The Calculus of Variations

Chase Foxen

Iowa State University

cefoxen@iastate.edu

April 5, 2024

Table of Contents

Definition

A **functional** is a function which assigns a real number to each function (or curve) belonging to a class of functions. That is, a function of functions.

Introduction

Definition

A **functional** is a function which assigns a real number to each function (or curve) belonging to a class of functions. That is, a function of functions.

Example

$S : C^1[a, b] \to \mathbb{R}$ defined by $S[f] = \int_a^b \sqrt{1 + f'(x)^2} \mathrm{d}x$

Introduction

Definition

A **functional** is a function which assigns a real number to each function (or curve) belonging to a class of functions. That is, a function of functions.

Example

$S : C^1[a, b] \to \mathbb{R}$ defined by $S[f] = \int_a^b \sqrt{1 + f'(x)^2}\mathrm{d}x$

Definition

Variational calculus is the study of finding minima and maxima of functionals using variations of the function (i.e., perturbations of the input functions).

Historical Notes

The calculus of variations begins with the *brachistochrone problem* in 1696. Johann Bernoulli proposed a problem asking to find the path of shortest time under gravity between two points.

This problem had previously been examined by Galileo Galilei in the 1630s and his solution was that the arc of a circle is the fastest route between two points.

Figure 1: Johann Bernoulli in 1740 [3]

Historical Notes

Later in the 1750s, Joseph-Louis Lagrange, under the mentorship of Leonhard Euler, would push for an analytical approach to variational problems, over the more popular geometric methods.

This new perspective would lead Lagrange and Euler to discover the Euler-Lagrange Equation, a fundamental result in the calculus of variations.

Figure 2: Joseph-Louis Lagrange [4]

Figure 3: Leonhard Euler [5]

Definition

A set $\mathbf{V}$ of elements $x, y, z, ...$ of any kind for which operations of addition and multiplication by real numbers $\alpha, \beta, ...$ are defined and obey the following axioms:

1. $x + y = y + x$ (Symmetry)
2. $(x + y) + z = x + (y + z)$ (Associativity)
3. $\exists 0 \in \mathbf{V} \ni x + 0 = x \;\; \forall x \in \mathbf{V}$ (Additive Identity)
4. $\forall x \in \mathbf{V}, \;\; \exists(-x) \in \mathbf{V} \;\; \ni x + (-x) = 0$ (Additive Inverse)
5. $\exists 1 \in \mathbf{V} \;\; \ni 1x = x$ (Multiplicative Identity)
6. $\alpha(\beta x) = (\alpha\beta)x$ (Associativity of Scalar Multiplication)
7. $(\alpha + \beta)x = \alpha x + \beta x$ (Distributivity I)
8. $\alpha(x + y) = \alpha x + \alpha y$ (Distributivity II)

For all $x, y, z \in \mathbf{V}$ and $\alpha, \beta \in \mathbb{R}$, is called a **linear space** or a **vector space**. [1]

Preliminaries

Example

The set of continuous functions over an interval $[a, b]$ is denoted $C[a, b]$ and is a linear space.

Example

The set of once continuously differentiable functions over an interval $[a, b]$ is denoted $C^1[a, b]$ and is a linear space.

We can take this up to any $n \in \mathbb{N}$ and we will have that the set of n-times continuously differentiable functions over an interval $[a, b]$ is a linear space. We call this $C^n[a, b]$.

Preliminaries

Definition

A linear space is said to be **normed** if each element $\mathbf{x} \in \mathbf{V}$ is assigned a nonnegative real number $\|\mathbf{x}\|$, called the norm of $\mathbf{x}$, such that

1. $\|\mathbf{x}\| = 0 \iff \mathbf{x} = 0$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for $\alpha \in \mathbb{R}$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Preliminaries

Definition

A linear space is said to be **normed** if each element $\mathbf{x} \in \mathbf{V}$ is assigned a nonnegative real number $\|\mathbf{x}\|$, called the norm of $\mathbf{x}$, such that

1. $\|\mathbf{x}\| = 0 \iff \mathbf{x} = 0$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for $\alpha \in \mathbb{R}$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Definition

In the space $C[a, b]$ we define the norm to be

$$\|\mathbf{y}\|_0 = \max_{x \in [a,b]} |y(x)|.$$

Preliminaries

Definition

In the space $C^1[a, b]$ we define the norm to be

$$\|\mathbf{y}\|_1 = \max_{x \in [a,b]} |y(x)| + \max_{x \in [a,b]} |y'(x)|.$$

Thus, two functions $y(x)$ and $z(x)$ in $C^1[a, b]$ are considered close together if both the functions themselves and their first derivatives are close together. So, $\|\mathbf{y} - \mathbf{z}\|_1 < \varepsilon$ implies $|y(x) - z(x)| < \varepsilon$ and $|y'(x) - z'(x)| < \varepsilon$ for all $x \in [a, b]$.

