Introduction to the Calculus of Variations

by Peter J. Olver
University of Minnesota

1. Introduction.

Minimization principles form one of the most wide-ranging means of formulating mathematical models governing the equilibrium configurations of physical systems. Moreover, many popular numerical integration schemes such as the powerful finite element method are also founded upon a minimization paradigm. In these notes, we will develop the basic mathematical analysis of nonlinear minimization principles on infinite-dimensional function spaces — a subject known as the "calculus of variations", for reasons that will be explained as soon as we present the basic ideas. Classical solutions to minimization problems in the calculus of variations are prescribed by boundary value problems involving certain types of differential equations, known as the associated Euler–Lagrange equations. The mathematical techniques that have been developed to handle such optimization problems are fundamental in many areas of mathematics, physics, engineering, and other applications. In this chapter, we will only have room to scratch the surface of this wide ranging and lively area of both classical and contemporary research.

The history of the calculus of variations is tightly interwoven with the history of mathematics, [**9**]. The field has drawn the attention of a remarkable range of mathematical luminaries, beginning with Newton and Leibniz, then initiated as a subject in its own right by the Bernoulli brothers Jakob and Johann. The first major developments appeared in the work of Euler, Lagrange, and Laplace. In the nineteenth century, Hamilton, Jacobi, Dirichlet, and Hilbert are but a few of the outstanding contributors. In modern times, the calculus of variations has continued to occupy center stage, witnessing major theoretical advances, along with wide-ranging applications in physics, engineering and all branches of mathematics.

Minimization problems that can be analyzed by the calculus of variations serve to characterize the equilibrium configurations of almost all continuous physical systems, ranging through elasticity, solid and fluid mechanics, electro-magnetism, gravitation, quantum mechanics, string theory, and many, many others. Many geometrical configurations, such as minimal surfaces, can be conveniently formulated as optimization problems. Moreover, numerical approximations to the equilibrium solutions of such boundary value problems are based on a nonlinear finite element approach that reduces the infinite-dimensional minimization problem to a finite-dimensional problem. See [**14**; Chapter 11] for full details.

Just as the vanishing of the gradient of a function of several variables singles out the critical points, among which are the minima, both local and global, so a similar "functional gradient" will distinguish the candidate functions that might be minimizers of the functional. The finite-dimensional calculus leads to a system of algebraic equations for the

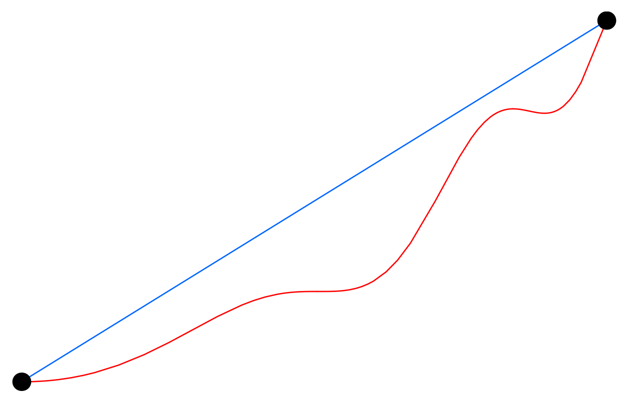

Figure 1. The Shortest Path is a Straight Line.

critical points; the infinite-dimensional functional analog results a boundary value problem for a nonlinear ordinary or partial differential equation whose solutions are the critical functions for the variational problem. So, the passage from finite to infinite dimensional nonlinear systems mirrors the transition from linear algebraic systems to boundary value problems.

2. Examples of Variational Problems.

The best way to appreciate the calculus of variations is by introducing a few concrete examples of both mathematical and practical importance. Some of these minimization problems played a key role in the historical development of the subject. And they still serve as an excellent means of learning its basic constructions.

