end of each season, so our cust fucntion is to maximize $r(T) = \int_0^T \dot{r}(t)\ dt$, or conversely, to minimize

$$\begin{aligned} r(T) &= -\int_0^t \dot{r}(t)\ dt \\ &= \int_0^t -a_2(1 - u(t))w(t) + b_w r(t)\ dt \end{aligned}$$

Let $y_1 = w, y_2 = r$ so that the Pontryagin Hamiltonian is given by

$$\begin{aligned} h &= f + p^\dagger g \\ &= -a_2(1-u)y_1 + b_2 y_2 + p_1(a_1 u y_1 - b_1 y_1) + p_2(a_2(1-u)y_1 - b_2 y_2) \\ &= [-a_2 y_1 + b_2 y_2 - p_1 b_1 y_1 + p_2 a_2 y_1 - p_2 b_2 y_2] + u y_1 \underbrace{(a_2 + p_1 a_1 - p_2 a_2)}_{\psi(t)} \end{aligned}$$

Thus the optimal u minimizes h, which implies that the control has the form

$$u^0(t) = \begin{cases} 1 & y_1(t)\psi(t) < 0 \\ 0 & y_1(t)\psi(t) > 0 \end{cases}$$

with the case where $\psi(t) = 0$ as of yet undetermined. Let us take a moment to examine the dynamics themselves to see if they shed light on the problem. Notice that $\dot{y_1}(t) = (au(t) - b_1)y_1(t)$, so if $y_1(0) > 0$, then $y_1(t) > 0, \forall t$. Thus $\text{sgn}\psi(t)$ determines $u^0(t)$. We will that $\dot{\psi}(t_s) > 0$ if $\psi(t_s) = 0$, so there is at most one switching time. Our other optimality condition implies that

$$\begin{aligned} \dot{p}_1 &= -h_{y_1} = a_2(1-u) + p_1(b_1 - a_1 u) - p_2 a_2(1-u) \\ \dot{p}_2 &= -h_{y_2} = -b_2 + p_2 b_2 \\ p_2(t) &= 1 - e^{-b_2(t-T)} \end{aligned}$$

where $p_1(T) = p_2(T) = 0$. Thus

$$\begin{aligned}
\dot{\psi} &= \dot{p}_1 a_1 - \dot{p}_2 a_2 \\
&= p_1(b_1 - a_1 u)a_1 - p_2 a_2(1-u)a_1 + a_2(1-u)a_1 + b_2 u_2 - p_2 b_2 a_2 \\
&= a_1 p_1(b_1 - a_1 u) + (a_2 p_2 - a_2)(-(1-u)a_1 - b_2) \\
&= (a_2 p_2 - a_2)(b_1 - a_1 u - a_1 + a_1 u - b_2) = (a_2 p_2 - a_2)(b_1 - a_1 - b_2) \\
&= a_2(p_2 - 1)(b_1 - a_1 - b_2) \\
&= \underbrace{-a_2}_{<0}\underbrace{e^{b_2(t-T)}}_{>0}\underbrace{(b_1 - a_1 - b_2)}_{<0} > 0
\end{aligned}$$

Thus there are three possibilities. Either $t_s \in [0, T]$ in which case u fluctuates between $u(t) = 0, u(t) = 1$, or t_s lies either before or after this interval. However, notice that assume that $t_s > T$ implies that $r(T) = 0$, so that all birthrate effort is dedicated towards workers, which would lead to extinction. In the case where $t_s < 0$ then all effort goes towards enlarging $r(T)$.

4. **HIV Treatment Schedule**

Let Q be the quiscent T-cell population, T the active T-cell population, I the infected T-cell population, and $u(t)$ the drug concentration for $0 \leq u(t) \leq 1$. We can then model the relationship between these factors as

$$\begin{aligned}
\frac{dQ}{dt} &= a_1 T - a_2 Q \\
\frac{dT}{dt} &= a_3 Q - a_4 T I f(u) \\
\frac{dI}{dt} &= a_5 T I f(u) - a_6 I
\end{aligned}$$

where $f(u) = \frac{\ell}{1 + \frac{D}{R}u(t)}$, and a_i is constant $\forall i$. Define $y = \begin{pmatrix} Q & T & I \end{pmatrix}^T$. We want to choose $u(t)$ that minimize the infected population, namely $\displaystyle\int_0^{t_f} I(t)^2 \, dt$ subject to the previous differential equation with initial conditions given by $Q(0) = Q_0, T(0) = T_0, I(0) = I_0$.

