Introduction to the Calculus of Variations

Jim Fischer

March 20, 1999

Abstract

This is a self-contained paper which introduces a fundamental problem in the calculus of variations, the problem of finding extreme values of *functionals*. The reader should have a solid background in one-variable calculus.

Contents

1 Introduction **1**

2 Partial Derivatives **2**

3 The Chain Rule **3**

4 Statement of the Problem **4**

5 The Euler-Lagrange Equation **5**

6 The Brachistochrone Problem **10**

7 Concluding Remarks **12**

1 Introduction

We begin with an introduction to partial differentiation of functions of several variables. After partial derivatives are introduced we discuss some forms of the chain rule. In section 4 we formulate the fundamental problem. In section 5 we state and prove the fundamental result (The Euler-Lagrange Equation). We conclude the paper with a solution of one of the most famous problems from the calculus of variations, *The Brachistochrone Problem.*

2 Partial Derivatives

Given a function of one variable say $f(x)$, we define the derivative of $f(x)$ at $x = a$ to be

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

provided this limit exists. For a function of several variables the *total* derivative of a function is not as easy to define, however, if we set all but one of the independent variables equal to a constant, we can define the *partial* derivative of a function by using a limit similar to the one above. For example, if f is a function of the variables x, y and z, we can set $x = a$ and $y = b$ and define the partial derivative of f with respect to z to be

$$\frac{\partial f}{\partial z}(x, y, z) = \lim_{h \to 0} \frac{f(a, b, z+h) - f(a, b, z)}{h},$$

wherever this limit exists. Note that the partial derivative is a function of all three variables x, y and z. The partial derivative of f with respect to z gives the instantaneous rate of change of f in the z direction. The definition for the partial derivative of f with respect to x or y is defined in a similar way. Computing partial derivatives is no harder than computing ordinary one-variable derivatives, one simply treats the fixed variables as constants.

Example 1. *Suppose* $f(x, y, z) = x^2 y^2 z^2 + y\cos(z)$*, then*

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y, z) &= 2xy^2z^2 \\
\frac{\partial f}{\partial y}(x, y, z) &= 2x^2yz^2 + \cos(z) \\
\frac{\partial f}{\partial z}(x, y, z) &= 2x^2y^2z - y\sin(z)
\end{aligned}$$

We can take higher order partial derivatives by continuing in the same manner. In the example above, first taking the partial derivative of f with respect to y and then with respect to z yields:

$$\frac{\partial^2 f}{\partial z \partial y} = 4x^2yz - \sin(z).$$

Such a derivative is called a *mixed partial derivative of* f. From a well known theorem of advanced calculus, if the second order partial derivatives of f exist in a neighborhood of the point (a, b, c) and are continuous at

(a, b, c), then the mixed partial derivatives do not depend on the order in which they are derived. That is, for example

$$\frac{\partial^2 f}{\partial z \partial y}(a, b, c) = \frac{\partial^2 f}{\partial y \partial z}(a, b, c).$$

This result was first proved by Leonard Euler in 1734.

Exercise 1. *Let* $f(x, y, z) = \sqrt{\dfrac{1 + y^2}{z^2}}$, *compute all three partial derivatives of* f.

Exercise 2. *For* f *in Example 1, compute a couple of mixed partial derivatives and verify that the order in which you differentiate does not matter.*

3 The Chain Rule

We begin with a review of the chain rule for functions of one variable. Suppose $f(x)$ is a differentiable function of x and $x = x(t)$ is differentiable of t. By the chain rule theorem, the composite function $z(t) = f \circ x(t)$ is a differentiable function of t and

$$\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt}. \tag{3.1}$$

For example, if $f(x) = \sin(x)$ and $x(t) = t^2$, then the derivative with respect to t of $z = \sin(t^2)$ is given by $\cos(t^2) \cdot 2t$. It turns out that there is a chain rule for functions of several variables. For example, suppose x and y are functions of t and consider the function $z = [x(t)]^2 + 3[y(t)]^3$. We can think of z as the composite of the function $f(x, y) = x^2 + 3y^3 with the functions\ x(t)$ and $y(t)$. By a chain rule theorem for functions of several variables,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \tag{3.2}$$

Note the similarity between (3.1) and (3.2). For functions of several variables, one needs to keep track of each of the independent variables separately, applying a chain rule to each. The hypothesis for the chain rule theorem require the function $z = f(x, y)$ to have continuous partial derivatives and for $x(t)$ and $y(t)$ to be differentiable.