Minimal Curves, Optics, and Geodesics

The *minimal curve problem* is to find the shortest path between two specified locations. In its simplest manifestation, we are given two distinct points

$$\mathbf{a} = (a, \alpha) \qquad \text{and} \qquad \mathbf{b} = (b, \beta) \qquad \text{in the plane} \quad \mathbb{R}^2, \tag{2.1}$$

and our task is to find the curve of shortest length connecting them. "Obviously", as you learn in childhood, the shortest route between two points is a straight line; see Figure 1. Mathematically, then, the minimizing curve should be the graph of the particular affine function†

$$y = c\,x + d = \frac{\beta - \alpha}{b - a}\,(x - a) + \alpha \tag{2.2}$$

that passes through or interpolates the two points. However, this commonly accepted "fact" — that (2.2) is the solution to the minimization problem — is, upon closer inspection, perhaps not so immediately obvious from a rigorous mathematical standpoint.

Let us see how we might formulate the minimal curve problem in a mathematically precise way. For simplicity, we assume that the minimal curve is given as the graph of

† We assume that $a \neq b$, i.e., the points $\mathbf{a}, \mathbf{b}$ do not lie on a common vertical line.

a smooth function $y = u(x)$. Then, the length of the curve is given by the standard arc length integral

$$J[u] = \int_a^b \sqrt{1 + u'(x)^2}\, dx, \tag{2.3}$$

where we abbreviate $u' = du/dx$. The function $u(x)$ is required to satisfy the boundary conditions

$$u(a) = \alpha, \qquad\qquad u(b) = \beta, \tag{2.4}$$

in order that its graph pass through the two prescribed points (2.1). The minimal curve problem asks us to find the function $y = u(x)$ that minimizes the arc length functional (2.3) among all "reasonable" functions satisfying the prescribed boundary conditions. The reader might pause to meditate on whether it is analytically obvious that the affine function (2.2) is the one that minimizes the arc length integral (2.3) subject to the given boundary conditions. One of the motivating tasks of the calculus of variations, then, is to rigorously prove that our everyday intuition is indeed correct.

Indeed, the word "reasonable" *is* important. For the arc length functional (2.3) to be defined, the function $u(x)$ should be at least piecewise C^1, i.e., continuous with a piecewise continuous derivative. Indeed, if we were to allow discontinuous functions, then the straight line (2.2) does not, in most cases, give the minimizer. Moreover, continuous functions which are not piecewise C^1 need not have a well-defined arc length. The more seriously one thinks about these issues, the less evident the "obvious" solution becomes. But before you get too worried, rest assured that the straight line (2.2) is indeed the true minimizer. However, a fully rigorous proof of this fact requires a careful development of the mathematical machinery of the calculus of variations.

A closely related problem arises in geometrical optics. The underlying physical principle, first formulated by the seventeenth century French mathematician Pierre de Fermat, is that, when a light ray moves through an optical medium, it travels along a path that minimizes the travel time. As always, Nature seeks the most economical† solution. In an inhomogeneous planar optical medium, the speed of light, $c(x, y)$, varies from point to point, depending on the optical properties. Speed equals the time derivative of distance traveled, namely, the arc length of the curve $y = u(x)$ traced by the light ray. Thus,

$$c(x, u(x)) = \frac{ds}{dt} = \sqrt{1 + u'(x)^2}\ \frac{dx}{dt}\ .$$

Integrating from start to finish, we conclude that the total travel time along the curve is equal to

$$T[u] = \int_0^T dt = \int_a^b \frac{dt}{dx}\, dx = \int_a^b \frac{\sqrt{1 + u'(x)^2}}{c(x, u(x))}\, dx. \tag{2.5}$$

Fermat's Principle states that, to get from point $\mathbf{a} = (a.\alpha)$ to point $\mathbf{b} = (b, \beta)$, the light ray follows the curve $y = u(x)$ that minimizes this functional subject to the boundary conditions

$$u(a) = \alpha, \qquad\qquad u(b) = \beta,$$

† Assuming time = money!

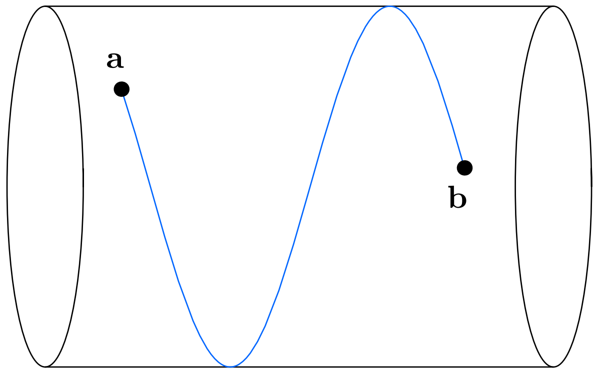

Figure 2. Geodesics on a Cylinder.