The Pontryagin Hamiltonian is given by

$$h = f + p^\dagger g = I^2 + p_1(a_1 T - a_2 Q) + p_2(a_3 Q - a_4 T I f(u)) + p_3(a_5 T I f(u) - a_6 I)$$

so by the minimal principle, u will minimize h subject to the optimality constraints. Let $h_1 = I^2 + p_1(a_1 T - a_2 Q) + p_2 a_3 Q - p_3 a_6 I$, which is all the terms of h that do not depend on u. Thus we can rewrite the Hamiltonian as

$$h = h_1 + (p_3 a_5 - p_2 a_4)\frac{T I \ell R}{R + Du}$$

Since $h_u \neq 0$ for any u, and all variables are non-negative,

$$\begin{cases} 0 & p_3 a_5 - p_2 a_4 < 0 \\ 1 & p_3 a_5 - p_2 a_4 > 1 \end{cases}$$

Notice that $p_3a_5 - p_2a_4$ acts as a switching function. To find the switching times, we need to solve optimality conditions

$$\begin{aligned} \dot{p}_1 &= -h_Q = p_1a_2 - p_2a_3 & p_1(T) &= 0 \\ \dot{p}_2 &= -h_T = -p_1a_1 + p_2a_4If(u) - p_3a_5If(u) & p_2(T) &= 0 \\ \dot{p}_3 &= -h_I = -2I + p_2a_4Tf(u) + p_3a_6 & p_3(T) &= 0 \end{aligned}$$

5. Minimize the cost functional $\int_0^T dt$ with system dynamics given by $\ddot{w}(t) = u(t)$ and initial conditions $w(0) = y_{10}, \dot{w}(0) = y_{20}$ and constraint $|u(t)| \leq 1$. As before, we break this into a system of first order equations,

$$\begin{aligned} \dot{y}_1(t) &= y_2(t) & y_1(0) &= y_{10} \\ \dot{y}_2(t) &= u(t) & y_2(0) &= y_{20} \end{aligned}$$

The Hamiltonian is $h = 1 + p_1y_2 + p_2u$. By Pontryagin's Minimum principle, the optimal u will minimize h so

$$u^0(t) = -\text{sgn}(p_2(t))$$

To find the form of $p_2(t)$, we must consider the optimality conditions

$$\begin{aligned} \dot{p}_1 &= -h_{y_1} = 0 \\ \dot{p}_2 &= -h_{y_2} = -p_1 \end{aligned}$$

This can easily be solved to find that $p_2(t) = at + b$. Notice that if $p_i(t) = 0$ on an interval, then $c = k = 0$ which would cause $h = 1$. However, since the final time is fixed and the system is time invariant, the minimum principle implies that $h = 0$ for all time. Thus we cannot possible have that p_1, p_2 are both simultaneously zero on the same interval. Since p_2 is linear, there is at most one switch. The precise number of switches will depend on the initial conditions.

At the final time T, we have that $y_1(T) = y_2(T) = 0$. If $u(t) = 1$ then $\dot{y}_2(t) = 1$ which implies that $y_2(t) = t + c_2$ and $y_1(t) = \frac{t^2}{2} + c_2t + c_1$. The boundary conditions then imply that $c_2 = -T$ and $c_1 = \frac{T^2}{2}$. Thus

$$y_1(t) = \frac{1}{2}t^2 - Tt + \frac{1}{2}T^2 = \frac{1}{2}(t-T)^2, \qquad y_2(t) = t - T$$

Similarily, if $u(T) = -1$ then

$$y_1(t) = -\frac{1}{2}(t-T)^2, \qquad y_2(t) = -(t-T)$$

We can visualize the corresponding phase portrait by recognizing that $y_1 = \text{sgn}(u)y_2^2$.

In the general case, we had $y_2(t) = \frac{1}{2}t^2 + c_2t + c_1$ and $y_1(t) = t + c_2$ so that we can write

$$\begin{aligned} y_2^2 &= t^2 + 2c_2t + c_2^2 \\ &= 2\left(\frac{1}{2}t^2 + c_2t + c_1\right)\underbrace{-2c_1 + c_2^2}_{k} \\ &= 2y_1 + k \end{aligned}$$

which gives a family of curves in the parameter k.