Definition

In the space $C^1[a, b]$ we define the norm to be

$$\|\mathbf{y}\|_1 = \max_{x \in [a,b]} |y(x)| + \max_{x \in [a,b]} |y'(x)|.$$

Thus, two functions $y(x)$ and $z(x)$ in $C^1[a, b]$ are considered close together if both the functions themselves and their first derivatives are close together. So, $\|\mathbf{y} - \mathbf{z}\|_1 < \varepsilon$ implies $|y(x) - z(x)| < \varepsilon$ and $|y'(x) - z'(x)| < \varepsilon$ for all $x \in [a, b]$.

Definition

In the space $C^n[a, b]$ we define the norm to be

$$\|\mathbf{y}\|_n = \sum_{i=0}^{\infty} \max_{x \in [a,b]} |y^{(i)}(x)|$$

Definition

The functional $J[\mathbf{y}]$ is said to be **continuous** at the point $\hat{\mathbf{y}} \in \mathbf{V}$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|J[\mathbf{y}] - J[\hat{\mathbf{y}}]| < \varepsilon,$$

when $\|\mathbf{y} - \hat{\mathbf{y}}\| < \delta$.

Seems very familiar, just instead of some $x \in \mathbb{R}$ we consider some function $\mathbf{y} \in \mathbf{V}$, where $\mathbf{V}$ is a function space.

Definition

The functional $J[\mathbf{y}]$ is said to be **continuous** at the point $\hat{\mathbf{y}} \in \mathbf{V}$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|J[\mathbf{y}] - J[\hat{\mathbf{y}}]| < \varepsilon,$$

when $\|\mathbf{y} - \hat{\mathbf{y}}\| < \delta$.

Seems very familiar, just instead of some $x \in \mathbb{R}$ we consider some function $\mathbf{y} \in \mathbf{V}$, where $\mathbf{V}$ is a function space.

Remark

Generally, $C[a, b]$ is inadequate for the study of variational problems. One of the basic types of functionals is of the form $J[y] = \int_a^b F(x, \mathbf{y}, \mathbf{y}')\mathrm{d}x$. So, we need $\mathbf{y}'$ to be close to $\hat{\mathbf{y}}'$ in addition to $\mathbf{y}$ and $\hat{\mathbf{y}}$ being close. [1]

Variational Calculus

Definition

Given a normed linear space **V**, let $\Phi[h]$ be a functional defined on **V**. Then $\Phi[h]$ is said to be a **continuous linear functional** if

1. $\Phi[\alpha h] = \alpha\Phi[h]\ \forall h \in \mathbf{V}$ and $\forall \alpha \in \mathbb{R}$
2. $\Phi[h_1 + h_2] = \Phi[h_1] + \Phi[h_2]$ for any $h_1, h_2 \in \mathbf{V}$
3. $\Phi[h]$ is continuous for all $h \in \mathbf{V}$

Variational Calculus

Definition

Given a normed linear space **V**, let $\Phi[h]$ be a functional defined on **V**. Then $\Phi[h]$ is said to be a **continuous linear functional** if

1. $\Phi[\alpha h] = \alpha\Phi[h]$ $\forall h \in \mathbf{V}$ and $\forall \alpha \in \mathbb{R}$
2. $\Phi[h_1 + h_2] = \Phi[h_1] + \Phi[h_2]$ for any $h_1, h_2 \in \mathbf{V}$
3. $\Phi[h]$ is continuous for all $h \in \mathbf{V}$

Example

The integral

$$\Phi[h] = \int_a^b \alpha(x)h(x)\mathrm{d}x,$$

where $\alpha(x)$ is a fixed function in $C[a, b]$, defines a linear functional in $C[a, b]$.

Example

(cont.) To see this take some scalars $\beta, \gamma \in \mathbb{R}$ and $h_1, h_2 \in \mathbf{V}$, and it follows

$$\Phi[\beta h_1 + \gamma h_2] = \int_a^b \alpha(x)\Big(\beta h_1(x) + \gamma h_2(x)\Big)\mathrm{d}x \tag{1}$$

$$= \int_a^b \beta\alpha(x)h_1(x) + \gamma\alpha(x)h_2(x)\mathrm{d}x \tag{2}$$

$$= \beta \int_a^b \alpha(x)h_1(x)\mathrm{d}x + \gamma \int_a^b \alpha(x)h_2(x)\mathrm{d}x \tag{3}$$

$$= \beta\Phi[h_1] + \gamma\Phi[h_2]. \tag{4}$$

Hence, it is a linear functional.