If the medium is homogeneous, e.g., a vacuum‡, then $c(x,y) \equiv c$ is constant, and $T[u]$ is a multiple of the arc length functional (2.3), whose minimizers are the "obvious" straight lines traced by the light rays. In an inhomogeneous medium, the path taken by the light ray is no longer evident, and we are in need of a systematic method for solving the minimization problem. Indeed, all of the known laws of geometric optics, lens design, focusing, refraction, aberrations, etc., will be consequences of the geometric and analytic properties of solutions to Fermat's minimization principle, [**3**].

Another minimization problem of a similar ilk is to construct the *geodesics* on a curved surface, meaning the curves of minimal length. Given two points $\mathbf{a}, \mathbf{b}$ lying on a surface $S \subset \mathbb{R}^3$, we seek the curve $C \subset S$ that joins them and has the minimal possible length. For example, if S is a circular cylinder, then there are three possible types of geodesic curves: straight line segments parallel to the center line; arcs of circles orthogonal to the center line; and spiral helices, the latter illustrated in Figure 2. Similarly, the geodesics on a sphere are arcs of great circles. In aeronautics, to minimize distance flown, airplanes follow geodesic circumpolar paths around the globe. However, both of these claims are in need of mathematical justification.

In order to mathematically formulate the geodesic minimization problem, we suppose, for simplicity, that our surface $S \subset \mathbb{R}^3$ is realized as the graph† of a function $z = F(x,y)$. We seek the geodesic curve $C \subset S$ that joins the given points

$$\mathbf{a} = (a, \alpha, F(a,\alpha)), \qquad \text{and} \qquad \mathbf{b} = (b, \beta, F(b,\beta)), \quad \text{lying on the surface} \quad S.$$

Let us assume that C can be parametrized by the x coordinate, in the form

$$y = u(x), \qquad\qquad z = v(x) = F(x, u(x)),$$

‡ In the absence of gravitational effects due to general relativity.

† Cylinders are not graphs, but can be placed within this framework by passing to cylindrical coordinates. Similarly, spherical surfaces are best treated in spherical coordinates. In differential geometry, [**6**], one extends these constructions to arbitrary parametrized surfaces and higher dimensional manifolds.

where the last equation ensures that it lies in the surface S. In particular, this requires $a \neq b$. The length of the curve is supplied by the standard three-dimensional arc length integral. Thus, to find the geodesics, we must minimize the functional

$$\begin{aligned} J[u] &= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}\, dx \\ &= \int_a^b \sqrt{1 + \left(\frac{du}{dx}\right)^2 + \left(\frac{\partial F}{\partial x}(x, u(x)) + \frac{\partial F}{\partial u}(x, u(x))\, \frac{du}{dx}\right)^2}\, dx, \end{aligned} \tag{2.6}$$

subject to the boundary conditions $u(a) = \alpha$, $u(b) = \beta$. For example, geodesics on the paraboloid

$$z = \tfrac{1}{2}\, x^2 + \tfrac{1}{2}\, y^2 \tag{2.7}$$

can be found by minimizing the functional

$$J[u] = \int_a^b \sqrt{1 + (u')^2 + (x + u\, u')^2}\, dx. \tag{2.8}$$

Minimal Surfaces

The minimal surface problem is a natural generalization of the minimal curve or geodesic problem. In its simplest manifestation, we are given a simple closed curve $C \subset \mathbb{R}^3$. The problem is to find the surface of least total area among all those whose boundary is the curve C. Thus, we seek to minimize the surface area integral

$$\text{area } S = \iint_S dS$$

over all possible surfaces $S \subset \mathbb{R}^3$ with the prescribed boundary curve $\partial S = C$. Such an area–minimizing surface is known as a *minimal surface* for short. For example, if C is a closed plane curve, e.g., a circle, then the minimal surface will just be the planar region it encloses. But, if the curve C twists into the third dimension, then the shape of the minimizing surface is by no means evident.