Consider the problem of finding the time-optimal control of an n-dimensional system $\dot{y}(t) = Ay(t) + Bu(t)$ with initial and terminal conditions $y(0) = y_0, y(T) = 0$ respectively. Assume that each component of $u(t) \in \mathbb{R}^m$ satisfies

$$M_1 \leq u_k(t) \leq M_2, \qquad 1 \leq k \leq m$$

For this class of problems, the Pontryagin Hamiltonian wil be

$$h = f + p^\dagger g = 1 + p^\dagger(Ay + Bu)$$

Note that since the final time is not fixed, the problem is time-invariant, so $h(t) = 0, \forall t$ and the costate variables cannot be simulateneously zero. If we let $U = \left\{ u \in \mathbb{R}^m \;\middle|\; M_1 \leq u_i \leq M_2, 1 \leq i \leq m \right\}$, the optimality conditions imply that

$$\dot{p} = -A^\dagger p, \qquad \dot{y} = Ay + Bu, \qquad u = \underset{u \in U}{\operatorname{argmin}} h(t, y, u)$$

In the case of one control ($m = 1$), B is a column vector. To minimize h, u will be a bang-bang control of the form

$$u^0(t) = \begin{cases} M_1 & p^\dagger B > 0 \\ M_2 p^\dagger B < 0 \end{cases}$$

In the more general case of multiple controls $m \geq 1$, we have

$$h = 1 + p^\dagger Ay + \sum_{k=1}^{m} p^\dagger B_k u_k(t)$$

and the minimizing control is still bang-bang with

$$u_k^0(t) = \begin{cases} M_1 & p^\dagger B_k > 0 \\ M_2 p^\dagger B_k < 0 \end{cases}$$

The switching points will occur whenever $p^\dagger(t)B_k = 0$. Singular controls will occur whenever $p^\dagger(t)B_k = 0$ over a non-zero measure set.

Theorem 3.15. *If the vectors $B_k, AB_k, \ldots, A^{n-1}B_k$ are linearly indpendent then the optimal control $u_k^0(t)$ for the time-optimal problem is M_1 or $M_2, \forall t \geq 0$, except possibly at switching points.*

Proof. The Pontryagin's Minimum Principle implie that $u_k^0(t) = M_1$ or M_2 whenever $p^\dagger(t)B_k \neq 0$. This only fails to determine $u_k(t)$ when the switching function is precisely zero on some interval $[t_1, t_2]$. Let $\psi(t) = p^\dagger(t)B_k$, and notice that

$$\begin{aligned} \frac{d}{dt}\psi(t) &= \dot{p}^\dagger(t)B_k = -p(t)^\dagger AB_k \\ \frac{d^2}{dt^2}\psi(t) &= p(t)^\dagger A^2 B_k \\ \frac{d^r}{dt^r}\psi(t) &= p(t)^\dagger A^r \mod n B_k \end{aligned}$$

Thus we have a system of equations

$$p(t)^\dagger \underbrace{\begin{pmatrix} B_k & AB_k & \cdots & A^{n-1}B_k \end{pmatrix}}_{R(A,B)}$$

However, since $p(t)$ can never vanish, if $\psi(t) = 0$ on an interval, the matrix $R(A, B)$ must be singular implying that $\left\{ B_k, AB_k, \ldots A^{n-1}B_k \right\}$ are not a linearily independent set. □

Note that the linear independence of $\{B_k, AB_k, \dots A^{n-1}B_k\}$ is equivalent to rank$R(A,B) = n$ or $R(A,B)$ has full row rank. Hence if (A, B_k) is controllable, the time optimal control $u_k^0(t)$ is a bang-bang controller.