Lemma 1

If $\alpha(x)$ is continuous in $[a, b]$, and if

$$\int_a^b \alpha(x)h(x)\mathrm{d}x = 0$$

for every function $h(x) \in C[a, b]$ such that $h(a) = h(b) = 0$, then $\alpha(x) = 0$ for all $x \in [a, b]$.

This is often referred to as the "Fundamental Lemma of the Calculus of Variations".

Proof. (Lemma 1)
Suppose that the function $\alpha(x)$ is nonzero, WLOG say positive, at some point in the interval $[a, b]$. Then $\alpha(x)$ is also positive on some subinterval $[x_1, x_2]$ contained in $[a, b]$ since $\alpha(x)$ is continuous. If we set

$$h(x) = (x - x_1)(x - x_2)$$

for $x \in [x_1, x_2]$ and $h(x) = 0$ otherwise, then $h(x)$ obviously satisfies the conditions stated.

Proof. (Lemma 1)
Suppose that the function $\alpha(x)$ is nonzero, WLOG say positive, at some point in the interval $[a, b]$. Then $\alpha(x)$ is also positive on some subinterval $[x_1, x_2]$ contained in $[a, b]$ since $\alpha(x)$ is continuous. If we set

$$h(x) = (x - x_1)(x - x_2)$$

for $x \in [x_1, x_2]$ and $h(x) = 0$ otherwise, then $h(x)$ obviously satisfies the conditions stated. However,

$$\int_a^b \alpha(x)h(x)\mathrm{d}x = \int_{x_1}^{x_2} \alpha(x)(x - x_1)(x - x_2)\mathrm{d}x > 0,$$

since the integrand is positive (except at x_1 and x_2). Hence, a contradiction.

Variational Calculus

Proof.

Suppose that the function $\alpha(x)$ is nonzero, WLOG say positive, at some point in the interval $[a, b]$. Then $\alpha(x)$ is also positive on some subinterval $[x_1, x_2]$ contained in $[a, b]$ since $\alpha(x)$ is continuous. If we set

$$h(x) = (x - x_1)(x - x_2)$$

for $x \in [x_1, x_2]$ and $h(x) = 0$ otherwise, then $h(x)$ obviously satisfies the conditions stated. However,

$$\int_a^b \alpha(x)h(x)\mathrm{d}x = \int_{x_1}^{x_2} \alpha(x)(x - x_1)(x - x_2)\mathrm{d}x > 0,$$

since the integrand is positive (except at x_1 and x_2). Hence, a contradiction. [1] □

Lemma 2

If $\alpha(x)$ is continuous in $[a, b]$, and if

$$\int_a^b \alpha(x)h'(x)\mathrm{d}x = 0$$

for every function $h(x) \in C^1[a, b]$ such that $h(a) = h(b) = 0$, then $\alpha(x) = c$ for all $x \in [a, b]$, where c is a constant.

Proof. (Lemma 2)

Let c be the constant defined in the condition

$$\int_a^b [\alpha(x) - c]\mathrm{d}x = 0,$$

and let

$$h(x) = \int_a^x [\alpha(\xi) - c]\mathrm{d}\xi$$

so that $h(x)$ automatically belongs to $C^1[a, b]$ and satisfies the condition $h(a) = h(b) = 0$.

Proof. (Lemma 2)

Let c be the constant defined in the condition

$$\int_a^b [\alpha(x) - c]\mathrm{d}x = 0,$$

and let

$$h(x) = \int_a^x [\alpha(\xi) - c]\mathrm{d}\xi$$

so that $h(x)$ automatically belongs to $C^1[a, b]$ and satisfies the condition $h(a) = h(b) = 0$. Then, on the one hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b \alpha(x)h'(x)\mathrm{d}x - c[h(b) - h(a)] = 0,$$

while on the other hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b [\alpha(x) - c]^2\mathrm{d}x.$$

Proof. (Lemma 2) (cont.)
Then, on the one hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b \alpha(x)h'(x)\mathrm{d}x - c[h(b) - h(a)] = 0,$$

while on the other hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b [\alpha(x) - c]^2\mathrm{d}x.$$

It follows from Lemma 1 that $\alpha(x) - c = 0$, i.e., $\alpha(x) = c$ for all $x \in [a, b]$.

Proof.

Let c be the constant defined in the condition

$$\int_a^b [\alpha(x) - c]\mathrm{d}x = 0,$$

and let

$$h(x) = \int_a^x [\alpha(\xi) - c]\mathrm{d}\xi$$

so that $h(x)$ automatically belongs to $C^1[a, b]$ and satisfies the condition $h(a) = h(b) = 0$. Then, on the one hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b \alpha(x)h'(x)\mathrm{d}x - c[h(b) - h(a)] = 0,$$

while on the other hand,

$$\int_a^b [\alpha(x) - c]h'(x)\mathrm{d}x = \int_a^b [\alpha(x) - c]^2\mathrm{d}x.$$

It follows from Lemma 1 that $\alpha(x) - c = 0$, i.e., $\alpha(x) = c$ for all $x \in [a, b]$. [1] □

Lemma 3

If $\alpha(x)$ and $\beta(x)$ are continuous in $[a, b]$, and if

$$\int_a^b [\alpha(x)h(x) - \beta(x)h'(x)]\mathrm{d}x = 0$$

for every function $h(x) \in C^1[a, b]$ such that $h(a) = h(b) = 0$, then $\beta(x)$ is differentiable and $\beta'(x) = \alpha(x)$ for all $x \in [a, b]$.