Physically, if we bend a wire in the shape of the curve C and then dip it into soapy water, the surface tension forces in the resulting soap film will cause it to minimize surface area, and hence be a minimal surface†. Soap films and bubbles have been the source of much fascination, physical, æsthetical and mathematical, over the centuries, [**9**]. The minimal surface problem is also known as *Plateau's Problem*, named after the nineteenth century French physicist Joseph Plateau who conducted systematic experiments on such soap films. A satisfactory mathematical solution to even the simplest version of the minimal surface problem was only achieved in the mid twentieth century, [**11**, **12**]. Minimal surfaces and

† More accurately, the soap film will realize a local but not necessarily global minimum for the surface area functional. Non-uniqueness of local minimizers can be realized in the physical experiment — the same wire may support more than one stable soap film.

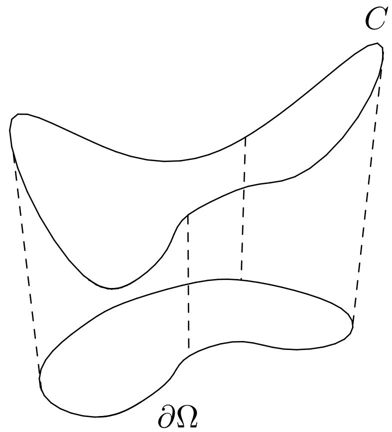

Figure 3. Minimal Surface.

related variational problems remain an active area of contemporary research, and are of importance in engineering design, architecture, and biology, including foams, domes, cell membranes, and so on.

Let us mathematically formulate the search for a minimal surface as a problem in the calculus of variations. For simplicity, we shall assume that the bounding curve C projects down to a simple closed curve $\Gamma = \partial\Omega$ that bounds an open domain $\Omega \subset \mathbb{R}^2$ in the (x, y) plane, as in Figure 3. The space curve $C \subset \mathbb{R}^3$ is then given by $z = g(x, y)$ for $(x, y) \in \Gamma = \partial\Omega$. For "reasonable" boundary curves C, we expect that the minimal surface S will be described as the graph of a function $z = u(x, y)$ parametrized by $(x, y) \in \Omega$. According to the basic calculus, the surface area of such a graph is given by the double integral

$$J[u] = \iint_\Omega \sqrt{1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}\; dx\, dy. \tag{2.9}$$

To find the minimal surface, then, we seek the function $z = u(x, y)$ that minimizes the surface area integral (2.9) when subject to the Dirichlet boundary conditions

$$u(x, y) = g(x, y) \qquad \text{for} \qquad (x, y) \in \partial\Omega. \tag{2.10}$$

As we will see, (5.10), the solutions to this minimization problem satisfy a complicated nonlinear second order partial differential equation.

A simple version of the minimal surface problem, that still contains some interesting features, is to find minimal surfaces with rotational symmetry. A *surface of revolution* is obtained by revolving a plane curve about an axis, which, for definiteness, we take to be the x axis. Thus, given two points $\mathbf{a} = (a, \alpha)$, $\mathbf{b} = (b, \beta) \in \mathbb{R}^2$, the goal is to find the curve $y = u(x)$ joining them such that the surface of revolution obtained by revolving the curve around the x-axis has the least surface area. Each cross-section of the resulting surface is a circle centered on the x axis. The area of such a surface of revolution is given by

$$J[u] = \int_a^b 2\pi\,|\,u\,|\,\sqrt{1 + (u')^2}\; dx. \tag{2.11}$$

We seek a minimizer of this integral among all functions $u(x)$ that satisfy the fixed boundary conditions $u(a) = \alpha$, $u(b) = \beta$. The minimal surface of revolution can be physically realized by stretching a soap film between two circular wires, of respective radius α and β, that are held a distance $b - a$ apart. Symmetry considerations will require the minimizing surface to be rotationally symmetric. Interestingly, the revolutionary surface area functional (2.11) is exactly the same as the optical functional (2.5) when the light speed at a point is inversely proportional to its distance from the horizontal axis: $c(x, y) = 1/(2\,\pi\,|\,y\,|)$.