Theorem 3.16. *Consider the linear time-optimal contorl problem. If all the eigenvalues of A are real, and the component u_k^0 of the optimal control is bang-bang, there are at most $n-1$ switches in the value of u_k.*

Before we continue with the proof of this theorem, consider the following lemma

Lemma 3.17. *The sum $\sum_{k=1}^{n} \alpha_k e^{-\lambda_k t}$ for $\alpha_k, \lambda_k \in \mathbb{R}$ has at most $n-1$ roots.*

Proof of Theorem. For the sake of simplicity, assume that the eigenvalues $\lambda_1, \dots, \lambda_n$ are all distinct, there is only one control variable ($m = 1$), and that $|u(t)| \leq 1$. By the spectral theorem, we know that $\exists Q$ such that

$$-A = Q^{-1} \begin{pmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_n \end{pmatrix} Q$$

From Pontryagin's Minimum Principle, we have that $u^0(t) = -\text{sgn}(p(t)^\dagger B)$. Now $\dot{p}(t) = -A^\dagger p(t)$. We can exploit the matrix exponential to find the solution to this differential equation $p(t) = \exp(-A^\dagger t)p_0$. for some $p_0 \in \mathbb{R}^n$. But due to our diagonalization, we can very easily write out $\exp -A^\dagger t$ explicitly as

$$\exp(-At) = Q^{-1} \begin{pmatrix} e^{-\lambda_1 t} & & & \\ & e^{-\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{-\lambda_n t} \end{pmatrix} Q$$

Thus th eoptimal control is given by

$$\begin{aligned} u^0(t) &= -\text{sgn}\left(p_0^\dagger \exp(-At)B\right) \\ &= -\text{sgn}\left(p_0^\dagger Q^{-1} \begin{pmatrix} e^{-\lambda_1 t} & & & \\ & e^{-\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{-\lambda_n t} \end{pmatrix} QB\right) \\ &= -\text{sgn}\left(\sum_{k=1}^{n} \alpha_k e^{-\lambda_k t}\right) \end{aligned}$$

By Lemma 3.17, it follows that there are at most $n-1$ switches. □

Example:

Consider the time optimal control of a simple harmonic oscillator. We are given control of the forcing motion, and want to drive the system to the origin. The system dynamics are given by

$$\ddot{w}(t) + w(t) = u(t) \quad \Leftrightarrow \quad \begin{array}{l} \dot{y}_1(t) = y_2(t) \\ \dot{y}_2(t) = -y_1(t) + u(t) \end{array}$$

Let us impose the conditions that $y(0) = (y_{10}, y_{20}), y(T) = (0,0)$, and that $|u(t)| \leq 1$. The system is controllable as it can be shown that the controllability matrix is

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which has full row rank. The optimal control is thus bang-bang, but the eigenvalues are $\pm i$.

The Pontryagin Hamiltonian is $h = 1 + p_1 y_2 - y_1 p_2 + p_2 u$. The optimality equations are then

$$\begin{aligned} \dot{p}_1 &= -h_{y_1} = p_2 \\ \dot{p}_2 &= -h_{y_2} = -p_1 \end{aligned}$$

From the PMP, the optimal control u minimizes h so

$$u^0(t) = -\text{sgn} p_2(t)$$

Thus in order to completely characterize the optimal control, we need to find more information about $p_2(t)$. This is a fairly elementary differential equation, and can easily be solved in terms of $p_2(t)$ to find that

$$p_2(t) = c_1 \cos(t) + c_2 \sin(t) = A \cos(t + \phi)$$

Note that this implies that switches should occur with a period of π. Only the very first switch may have a switching time that is strictly less than π. From the differential equation,

$$(y_1 - u)\dot{y}_1 + \dot{y}_2 y_2 = 0$$

Thus if u is constant on an interval, we get

$$(y_1 - u)^2 + y_2^2 = c^2$$

on such an interval. These are circles with centre $(u, 0)$ and radius c. That is, $(1,0)$ if $u = 1$ or $(-1,0)$ if $u = -1$. We can express this in parametric from by

$$y_1 - u = c\cos\theta, \qquad y_2 = c\sin\theta$$

Substituting into the differntial equation yields

$$-c\dot{\theta}\sin\theta = c\sin\theta, \qquad \dot{\theta} = 1$$

Thus the time take between any two point is the angle between those points.