4 Statement of the Problem

We begin with a simple example. Let P and Q be two points in the xy-plane and consider the collection of all smooth curves which connect P to Q. Let $y(x)$ be such a curve with $P = (a, y(a))$ and $Q = (b, y(b))$. The arc-length of the curve $y(x)$ is given by the integral

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

Suppose now that we wish to determine which curve will minimize the above integral. Certainly our knowledge of ordinary geometry suggests that the curve which minimizes the arc-length is the straight line connecting P to Q. However, what if instead we were interested in finding which curve minimizes a different integral? For example, consider the integral

$$\int_a^b \sqrt{\frac{1 + [y'(x)]^2}{y(x)}}\, dx.$$

It is not obvious what choice of $y(x)$ will result in minimizing this integral. Further, it is not at all obvious that such a minimum exists!

One way to proceed is to notice that the above integrals can be viewed as special kinds of functions, functions whose inputs are functions and whose outputs are real numbers. For example we could write

$$F[y(x)] = \int_a^b \sqrt{1 + [y'(x)]^2}\, dx$$

More generally we could write:

$$F[y(x)] = \int_a^b f(x, y(x), y'(x))\, dx$$

A function like F is actually called a **functional**, this name is used to distinguish F from ordinary real-valued functions whose domains consist of ordinary variables. The function f in the integral is to be viewed as an ordinary function of the variables x, y and y' (this should become more clear in the next section).[1] One of the fundamental problems of which the calculus of variations is concerned, is locating the extrema of functionals.

[1] We don't call $f(x, y, y')$ a functional because its range is not $\mathbb{R}$.

Before we formally state the problem, we need to specify the domain of F more precisely. Consider the interval $[a, b] \subset \mathbb{R}$ and define $C^1_{[a,b]}$ to be the set

$$C^1_{[a,b]} = \{y(x) |\ y : [a, b] \mapsto \mathbb{R},\ x \text{ has a continuous first derivative on [a,b]}\}.$$

We will consider only functionals which have certain desirable properties. Let F be a functional whose domain is $C^1_{[a,b]}$,

$$F[y(x)] = \int_a^b f(x, y(x), y'(x))\, dx.$$

We will require that the function f in the integral have continuous partial derivatives of x, y and y'. We require the continuity of derivatives because we will need to apply chain rules and the Leibniz rule for differentiation. We now state the fundamental problem.

> **Problem:** Let F be a functional defined on $C^2_{[a,b]}$, with $F[y(x)]$ given by
>
> $$F[y(x)] = \int_a^b f(x, y(x), y'(x))\, dx.$$
>
> Suppose the functional F obtains a minimum (or maximum) value[2]. How do we determine the curve $y(x)$ which produces such a minimum (maximum) value for F?

In the next section we will show that the minimizing curve $y(x)$ must satisfy a differential equation known as the *Euler-Lagrange Equation.*

5 The Euler-Lagrange Equation

We begin this section with the fundamental result:

Theorem 1. *If $y(x)$ is a curve in $C^2_{[a,b]}$ which minimizes the functional*

$$F[y(x)] = \int_a^b f(x, y(x), y'(x))\, dx,$$

then the following differential equation must be satisfied:

$$\frac{\partial f}{\partial x} - \frac{d}{dx}\left(\frac{\partial f}{\partial x'}\right) = 0.$$

This equation is called the Euler-Lagrange Equation.

[2]This is an important assumption for there do exist Functionals which have no extrema.