Isoperimetric Problems and Constraints

The simplest *isoperimetric problem* is to construct the simple closed plane curve of a fixed length ℓ that encloses the domain of largest area. In other words, we seek to maximize

$$\text{area}\ \Omega = \iint_{\Omega} dx\,dy \qquad \text{subject to the constraint} \qquad \text{length}\ \partial\Omega = \oint_{\partial\Omega} ds = \ell,$$

over all possible domains $\Omega \subset \mathbb{R}^2$. Of course, the "obvious" solution to this problem is that the curve must be a circle whose perimeter is ℓ, whence the name "isoperimetric". Note that the problem, as stated, does not have a unique solution, since if Ω is a maximizing domain, any translated or rotated version of Ω will also maximize area subject to the length constraint.

To make progress on the isoperimetric problem, let us assume that the boundary curve is parametrized by its arc length, so $\mathbf{x}(s) = (\,x(s), y(s)\,)$ with $0 \leq s \leq \ell$, subject to the requirement that

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1. \tag{2.12}$$

We can compute the area of the domain by a line integral around its boundary,

$$\text{area}\ \Omega = \oint_{\partial\Omega} x\,dy = \int_0^{\ell} x\,\frac{dy}{ds}\,ds, \tag{2.13}$$

and thus we seek to maximize the latter integral subject to the arc length constraint (2.12). We also impose periodic boundary conditions

$$x(0) = x(\ell), \qquad\qquad y(0) = y(\ell), \tag{2.14}$$

that guarantee that the curve $\mathbf{x}(s)$ closes up. (Technically, we should also make sure that $\mathbf{x}(s) \neq \mathbf{x}(s')$ for any $0 \leq s < s' < \ell$, ensuring that the curve does not cross itself.)

A simpler isoperimetric problem, but one with a less evident solution, is the following. Among all curves of length ℓ in the upper half plane that connect two points $(-a, 0)$ and $(a, 0)$, find the one that, along with the interval $[\,-a, a\,]$, encloses the region having the largest area. Of course, we must take $\ell \geq 2a$, as otherwise the curve will be too short to connect the points. In this case, we assume the curve is represented by the graph of a non-negative function $y = u(x)$, and we seek to maximize the functional

$$\int_{-a}^{a} u\,dx \qquad \text{subject to the constraint} \qquad \int_{-a}^{a} \sqrt{1 + u'^2}\,dx = \ell. \tag{2.15}$$

In the previous formulation (2.12), the arc length constraint was imposed at every point, whereas here it is manifested as an integral constraint. Both types of constraints, pointwise and integral, appear in a wide range of applied and geometrical problems. Such constrained variational problems can profitably be viewed as function space versions of constrained optimization problems. Thus, not surprisingly, their analytical solution will require the introduction of suitable Lagrange multipliers.

3. The Euler–Lagrange Equation.

Even the preceding limited collection of examples of variational problems should already convince the reader of the tremendous practical utility of the calculus of variations. Let us now discuss the most basic analytical techniques for solving such minimization problems. We will exclusively deal with classical techniques, leaving more modern direct methods — the function space equivalent of gradient descent and related methods — to a more in–depth treatment of the subject, [**5**].

Let us concentrate on the simplest class of variational problems, in which the unknown is a continuously differentiable scalar function, and the functional to be minimized depends upon at most its first derivative. The basic minimization problem, then, is to determine a suitable function $y = u(x) \in \mathrm{C}^1[a, b]$ that minimizes the *objective functional*

$$J[u] = \int_a^b L(x, u, u')\, dx. \tag{3.1}$$

The integrand is known as the *Lagrangian* for the variational problem, in honor of Lagrange. We usually assume that the Lagrangian $L(x, u, p)$ is a reasonably smooth function of all three of its (scalar) arguments x, u, and p, which represents the derivative u'. For example, the arc length functional (2.3) has Lagrangian function $L(x, u, p) = \sqrt{1 + p^2}$, whereas in the surface of revolution problem (2.11), $L(x, u, p) = 2\pi \,|\, u \,|\sqrt{1 + p^2}$. (In the latter case, the points where $u = 0$ are slightly problematic, since L is not continuously differentiable there.)