3.6 Summary of Optimal Control

In general, we want to consider the problem of minimizing the cost functional $J(u) = \int_0^T f(t, y, u)\ dt$ subject to the system dynamics

$$\dot{y} = g(t, y, u), \qquad y(0) = y_0$$

In order to accomodate this problem, we switch the the problem of the augmented cost functional

$$\tilde{J}(y, u, p) = \int_0^T f(t, y, u) + p(g(t, y, u) - y)$$

The following table describes the optimality equations:

Euler-Lagrange for $\tilde{J}$	Optimality Equations	Reformulation
$\frac{\partial \tilde{J}}{\partial y_i} - \frac{d}{dt}\frac{\partial \tilde{J}}{\partial \dot{y}_i} = 0$	$\frac{\partial f}{\partial y_i} + \sum_{j=1}^n p_n \frac{\partial g_j}{\partial y_i} - \dot{p}_i = 0$	$\dot{p}_i = -h_y^\dagger$
$\frac{\partial \tilde{J}}{\partial p_i} - \frac{d}{dt}\frac{\partial \tilde{J}}{\partial \dot{p}_i} = 0$	$\dot{y}_i = g_i(t, y, u)$	$\dot{y}_i = g_i$
$\frac{\partial \tilde{J}}{\partial u} - \frac{d}{dt}\frac{\partial \tilde{J}}{\partial \dot{u}} = 0$	$\frac{\partial f}{\partial u} + \sum_{j=1}^n \frac{\partial g_i}{\partial u} = 0$	$h_u = 0$

We have spent a great deal of time considering linear quadratic control and time-optimal control.

4 Midterm Review

4.1 Free Endpoint Boundary Conditions

Allow $y(b)$ to lie on the curve $\phi(x)$. Since $y(b)$ is not fixed, we must utilize the transversality condition:

$$\begin{aligned} &-H(b) + p(b)\varphi'(b) = 0 \\ &H = -f + f_z y' \\ &p = f_z \end{aligned}$$

The special cases occur when $y(b)$ is fixed, so that $p(b) = 0$, and when b is free, then $H(b) = 0$.

Example:

Find a smooth curve that will minimize distance from the origin to $y = x^2 - 1$. Our Lagrangian is given by $f[y] = \sqrt{1 + y'(x)^2}$. Without being explicitly stated, we can imply the following boundary conditions from the problem statement:

$$y(0) = 0, \qquad y(b) = b^2 - 1$$

Hence the Euler-Lagrange equation simplifies to

$$\begin{aligned} f_z &= c_1 \\ \frac{y'}{(1 + y'^2)^{\frac{1}{2}}} &= c_1 \end{aligned}$$

Rearranging we get that

$$y'(x) = \pm\sqrt{\frac{c_1}{1-c_1}} = c$$
$$y(x) = cx + d$$

Now the endpoint $y(0) = 0$ implies that $d = 0$, and the transversality condition simplifies as

$$\begin{aligned} -H[y(b)] + f_z[y(b)]\varphi'(b) &= 0 \\ \frac{1}{\sqrt{1+y'(b)^2}}\left[y'(b)\varphi'(b)+1\right] &= 0 \qquad\qquad y'(b) = c, \varphi'(b) = 2b \\ \frac{1}{\sqrt{1+c^2}}(2bc+1) &= 0 \end{aligned}$$

Thus we have $2bc + 1 = 0$. Furthermore, since $y(b)$ lies on φ we must have that $y(b) = \varphi(b)$ so $cb = b^2 - 1$. Hence we get that

$$2(b^2-1)+1 = 0$$
$$b = \pm\frac{1}{\sqrt{2}}$$

We check both conditions and find that both yield a value of 5 for the distance, so both are minimizing solutions.

Example: Find a stationary function of

$$J(y) = \int_0^4 \left(y'(x)-1\right)^2\left(y'(x)+1\right)^2 dx$$

subject to $y(0) = 0, y(4) = 2$ that has precisely one corner.

The Euler-Lagrange equation simplifies to $f_z = c = 4y'(y'^2 - 1)$. The Weierstrass-Erdmann conditions imply that both P and H are continuous at the corners

$$\begin{aligned} P &= 4y'(y'^2-1) = c \\ H &= -f + f_z y' \\ &= (y'-1)(y'-1)(3y'^2+1) \end{aligned}$$

It can be seen that the only reasonable possibility for switches at corners are the changes $y' = \pm 1$. We also require y to be continuous at the corner.

Case 1 Assume that $y'_- = 1, y'_+ = -1$. Then

$$\begin{array}{ll} y(0) = 0 & y_1(x) = x \\ y(4) = 2 & y_2(x) = -x + 6 \end{array}$$

At a corner $c = -c + 6$ so $c = 3$.