Before proving this theorem, we consider an example.

Example 2. *If* $F[y(x)] = \int_a^b \sqrt{1+[y'(x)]^2}\,dx$, *then the Euler-Lagrange Equation is given by:*

$$\begin{aligned}
0 &= \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \\
&= 0 - \frac{d}{dx}\left(\frac{y'(x)}{\sqrt{1+[y'(x)]^2}}\right) \\
&= -\frac{\sqrt{1+[y'(x)]^2}y''(x) - [y'(x)]^2 y''(x)\left(1+[y'(x)]^2\right)^{-\frac{1}{2}}}{1+[y'(x)]^2} \\
&= -\frac{\left(1+[y'(x)]^2\right)y''(x) - [y'(x)]^2 y''(x)}{(1+[y'(x)])^{\frac{3}{2}}} \\
&= -\frac{y''(x)}{(1+[y'(x)])^{\frac{3}{2}}}
\end{aligned}$$

Exercise 3. *Show that the solution to*

$$0 = \frac{y''(x)}{(1+[y'(x)])^{\frac{3}{2}}}$$

is a straight line. That is $y(x) = Ax + B$. *Is this a proof that the shortest path between two points is a straight line?*

The proof of Theorem 1 relies on three things, the Leibniz rule, integration by parts and Lemma 1. It is assumed that the reader is familiar with integration by parts, we will discuss the Leibniz rule later, and we state and prove Lemma 1 now.

Lemma 1. *Let* $M(x)$ *be a continuous function on the interval* $[a,b]$. *Suppose that for any continuous function* $h(x)$ *with* $h(a) = h(b) = 0$ *we have*

$$\int_a^b M(x)h(x)\,dx = 0.$$

Then $M(x)$ *is identically zero*[3] *on* $[a,b]$.

[3]Actually the function is zero almost everywhere. This means that the set of x values where the function is not zero has a length of zero.

Proof of Lemma 1: Since $h(x)$ can be any continuous function with $h(a) = h(b) = 0$, we choose $h(x) = -M(x)(x-a)(x-b)$. Clearly $h(x)$ is continuous since M is continuous. Also, $M(x)h(x) \geq 0$ on $[a,b]$ (check this). But, if the definite integral of a non-negative function is zero then the function itself must be zero. So we conclude that

$$\begin{aligned} 0 &= M(x)h(x) \\ &= [M(x)]^2[-(x-a)(x-b)]. \end{aligned}$$

This and the fact that $[-(x-a)(x-b)] > 0$ on (a,b) implies that $[M(x)]^2 = 0$ on $[a,b]$. Finally, $[M(x)]^2 = 0$ on $[a,b]$ implies that $M(x) = 0$ on $[a,b]$. □

Proof of Theorem 1: Suppose $y(x)$ is a curve which minimizes the functional F. That is, for any other permissible curve $g(x)$, $F[y(x)] \leq F[g(x)]$. The basic idea in this proof will be to construct a function of one real variable say $H(\epsilon)$ which has the following properties:

1. $H(\epsilon)$ is a differentiable function near $\epsilon = 0$.
2. $H(0)$ is a local minimum for H.

After constructing H, we show that Property 2 implies the Euler-Lagrange equation must be satisfied.

We begin by constructing a variation of $y(x)$. Let ϵ be a small real number (positive or negative), and consider the new function given by:

$$y_\epsilon(x) = y(x) + \epsilon h(x)$$

where $h(x) \in C^2_{[a,b]}$ and $h(a) = h(b) = 0$.

We can now define the function H to be

$$H(\epsilon) = F[y_\epsilon(x)].$$

Since $x_0(t) = y(x)$ and $y(x)$ minimizes $F[y(x)]$, it follows that 0 minimizes $H(\epsilon)$. Now, since $H(0)$ is a minimum value for H, we know from ordinary calculus that $H'(0) = 0$. The function H can be differentiated by using the Leibniz rule[4]:

$$\begin{aligned} \frac{d}{d\epsilon}\left(H(\epsilon)\right) &= \frac{d}{d\epsilon}\int_a^b f(x, y_\epsilon, y'_\epsilon)\,dx \\ &= \int_a^b \frac{\partial}{\partial\epsilon} f(x, y_\epsilon, y'_\epsilon)\,dx \end{aligned}$$

[4]For a proof of the Leibniz rule, check out a text on advanced calculus.