In order to uniquely specify a minimizing function, we must impose suitable boundary conditions. All of the usual suspects — Dirichlet (fixed), Neumann (free), as well as mixed and periodic boundary conditions — are also relevant here. In the interests of brevity, we shall concentrate on the Dirichlet boundary conditions

$$u(a) = \alpha, \qquad\qquad u(b) = \beta, \tag{3.2}$$

although some of the exercises will investigate other types of boundary conditions.

The First Variation

The (local) minimizers of a (sufficiently nice) objective function defined on a finite-dimensional vector space are initially characterized as critical points, where the objective function's gradient vanishes. An analogous construction applies in the infinite-dimensional context treated by the calculus of variations. Every sufficiently nice minimizer of a sufficiently nice functional $J[u]$ is a "critical function", Of course, not every critical point turns out to be a minimum — maxima, saddles, and many degenerate points are also critical.

The characterization of nondegenerate critical points as local minima or maxima relies on the second derivative test, whose functional version, known as the second variation, will be is the topic of the following Section 4.

But we are getting ahead of ourselves. The first order of business is to learn how to compute the gradient of a functional defined on an infinite-dimensional function space. The general definition of the gradient requires that we first impose an inner product $\langle\, u\,;v\,\rangle$ on the underlying function space. The gradient $\nabla J[u]$ of the functional (3.1) will then be defined by the same basic directional derivative formula:

$$\langle\, \nabla J[u]\,;v\,\rangle = \frac{d}{d\varepsilon}\, J[u+\varepsilon\, v]\,\Big|_{\varepsilon=0}. \tag{3.3}$$

Here $v(x)$ is a function that prescribes the "direction" in which the derivative is computed. Classically, v is known as a *variation* in the function u, sometimes written $v=\delta u$, whence the term "calculus of variations". Similarly, the gradient operator on functionals is often referred to as the *variational derivative*, and often written δJ. The inner product used in (3.3) is usually taken (again for simplicity) to be the standard L^2 inner product

$$\langle\, f\,;g\,\rangle = \int_a^b f(x)\, g(x)\, dx \tag{3.4}$$

on function space. Indeed, while the formula for the gradient will depend upon the underlying inner product, the characterization of critical points does not, and so the choice of inner product is not significant here.

Now, starting with (3.1), for each fixed u and v, we must compute the derivative of the function

$$h(\varepsilon) = J[u+\varepsilon\, v] = \int_a^b L(x, u+\varepsilon\, v, u' + \varepsilon\, v'\,)\, dx. \tag{3.5}$$

Assuming sufficient smoothness of the integrand allows us to bring the derivative inside the integral and so, by the chain rule,

$$\begin{aligned} h'(\varepsilon) = \frac{d}{d\varepsilon}\, J[u+\varepsilon\, v] &= \int_a^b \frac{d}{d\varepsilon}\, L(x, u+\varepsilon\, v, u' + \varepsilon\, v'\,)\, dx \\ &= \int_a^b \left[\, v\, \frac{\partial L}{\partial u}\,(x, u+\varepsilon\, v, u' + \varepsilon\, v') + v'\, \frac{\partial L}{\partial p}\,(x, u+\varepsilon\, v, u' + \varepsilon\, v')\,\right] dx. \end{aligned}$$

Therefore, setting $\varepsilon = 0$ in order to evaluate (3.3), we find

$$\langle\, \nabla J[u]\,;v\,\rangle = \int_a^b \left[\, v\, \frac{\partial L}{\partial u}\,(x,u,u') + v'\, \frac{\partial L}{\partial p}\,(x,u,u')\,\right] dx. \tag{3.6}$$

The resulting integral often referred to as the *first variation* of the functional $J[u]$. The condition

$$\langle\, \nabla J[u]\,;v\,\rangle = 0$$

for a minimizer is known as the *weak form* of the variational principle.