Case 2 Assume that $y'_1 = -1, y'_2 = 1$, so that

$$\begin{array}{ll} y(0) = 0 & y_1(x) = -x \\ y(4) = 2 & y_2(x) = x - 2 \end{array}$$

Hence we get that $-c = c - 2$ which implies that $c = 1$.

4.2 Optimal Control as an Extension of Variational Calculus

In our study of the calculus of variations, we were often presented with the task of minimizing the cost functional $J(y) = \int_a^b f(t, y, \dot{y})\ dt$ subject to the boundary conditoins $y(a) = y_a, y(b) = y_b$. Define $z_1 = t, z_2 = y, u = \dot{y}$ so that $\dot{z}_1 = 1, \dot{z}_2 = u$. Then we can view the equivalent formalism of minimizing

$$J(u) = \int_a^b f(z_1, z_2, u)\ dt \quad \text{subject to} \quad \begin{array}{ll} \dot{z}_1 = 1 & z_1(a) = a, z_a(b) = b \\ \dot{z}_2 = u & z_2(a) = y_a, z_2(b) = y_b \end{array}$$

The Pontryagin Hamiltonian is $h = f + p_1 + p_2 u$ and the optimality equations are

$$\dot{p}_1 = -f_{z_1}, \qquad \dot{p}_2 = -f_{z_2}, \qquad h_u = f_u + p_2$$

Assume that u is continuous so that

$$\frac{d}{dt} f_u + \dot{p}_2 = \frac{d}{dt} f_u - f_{z_2} = 0$$

which we notice is precisely the Euler-Lagrange equation.

It may turn out that recasting some variational calculus problems into an optimal control framework may simplify the problem of finding solutions.

Example:

Consider the task of crossing a river in minimal time. Let the initial and terminal points be $(0, 0)$ and $(\bar{y}_1, \bar{y}_2)$ respectively. The boat has a speed of 1, oriented at an angle of σ. The current is $0 < r(y_1) < 1$. Using the method of the calculus of variations, we can write the velocity as $v_x = \cos\sigma$ and $v_y = r(x) + \sin\sigma$. The cost functional is

$$T = \int_0^{x_1} \frac{dx}{v_x} = \int_0^{x_1} \frac{1}{\cos\sigma}\ dx$$

This is a rather involved and complicated problem in this framework. Let us now consider the equivalent optimal control problem. We want to find the time optimal control (hence minimizing $\int_0^T\ dt$) subject to the differential equation

$$\begin{aligned} v_x &= \dot{y}_1(t) = \cos\sigma \\ v_y &= \dot{y}_2(t) = r(y_1) + \sin\sigma \end{aligned}$$

Let $|\sigma| < \frac{\pi}{2}$. The Pontryagin Hamiltonian is

$$h = 1 + p_1 \cos\sigma + p_2(r(y_1) + \sin\sigma)$$

and the optimality conditions are given by

$$\dot{p}_1 = -p_2 r'(y_1), \qquad \dot{p}_2 = 0$$

By the PMP, u minimizes h so

$$h_\sigma = -p_1 \sin\sigma + p_2 \cos\sigma = 0$$

implies that $\sigma = \arctan\left(\frac{p_2}{p_1}\right)$.

For simplicity sake, assume that $r(y_1) = r$ a constant. Then both p_1 and p_2 are similarily constant which implies that $\tan\sigma$ is also constant. Hence the time optimal path is a straight line.

Example: Steering with Resistance

Consider the differential equation given by $\ddot{\omega} = -\epsilon\dot{\omega} + u$ or

$$\begin{pmatrix}\dot{y}_1\\ \dot{y}_2\end{pmatrix} = \begin{pmatrix}0 & 1\\ 0 & -\epsilon\end{pmatrix}\begin{pmatrix}y_1\\ y_2\end{pmatrix} + \begin{pmatrix}0\\ 1\end{pmatrix}u$$

We want to minimize the time required to steer from $y_1(0) > 0, y(2) = 0$ to $y_1(T) = y_2(T) = 0$ subject to the constraint $|u(t)| \leq 1$.