This will come in handy later...

Proof. (Lemma 3)
Setting

$$A(x) = \int_a^x \alpha(\xi)\mathrm{d}\xi$$

and integrating by parts we find that

$$\int_a^b \alpha(x)h(x)\mathrm{d}x = -\int_a^b A(x)h'(x)\mathrm{d}x,$$

that is,

$$\int_a^b [\alpha(x)h(x) - \beta(x)h'(x)]\mathrm{d}x = 0$$

can be rewritten as

$$\int_a^b [-A(x) + \beta(x)]h'(x)\mathrm{d}x = 0.$$

Proof. (Lemma 3) (cont.)
But, according to Lemma 2, this implies

$$\beta(x) - A(x) = c,$$

where c is a constant. Hence, by the definition of $A(x)$,

$$\beta'(x) = \alpha(x),$$

for all $x \in [a, b]$, as asserted. Note that the differentiability $\beta(x)$ was not assumed beforehand.

Proof.

Setting

$$A(x) = \int_a^x \alpha(\xi)\mathrm{d}\xi$$

and integrating by parts we find that

$$\int_a^b \alpha(x)h(x)\mathrm{d}x = -\int_a^b A(x)h'(x)\mathrm{d}x,$$

that is,

$$\int_a^b [\alpha(x)h(x) - \beta(x)h'(x)]\mathrm{d}x = 0$$

can be rewritten as

$$\int_a^b [-A(x) + \beta(x)]h'(x)\mathrm{d}x = 0.$$

But, according to Lemma 2, this implies

$$\beta(x) - A(x) = c,$$

where c is a constant. Hence, by the definition of $A(x)$,

$$\beta'(x) = \alpha(x),$$

for all $x \in [a, b]$, as asserted. Note that the differentiability $\beta(x)$ was not assumed beforehand. [1] □

Variational Calculus

Definition

Let $J[y]$ be a functional defined on some normed linear space, and let

$$\Delta J[y; h] = J[y+h] - J[y]$$

be its **increment**, corresponding to the increment $h = h(x)$ of the "independent variable" $y = y(x)$.

Variational Calculus

Definition

Let $J[y]$ be a functional defined on some normed linear space, and let

$$\Delta J[y; h] = J[y + h] - J[y]$$

be its **increment**, corresponding to the increment $h = h(x)$ of the "independent variable" $y = y(x)$.

So, we just take some function h(x) and perturb y(x) by this amount. This will serve a similar purpose as the h in the difference quotient definition of the derivative.

Remark

If y is fixed, $\Delta J[y; h]$ is a functional of h. So, for a fixed y the increment can be denoted $\Delta J[h]$.

Definition

Suppose that

$$\Delta J[y;h] = \Phi[y;h] + \varepsilon \|h\|,$$

where $\Phi[y;h]$ is a linear functional and $\varepsilon \to 0$ as $\|h\| \to 0$. Then the functional $J[y]$ is said to be differentiable, and the principal linear part of the increment $\Delta J[y;h]$, the linear functional $\Phi[y;h]$, is called the **variation** or **differential** of $J[y]$ and is denoted by

$$\delta J[y].$$

Theorem

The variation of a differentiable functional is unique.

Proof. Theorem 1 First we note that if $\Phi[y; h]$ is a linear functional and if

$$\frac{\Phi[y; h]}{\|h\|} \to 0$$

as $\|h\| \to 0$, then $\Phi[y; h] \equiv 0$, i.e., $\Phi[y; h] = 0$ for all h.

Theorem

The variation of a differentiable functional is unique.

Proof. Theorem 1 First we note that if $\Phi[y;h]$ is a linear functional and if

$$\frac{\Phi[y;h]}{\|h\|} \to 0$$

as $\|h\| \to 0$, then $\Phi[y;h] \equiv 0$, i.e., $\Phi[y;h] = 0$ for all h. In fact, suppose $\Phi[y;h_0] \neq 0$ for some $h_0 \neq 0$. Then setting

$$h_n = \frac{h_0}{n} \ , \quad \lambda = \frac{\Phi[y;h_0]}{\|h_0\|},$$

we see that $\|h_n\| \to 0$ as $n \to \infty$, but

$$\lim_{n\to\infty} \frac{\Phi[y;h_n]}{\|h_n\|} = \lim_{n\to\infty} \frac{n\Phi[y;h_0]}{n\|h_0\|} = \lambda \neq 0,$$

contrary to the hypothesis.