To obtain an explicit formula for $\nabla J[u]$, the right hand side of (3.6) needs to be written as an inner product,

$$\langle \nabla J[u]\,; v \rangle = \int_a^b \nabla J[u]\, v \ dx = \int_a^b h\, v\, dx$$

between some function $h(x) = \nabla J[u]$ and the variation v. The first summand has this form, but the derivative v' appearing in the second summand is problematic. However, one can easily move derivatives around inside an integral through integration by parts. If we set

$$r(x) \equiv \frac{\partial L}{\partial p}\,(x, u(x), u'(x)),$$

we can rewrite the offending term as

$$\int_a^b r(x)\, v'(x)\, dx = \big[\, r(b)\, v(b) - r(a)\, v(a)\, \big] - \int_a^b r'(x)\, v(x)\, dx, \tag{3.7}$$

where, again by the chain rule,

$$r'(x) = \frac{d}{dx}\left(\frac{\partial L}{\partial p}\,(x,u,u') \right) = \frac{\partial^2 L}{\partial x\, \partial p}\,(x,u,u') + u'\, \frac{\partial^2 L}{\partial u\, \partial p}\,(x,u,u') + u''\, \frac{\partial^2 L}{\partial p^2}\,(x,u,u')\,. \tag{3.8}$$

So far we have not imposed any conditions on our variation $v(x)$. We are only comparing the values of $J[u]$ among functions that satisfy the prescribed boundary conditions, namely

$$u(a) = \alpha, \qquad\qquad u(b) = \beta.$$

Therefore, we must make sure that the varied function

$$\widehat{u}(x) = u(x) + \varepsilon\, v(x)$$

remains within this set of functions, and so

$$\widehat{u}(a) = u(a) + \varepsilon\, v(a) = \alpha, \qquad\qquad \widehat{u}(b) = u(b) + \varepsilon\, v(b) = \beta.$$

For this to hold, the variation $v(x)$ must satisfy the corresponding homogeneous boundary conditions

$$v(a) = 0, \qquad\qquad v(b) = 0. \tag{3.9}$$

As a result, both boundary terms in our integration by parts formula (3.7) vanish, and we can write (3.6) as

$$\langle \nabla J[u]\,; v \rangle = \int_a^b \nabla J[u]\, v \ dx = \int_a^b v \left[\frac{\partial L}{\partial u}\,(x,u,u') - \frac{d}{dx}\left(\frac{\partial L}{\partial p}\,(x,u,u') \right) \right] dx.$$

Since this holds for all variations $v(x)$, we conclude that

$$\nabla J[u] = \frac{\partial L}{\partial u}\,(x,u,u') - \frac{d}{dx}\left(\frac{\partial L}{\partial p}\,(x,u,u') \right). \tag{3.10}$$

This is our explicit formula for the functional gradient or variational derivative of the functional (3.1) with Lagrangian $L(x,u,p)$. Observe that the gradient $\nabla J[u]$ of a functional is a *function*.

The *critical functions* $u(x)$ are, by definition, those for which the functional gradient vanishes: satisfy

$$\nabla J[u] = \frac{\partial L}{\partial u}\,(x,u,u') - \frac{d}{dx}\,\frac{\partial L}{\partial p}\,(x,u,u') = 0. \tag{3.11}$$

In view of (3.8), the critical equation (3.11) is, in fact, a second order ordinary differential equation,

$$E(x,u,u',u'') = \frac{\partial L}{\partial u}\,(x,u,u') - \frac{\partial^2 L}{\partial x\,\partial p}\,(x,u,u') - u'\,\frac{\partial^2 L}{\partial u\,\partial p}\,(x,u,u') - u''\,\frac{\partial^2 L}{\partial p^2}\,(x,u,u') = 0, \tag{3.12}$$

known as the *Euler–Lagrange equation* associated with the variational problem (3.1), in honor of two of the most important contributors to the subject. Any solution to the Euler–Lagrange equation that is subject to the assumed boundary conditions forms a critical point for the functional, and hence is a potential candidate for the desired minimizing function. And, in many cases, the Euler–Lagrange equation suffices to characterize the minimizer without further ado.