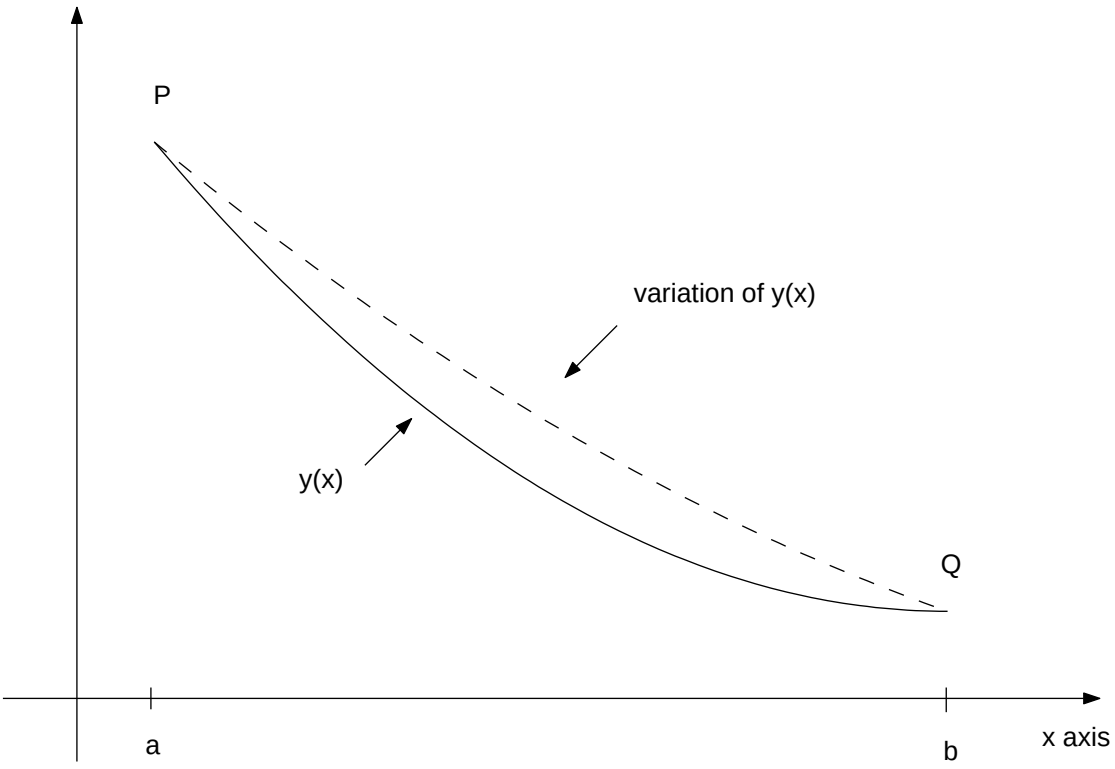

Figure 1: A variation of $y(x)$.

Applying the chain rule within the integral we obtain:

$$\frac{\partial}{\partial \epsilon} f(x, y_\epsilon, y'_\epsilon) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \epsilon} + \frac{\partial f}{\partial y_\epsilon}\frac{\partial y_\epsilon}{\partial \epsilon} + \frac{\partial f}{\partial y'_\epsilon}\frac{\partial x'_\epsilon}{\partial \epsilon} \tag{5.1}$$

$$= \frac{\partial f}{\partial y_\epsilon}\frac{\partial y_\epsilon}{\partial \epsilon} + \frac{\partial f}{\partial y'_\epsilon}\frac{\partial x'_\epsilon}{\partial \epsilon} \tag{5.2}$$

$$= \frac{\partial f}{\partial y_\epsilon} h(x) + \frac{\partial f}{\partial y'_\epsilon} h'(x) \tag{5.3}$$