Proof. Theorem 1 (cont.)

Now, suppose the differential of the functional $J[y]$ is not uniquely defined, so that

$$\Delta J[y;h] = \Phi_1[y;h] + \varepsilon_1 \|h\|,$$

$$\Delta J[y;h] = \Phi_2[y;h] + \varepsilon_2 \|h\|,$$

where $\Phi_1[y;h]$ and $\Phi_2[y;h]$ are linear functionals, and $\varepsilon_1, \varepsilon_2 \to 0$ as $\|h\| \to 0$.

Proof. Theorem 1 (cont.)
Now, suppose the differential of the functional $J[y]$ is not uniquely defined, so that

$$\Delta J[y; h] = \Phi_1[y; h] + \varepsilon_1 \|h\|,$$
$$\Delta J[y; h] = \Phi_2[y; h] + \varepsilon_2 \|h\|,$$

where $\Phi_1[y; h]$ and $\Phi_2[y; h]$ are linear functionals, and $\varepsilon_1, \varepsilon_2 \to 0$ as $\|h\| \to 0$. This implies

$$\Phi_1[y; h] - \Phi_2[y; h] = \varepsilon_1 \|h\| - \varepsilon_2 \|h\|,$$

and hence $\Phi_1[y; h] - \Phi_2[y; h]$ is an infinitesimal of order higher than 1 relative to $\|h\|$. But since $\Phi_1[y; h] - \Phi_2[y; h]$ is a linear functional it follows from the first part of the proof that $\Phi_1[y; h] - \Phi_2[y; h]$ vanishes identically, as asserted. [1] □

Variational Calculus

Let's break that last part down a bit...

Consider a function $f : \mathbb{R} \to \mathbb{R}$. We say $f(x) = \mathcal{O}(x)$ when $x \to 0$ if $\lim_{x\to 0} \frac{f(x)}{x} \leq C$, where $C \in \mathbb{R}$.

Variational Calculus

Let's break that last part down a bit...

Consider a function $f : \mathbb{R} \to \mathbb{R}$. We say $f(x) = \mathcal{O}(x)$ when $x \to 0$ if $\lim_{x\to 0} \frac{f(x)}{x} \leq C$, where $C \in \mathbb{R}$.
So, to get this to converge to zero we want the order to be higher than one (e.g., $\mathcal{O}(\|h\|^{\frac{3}{2}}), \mathcal{O}(\|h\|^2), \mathcal{O}(\|h\|^3)$).

Let's break that last part down a bit...

Consider a function $f : \mathbb{R} \to \mathbb{R}$. We say $f(x) = \mathcal{O}(x)$ when $x \to 0$ if $\lim_{x\to 0} \frac{f(x)}{x} \leq C$, where $C \in \mathbb{R}$.

So, to get this to converge to zero we want the order to be higher than one (e.g., $\mathcal{O}(\|h\|^{\frac{3}{2}}), \mathcal{O}(\|h\|^2), \mathcal{O}(\|h\|^3)$).

For $\Phi_1[y; h] - \Phi_2[y; h]$ to be an infinitesimal of order higher than one relative to $\|h\|$ tells us that

$$\lim_{\|h\|\to 0} \frac{\Phi_1[y; h] - \Phi_2[y; h]}{\|h\|} = \varepsilon_1 - \varepsilon_2.$$

We know $\varepsilon_1, \varepsilon_2 \to 0$ as $\|h\| \to 0$, hence

$$\Phi_1[y; h] - \Phi_2[y; h] = 0$$

as desired.

Variational Calculus

We are interested in maxima and minima of functionals, so let's define what it means for a functional to have an extremum.

Definition

We say that the functional $J[y]$ has a **relative extremum** for $y = \hat{y}$ if $J[y] - J[\hat{y}]$ does not change its sign for some neighborhood of the curve $y = \hat{y}(x)$.

We are interested in maxima and minima of functionals, so let's define what it means for a functional to have an extremum.

Definition

We say that the functional $J[y]$ has a **relative extremum** for $y = \hat{y}$ if $J[y] - J[\hat{y}]$ does not change its sign for some neighborhood of the curve $y = \hat{y}(x)$.

Definition

We say that the functional $J[y]$ has a **weak extremum** for $y = \hat{y}$ if there exists $\varepsilon > 0$ such that $J[y] - J[\hat{y}]$ has the same sign for all y in the domain of the definition of the functional which satisfy $\|y - \hat{y}\|_1 < \varepsilon$, where $\|\cdot\|_1$ denotes the norm in the space C^1.