It is easy to see that this system is controllable, hence the control will be bang-bang. Furthermore, since the eigenvalues of the internal dynamics are $0, -\epsilon \in \mathbb{R}$ we know that there is at most one switch. We have been told that $y_1(0) > 0$ so we choose $u(0) = -1$. We can solve the differential equations to find that $y_1(t) = A + Be^{-\epsilon t} + Ct$. Using the initial conditions we can solve explicitly to find that

$$y_{-1}(t) = y_{10} + \frac{1 - \epsilon t - e^{-\epsilon t}}{\epsilon^2}$$
$$y_{-2}(t) = \frac{e^{-\epsilon t} - 1}{\epsilon}$$

Since $y_2(t) < 0$ there must be exactly one switch. At the final time T we will be aa the origin. We can solve the differential equatoins again for $u = 1$ to find

$$y_{1+}(t) = \frac{e^{\epsilon(T-t)} - 1 - \epsilon(T-t)}{\epsilon^2}$$
$$y_{2+}(t) = \frac{1 - e^{\epsilon(T-t)}}{\epsilon}$$

Example:

Consider the problem of a forced damped oscillator, modelled as follows

$$\ddot{z}(t) + 2\xi\omega\dot{z}(t) + \omega^2 z(t) = u(t)$$

for $\omega > 0, 0 \leq \xi \leq 1$. In first order form, we can write this as

$$\begin{pmatrix}\dot{y}_1\\ \dot{y}_2\end{pmatrix} = \underbrace{\begin{pmatrix}0 & 1\\ -\omega^2 & -2\xi\omega\end{pmatrix}}_{A}\begin{pmatrix}y_1\\ y_2\end{pmatrix} + \underbrace{\begin{pmatrix}0\\ 1\end{pmatrix}}_{B}u$$

The eigenvalues of A are $\lambda = \omega(-\xi \pm i\sqrt{1-\xi^2})$. We want to find u that minimizes

$$\int_0^\infty y^\dagger \underbrace{\begin{pmatrix}q_1 & 0\\ 0 & q_1\end{pmatrix}}_{Q} y + u^2\ dt = \int_0^\infty q_1 y_1^2 + q_2 y_2^2 + u^2\ dt$$

This is an infinite time linear quadratic control problem. The optimal control is givnen by

$$u^0(t) = -R^{-1}B^\dagger P = -B^\dagger P \qquad \text{since } R = 1$$

where P solves the Algebraic Riccati Equation

$$A^\dagger P + PA - PBR^{-1}B^\dagger P + Q = 0$$

$$\begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -\omega^2 & -2\xi\omega \end{pmatrix} + \begin{pmatrix} 0 & -\omega^2 \\ 1 & -2\xi\omega \end{pmatrix}\begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix} - \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix} + \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} -2\omega^2 p_2 - p_2^2 + q_1 & -\omega^2 p_3 + p_1 - 2\xi\omega p_2 - p_2 p_3 \\ -\omega^2 p_3 + p_1 - 2\xi\omega p_2 - p_2 p_3 & 2p_2 - 4\xi\omega p_3 - p_3^2 + q_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

This yields three equations for p_1, p_2, p_3. Solving yields

$$\begin{aligned} p_1 &= \omega^2 p_3 + 2\xi\omega p_2 + p_2 p_3 \\ p_2 &= -\omega^2 \pm \sqrt{\omega^4 + q_1} \\ p_3 &= -2\xi\omega \pm \sqrt{4\xi^2\omega^2 + 2p_2 + q_2} \end{aligned}$$

We have two possibilities for p_2 and p_3. However, we want $p \geq 0$ so that

$$\begin{pmatrix} y_1 & y_2 \end{pmatrix} P \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \geq 0$$

This will hold for $\begin{pmatrix} 0 & y_2 \end{pmatrix}$ which implies that $p_3 \geq 0$. Similarily, $p_2 \geq 0$.

The optimal control is

$$u^0(t) = -B^\dagger P y(t) = -p_2 z(t) - p_3 \dot{z}(t)$$

Note that

$$\begin{aligned} \dot{y}(t) &= Ay(t) + B(-Ky(t)) \\ &= \begin{pmatrix} 0 & 1 \\ -\sqrt{\omega^4 + q_1} & -\sqrt{4\xi^2\omega^2 + 2p_2 + q_2} \end{pmatrix} \end{aligned}$$

Hence we can think of $\sqrt{\omega^4 + q_1}$ as ω_c^2 of the control system, and $\sqrt{4\xi^2\omega^2 + 2p_2 + q_2}$ as $2\xi_c\omega_c$ of the control system.