Exercise 4. *Show that equations* (5.1) *through* (5.3) *are true.*

From these computations, we have

$$H'(\epsilon) = \int_a^b \left(\frac{\partial f}{\partial y_\epsilon} h(x) + \frac{\partial f}{\partial y'_\epsilon} h'(x) \right) dx. \tag{5.4}$$

Evaluating this equation at $\epsilon = 0$ yields

$$0 = \int_a^b \left(\frac{\partial f}{\partial y} h(x) + \frac{\partial f}{\partial y'} h'(x) \right) dx. \tag{5.5}$$

At this point we would like to apply Lemma 1 but in order to do so, we must first apply integration by parts to the second term in the above integral. Once this is done, the following equation is obtained from equation (5.5):

$$0 = \int_a^b \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right] h(x)\, dx. \tag{5.6}$$

Finally, since this procedure works for any function $h(x)$ with $h(a) = h(b) = 0$, we can apply Lemma 1 and conclude that

$$0 = \frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right).$$

This completes the proof of Theorem 1. □

Exercise 5. *Verify equation* (5.6) *by doing the integration by parts in equation* (5.5).

Beltrami Identity

Often in applications, the function f which appears in the integrand does not depend directly on the variable x. In these situations, the Euler-Lagrange equation takes a particularly nice form. This simplification of the Euler-Lagrange equation is known as the *Beltrami Identity*. We present without proof the Beltrami Identity, it is not obvious how it arises from the Euler-Lagrange equation, however, its derivation is straight-forward.

The Beltrami Identity: If $\frac{\partial f}{\partial x} = 0$ then the Euler-Lagrange equation is equivalent to:

$$f - y' \frac{\partial f}{\partial y'} = C \tag{5.7}$$

where C is a constant.

Exercise 6. *Use the Beltrami identity to produce the differential equation in Example 2.*

6 The Brachistochrone Problem

Suppose P and Q are two points in the plane. Imagine there is a thin, flexible wire connecting the two points. Suppose P is above Q, and we let a frictionless bead travel down the wire impelled by gravity alone. By changing the shape of the wire we might alter the amount of time it takes for the bead to travel from P to Q. The brachistochrone problem (or quickest descent problem) is concerned with determining what shape (if any) will result in the bead reaching the point Q in the least amount of time. This problem was first introduced by J. Bernoulli in the mid 17th century, and was first solved by Isaac Newton. In this section we set up the relevant functional and then apply Theorem 1 to see what the differential equation associated with this problem looks like. Finally, we provide a solution to this differential equation.

First, we let a curve $y(x)$ that connects P and Q represent the wire. As before assume $P = (a, y(a))$ and $Q = (b, y(b))$. We will restrict ourselves to curves that belong to $C^2_{[a,b]}$. Given such a curve, the time it takes for the bead to go from P to Q is given by the functional[5]

$$F[y(x)] = \int_a^b \frac{ds}{v} \tag{6.1}$$

where $ds = \sqrt{1 + [y'(x)]^2}dx$ and $v(x) = y'(x)$. By using Newton's second law (Potential and Kinetic energies are equal) we obtain

$$\frac{1}{2}m[v(x)]^2 = mg(y(a) - y(x))$$

This allows us to rewrite the functional (6.1) as

$$F[y(x)] = \int_a^b \sqrt{\frac{1 + [y'(x)]^2}{2g(y(a) - y(x)}}\, dx$$

Assuming a minimum time exists, we can apply Theorem 1 to the functional F. Notice that the integrand does not depend directly on the variable x and therefore we can apply the Beltrami Identity. We can also make computations a little easier letting $P = (0, 0)$, the resulting equation is then

$$\sqrt{\frac{1 + [y'(x)]^2}{2gy(x)}} - \frac{1}{2}y'(x)\left(\sqrt{\frac{2gy(x)}{1 + [y'(x)]^2}}\right)\frac{y'(x)}{gy(x)} = C \tag{6.2}$$

[5]For convenience we use the down direction to represent positive y values.