Definition

We say that the functional $J[y]$ has a **strong extremum** for $y = \hat{y}$ if there exists $\varepsilon > 0$ such that $J[y] - J[\hat{y}]$ has the same sign for all y in the domain of the definition of the functional which satisfy $\|y - \hat{y}\|_0 < \varepsilon$, where $\| \cdot \|_0$ denotes the norm in the space C.

Theorem

A necessary condition for the differentiable functional $J[y]$ to have an extremum for $y = \hat{y}$ is that

$$\delta J[y; h] = 0$$

for $y = \hat{y}$ and all admissible h.

Proof. (Theorem 2)
According to the definition of the variation $\delta J[y]$, we have

$$\Delta J[\hat{y}; h] = \delta J[\hat{y}; h] + \varepsilon \|h\|,$$

where $\varepsilon \to 0$ as $\|h\| \to 0$.

Proof. (Theorem 2)
According to the definition of the variation $\delta J[y]$, we have

$$\Delta J[\hat{y}; h] = \delta J[\hat{y}; h] + \varepsilon \|h\|,$$

where $\varepsilon \to 0$ as $\|h\| \to 0$.Thus, for significantly small $\|h\|$, the sign of $\Delta J[\hat{y}; h]$ will be the same as $\delta J[\hat{y}; h]$. Now, suppose that $\delta J[\hat{y}; h_0] \neq 0$ for some admissible h_0. Then, for any $\alpha > 0$, no matter how small, we have

$$\delta J[\hat{y}; -\alpha h_0] = -\delta J[\hat{y}; \alpha h_0].$$

Proof. (Theorem 2)
According to the definition of the variation $\delta J[y]$, we have

$$\Delta J[\hat{y}; h] = \delta J[\hat{y}; h] + \varepsilon \|h\|,$$

where $\varepsilon \to 0$ as $\|h\| \to 0$.Thus, for significantly small $\|h\|$, the sign of $\Delta J[\hat{y}; h]$ will be the same as $\delta J[\hat{y}; h]$. Now, suppose that $\delta J[\hat{y}; h_0] \neq 0$ for some admissible h_0. Then, for any $\alpha > 0$, no matter how small, we have

$$\delta J[\hat{y}; -\alpha h_0] = -\delta J[\hat{y}; \alpha h_0].$$

Hence,

$$\Delta J[\hat{y}; h] = \delta J[\hat{y}; h] + \varepsilon ||h||$$

can be made to have either sign for arbitrarily small $\|h\|$. But this is impossible since by hypothesis $J[y]$ has a minimum for $y = \hat{y}$, i.e.,

$$\Delta J[\hat{y}; h] = J[\hat{y} + h] - J[\hat{y}] \geq 0$$

for all sufficiently small $\|h\|$. This contradiction proves the theorem. □

Problem

Let $F(x, y, z)$ be a function with continuous first and second partial derivatives with respect to all of its arguments. Then, among all functions $y(x)$ which satisfy the boundary conditions

$$y(a) = A \ , \ \ y(b) = B,$$

find the function for which the functional

$$J[y] = \int_a^b F(x, y, y')\mathrm{d}x$$

has a weak extremum. [1]

Solution

Suppose we give $y(x)$ an increment $h(x)$, where, in order for the function

$$\bar{y} = y(x) + h(x)$$

to continue to satisfy the boundary conditions, we require $h(a) = h(b) = 0$.

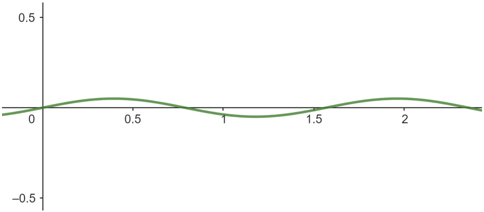

Figure 4: For example h(x)=0.05sin(4x)

The Euler-Lagrange Equation

Solution (cont.)

Then, since the corresponding increment of the functional above equals

$$\Delta J = J[y+h] - J[y] = \int_a^b F(x, y+h, y'+h) - F(x, y, y')\mathrm{d}x,$$

it follows by using Taylor's Theorem that

$$\Delta J = \int_a^b \Big(\frac{\partial F}{\partial y}(x, y, y')h + \frac{\partial F}{\partial y'}(x, y, y')h'\Big)\mathrm{d}x + \cdots$$

where the dots denote terms of order higher than one relative to h and h'.

Solution (cont.)

The integral above represents the principal linear part of the increment ΔJ, and hence, the variation of $J[y]$ is

$$\delta J = \int_a^b \Big(\frac{\partial F}{\partial y}(x, y, y')h + \frac{\partial F}{\partial y'}(x, y, y')h' \Big) \mathrm{d}x.$$

Solution (cont.)

The integral above represents the principal linear part of the increment ΔJ, and hence, the variation of $J[y]$ is

$$\delta J = \int_a^b \Big(\frac{\partial F}{\partial y}(x, y, y')h + \frac{\partial F}{\partial y'}(x, y, y')h'\Big)\mathrm{d}x.$$

According to Theorem 2, a necessary condition for $J[y]$ to have an extremum for $y = y(x)$ is that

$$\delta J = \int_a^b \Big(\frac{\partial F}{\partial y}h + \frac{\partial F}{\partial y'}h'\Big)\mathrm{d}x = 0$$

for all admissible h.

The Euler-Lagrange Equation

Solution (cont.)

But, according to Lemma 3, this implies that

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y'} = 0,$$

a result known as the **Euler-Lagrange Equation**.

Solution (cont.)

But, according to Lemma 3, this implies that

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y'} = 0,$$

a result known as the **Euler-Lagrange Equation**.

Recall...

Lemma 3

If $\alpha(x)$ and $\beta(x)$ are continuous in $[a, b]$, and if

$$\int_a^b [\alpha(x)h(x) - \beta(x)h'(x)]\mathrm{d}x = 0$$

for every function $h(x) \in C^1[a, b]$ such that $h(a) = h(b) = 0$, then $\beta(x)$ is differentiable and $\beta'(x) = \alpha(x)$ for all $x \in [a, b]$.

The Euler-Lagrange Equation

From the previous solution we have proven the next theorem. [1]

Theorem

Let $J[y]$ be a functional of the form

$$J[y] = \int_a^b F(x, y, y')\mathrm{d}x,$$

where $y(x) \in C^1[a, b]$ such that

$$y(a) = A \ , \ \ y(b) = B.$$

Then a necessary condition for $J[y]$ to have an extremum for a given function $y(x)$ is that $y(x)$ satisfy the Euler-Lagrange Equation

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y'} = 0.$$

The Brachistochrone

Question...
What is the fastest route a ball starting at rest can take from point a to point b when friction is neglected?

We set up our axes so that y points down, as this will be more convenient later.

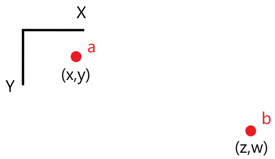

Figure 5: Point a and Point b

Let's apply some physics to this problem. We start with the conservation of energy (this is legal because, without friction, energy is conserved).

Define $a = (x_1, y_1)$ and $b = (x_2, y_2)$. So,

$$KE - PE = 0$$

The Brachistochrone

Let's apply some physics to this problem. We start with the conservation of energy (this is legal because, without friction, energy is conserved).

Define $a = (x_1, y_1)$ and $b = (x_2, y_2)$. So,

$$KE - PE = 0$$
$$\frac{1}{2}mv^2 - mgy = 0$$
$$\frac{1}{2}v^2 = gy$$

The Brachistochrone

Let's apply some physics to this problem. We start with the conservation of energy (this is legal because, without friction, energy is conserved).

Define $a = (x_1, y_1)$ and $b = (x_2, y_2)$. So,

$$
\begin{aligned}
KE - PE &= 0 \\
\frac{1}{2}mv^2 - mgy &= 0 \\
\frac{1}{2}v^2 &= gy \\
v &= \sqrt{2gy}
\end{aligned}
$$

Where $g = 9.81\frac{m}{s^2}$ and $y = (y_2 - y_1)m$.

The Brachistochrone

Since velocity is simply distance over time, we find

$$time(a \to b) = \int_a^b \frac{\mathrm{d}s}{v}.$$

The Brachistochrone

Since velocity is simply distance over time, we find

$$time(a \to b) = \int_a^b \frac{\mathrm{d}s}{v}.$$

The infinitesimal distance $\mathrm{d}s$ is defined as

$$\mathrm{d}s = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \sqrt{\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + 1}\,\mathrm{d}x.$$

Since velocity is simply distance over time, we find

$$time(a \to b) = \int_a^b \frac{\mathrm{d}s}{v}.$$

The infinitesimal distance $\mathrm{d}s$ is defined as

$$\mathrm{d}s = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \sqrt{\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + 1}\,\mathrm{d}x.$$

Hence,

$$time(a \to b) = \frac{1}{\sqrt{2g}} \int_0^{y_2} \frac{\sqrt{y'(x)^2 + 1}}{\sqrt{y}} \mathrm{d}x.$$

(Where for simplicity we center our frame of reference at a so that $a = (0, 0)$.)