Equation (6.2) simplifies to

$$\left(1 + [y'(x)]^2\right) y(x) = \frac{1}{2gC^2} = k^2 \tag{6.3}$$

Finally, equation (6.3) is well known and the solution is a cycloid.[6] The parametric equations of the cycloid are given by:

$$x(\theta) = \frac{1}{2}k^2 (\theta - \sin\theta)$$
$$y(\theta) = \frac{1}{2}k^2 (1 - \cos\theta)$$

Example 3. *With* $P = (0,0)$ *and* $Q = (\frac{\pi}{2}, -1)$,$k = 1$ *and* $0 \leq \theta \leq \frac{\pi}{2}$. *Figure 2 shows the cycloid solution to the Brachistochrone problem.*

A strange property of the cycloid is the followig: If we let the frictionless bead start from rest at any point on the cycloid, the amount of time it takes to reach point Q is always the same.

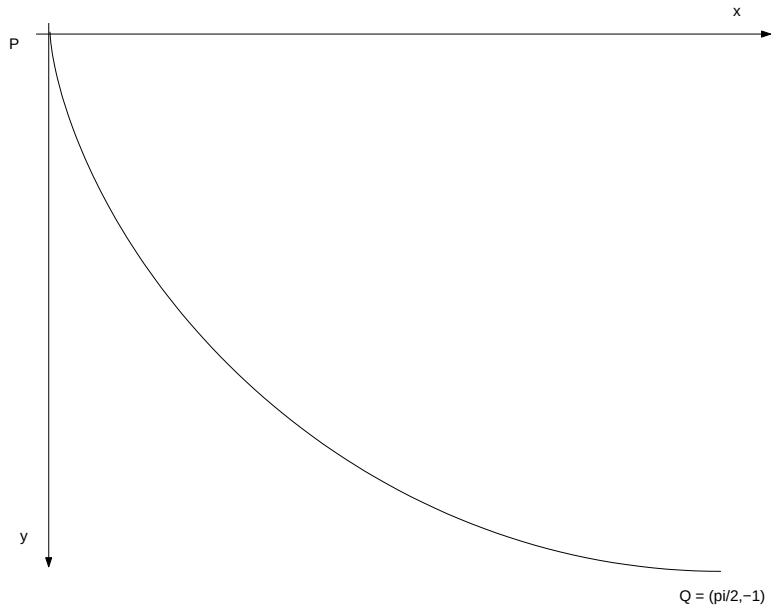

Figure 2: The Cycloid Solution for Example 3.

[6]The Cycloid curve is the path of a fixed point on the rim of a wheel as the wheel is rotated along a straight line.

7 Concluding Remarks

After introducing the notion of partial derivatives and the chain rule for functions of several variables, we were able to state a problem that the calculus of variations is concerned. This is the problem of identifying extreme values for functionals (which are functions of functions). In section 5 we showed that under the assumption that a minimum (or maximum) solution exists, the solution must satisfy the Euler-Lagrange equation. The analysis which lead to this equation relied heavily on the fact from ordinary calculus that the derivative of a function at an extreme value is zero (provided that the derivative exists there).

We admit that much of the analytical details have been omitted and encourage the interested reader to look further into these matters. It turns out that problems like the brachistochrone can be extended to situations with an arbitrary number of variables as well as to regions which are non-Euclidean. In fact, many books about General Relativity deduce the Einstein field equations via a variational approach which is based on the ideas that were discussed in this paper.

References

[1] Widder, D., *Advanced Calculus*, Prentice-Hall, 1961.

[2] Boyer, C., *A History of Mathematics*, John Wiley and Sons, 1991.

[3] Troutman, J., *Variational Calculus with Elementary Convexity*, Springer-Verlag, 1980.