Therefore we obtain the functional

$$T[y, y'] = \int_0^{y_2} \frac{\sqrt{y'(x)^2 + 1}}{\sqrt{y}} \mathrm{d}x.$$

Now, to minimize this functional we want to apply the Euler-Lagrange Equation

$$\frac{\partial T}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial T}{\partial y'} = 0.$$

The Brachistochrone

Therefore we obtain the functional

$$T[y, y'] = \int_0^{y_2} \frac{\sqrt{y'(x)^2 + 1}}{\sqrt{y}} \mathrm{d}x.$$

Now, to minimize this functional we want to apply the Euler-Lagrange Equation

$$\frac{\partial T}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial T}{\partial y'} = 0.$$

We note that the functional is only a functional of y and y'. So we opt to use the first integral of the Euler-Lagrange Equation, which is valid for functionals dependent solely on y and y'.

The Brachistochrone

Therefore we obtain the functional

$$T[y, y'] = \int_0^{y_2} \frac{\sqrt{y'(x)^2 + 1}}{\sqrt{y}} \mathrm{d}x.$$

Now, to minimize this functional we want to apply the Euler-Lagrange Equation

$$\frac{\partial T}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial T}{\partial y'} = 0.$$

We note that the functional is only a functional of y and y'. So we opt to use the first integral of the Euler-Lagrange Equation, which is valid for functionals dependent solely on y and y'. We take the equality

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial T}{\partial y'} - T\right) &= y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial T}{\partial y'}\right) - \frac{\partial T}{\partial x} - \frac{\partial T}{\partial y}y' \\ &= -y'\left(\frac{\partial T}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial T}{\partial y'}\right) - \frac{\partial T}{\partial x},\end{aligned}$$

The Brachistochrone

Since $T[y, y']$ is independent of x we know $\frac{\partial T}{\partial x} = 0$. Also, since we are searching for the case where the Euler-Lagrange Equation is satisfied, the first term goes to zero and we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial T}{\partial y'} - T\right) = 0,$$

The Brachistochrone

Since $T[y, y']$ is independent of x we know $\frac{\partial T}{\partial x} = 0$. Also, since we are searching for the case where the Euler-Lagrange Equation is satisfied, the first term goes to zero and we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial T}{\partial y'} - T\right) = 0,$$

that is,

$$y'\frac{\partial T}{\partial y'} - T = C,$$

for some $C \in \mathbb{R}$ [2].

The Brachistochrone

We find

$$\frac{\partial T}{\partial y'} = \frac{y'}{\sqrt{y}\sqrt{y'^2 + 1}}$$

The Brachistochrone

We find

$$\frac{\partial T}{\partial y'} = \frac{y'}{\sqrt{y}\sqrt{y'^2+1}}$$

So by the previous result we see

$$y'\frac{\partial T}{\partial y'} - T = \frac{y'^2}{\sqrt{y}\sqrt{y'^2+1}} - \frac{\sqrt{y'(x)^2+1}}{\sqrt{y}} = C.$$

The Brachistochrone

We find

$$\frac{\partial T}{\partial y'} = \frac{y'}{\sqrt{y}\sqrt{y'^2+1}}$$

So by the previous result we see

$$y'\frac{\partial T}{\partial y'} - T = \frac{y'^2}{\sqrt{y}\sqrt{y'^2+1}} - \frac{\sqrt{y'(x)^2+1}}{\sqrt{y}} = C.$$

We set this constant $C = \frac{1}{\sqrt{2a}}$ where $a \in \mathbb{R}$, solve for y', and integrate to obtain

$$x = \int \frac{\sqrt{y}}{\sqrt{2a-y}} \mathrm{d}y.$$

To solve this we use the substitution

$$y = 2a\sin^2\left(\frac{\theta}{2}\right)$$

and see

$$x = 2a\int \sin^2\left(\frac{\theta}{2}\right)\mathrm{d}\theta = a(\theta - \sin(\theta))$$

Hence, we have obtained the parameterized solution to the Euler-Lagrange Equation [2]

$$x = a(\theta - \sin(\theta))\ ,\ y = a\Big(1 - \cos\theta\Big).$$

This parameterization describes the curve of the **cycloid**, the fastest route from point a to b under gravity (without friction). This is the arc traced by a point on a rolling circle, which is not the arc of a circle! So, Galileo was close, but not quite there.

Figure 6: The Cycloid

References

[1] Gelfand, I. M., & Fomin, S. V. (1963). *Calculus of Variations*. Dover Publications.

[2] Weinstock, R. (1951). *Calculus of Variations with Applications to Physics & Engineering*. Dover Publications.

[3] Johann Bernoulli. (2024, March 15). In *Wikipedia*. https://en.wikipedia.org/wiki/Johann_Bernoulli

[4] Joseph-Louis Lagrange. (2024, March 20). In *Wikipedia*. https://en.wikipedia.org/wiki/Joseph-Louis_Lagrange

[5] Leonhard Euler. (2024, March 27). In *Wikipedia*. https://en.wikipedia.org/wiki/Leonhard_Euler