Theorem 3.1. *Suppose the Lagrangian function is at least twice continuously differentiable:* $L(x,u,p) \in \mathrm{C}^2$*. Then any* C^2 *minimizer* $u(x)$ *to the corresponding functional* $J[u] = \int_a^b L(x,u,u')\,dx$*, subject to the selected boundary conditions, must satisfy the associated Euler–Lagrange equation* (3.11)*.*

Let us now investigate what the Euler–Lagrange equation tells us about the examples of variational problems presented at the beginning of this section. One word of caution: there do exist seemingly reasonable functionals whose minimizers are not, in fact, C^2, and hence do not solve the Euler–Lagrange equation in the classical sense; see [**2**] for examples. Fortunately, in most variational problems that arise in real-world applications, such pathologies do not appear.

Curves of Shortest Length — Planar Geodesics

Let us return to the most elementary problem in the calculus of variations: finding the curve of shortest length connecting two points $\mathbf{a} = (a,\alpha)$, $\mathbf{b} = (b,\beta) \in \mathbb{R}^2$ in the plane. As we noted in Section 3, such planar geodesics minimize the arc length integral

$$J[u] = \int_a^b \sqrt{1+(u')^2}\;dx \qquad \text{with Lagrangian} \qquad L(x,u,p) = \sqrt{1+p^2}\,,$$

subject to the boundary conditions

$$u(a) = \alpha, \qquad\qquad u(b) = \beta.$$

Since

$$\frac{\partial L}{\partial u} = 0, \qquad \frac{\partial L}{\partial p} = \frac{p}{\sqrt{1+p^2}}\,,$$

the Euler–Lagrange equation (3.11) in this case takes the form

$$0 \;=\; -\,\frac{d}{dx}\,\frac{u'}{\sqrt{1+(u')^2}} \;=\; -\,\frac{u''}{(1+(u')^2)^{3/2}}.$$

Since the denominator does not vanish, this is the same as the simplest second order ordinary differential equation

$$u'' = 0. \tag{3.13}$$

We deduce that the solutions to the Euler–Lagrange equation are all affine functions, $u = c\,x + d$, whose graphs are straight lines. Since our solution must also satisfy the boundary conditions, the only critical function — and hence the sole candidate for a minimizer — is the straight line

$$y = \frac{\beta - \alpha}{b - a}\,(x - a) + \alpha \tag{3.14}$$

passing through the two points. Thus, the Euler–Lagrange equation helps to reconfirm our intuition that straight lines minimize distance.

Be that as it may, the fact that a function satisfies the Euler–Lagrange equation and the boundary conditions merely confirms its status as a critical function, and does not guarantee that it is the minimizer. Indeed, any critical function is also a candidate for *maximizing* the variational problem, too. The nature of a critical function will be elucidated by the second derivative test, and requires some further work. Of course, for the minimum distance problem, we "know" that a straight line cannot maximize distance, and must be the minimizer. Nevertheless, the reader should have a small nagging doubt that we may not have completely solved the problem at hand . . .

Minimal Surface of Revolution

Consider next the problem of finding the curve connecting two points that generates a surface of revolution of minimal surface area. For simplicity, we assume that the curve is given by the graph of a *non-negative* function $y = u(x) \geq 0$. According to (2.11), the required curve will minimize the functional

$$J[u] = \int_a^b u\,\sqrt{1+(u')^2}\;dx, \qquad \text{with Lagrangian} \qquad L(x,u,p) = u\,\sqrt{1+p^2}\,, \tag{3.15}$$

where we have omitted an irrelevant factor of 2π and used positivity to delete the absolute value on u in the integrand. Since

$$\frac{\partial L}{\partial u} = \sqrt{1+p^2}\,, \qquad \frac{\partial L}{\partial p} = \frac{u\,p}{\sqrt{1+p^2}}\,,$$

the Euler–Lagrange equation (3.11) is

$$\sqrt{1+(u')^2} \;-\; \frac{d}{dx}\,\frac{u\,u'}{\sqrt{1+(u')^2}} \;=\; \frac{1+(u')^2-u\,u''}{(1+(u')^2)^{3/2}} = 0. \tag{3.16}$$

Therefore, to find the critical functions, we need to solve a nonlinear second order ordinary differential equation — and not one in a familiar form.

Fortunately, there is a little trick[†] we can use to find the solution. If we multiply the equation by u', we can then rewrite the result as an exact derivative

$$u'\left[\frac{1+(u')^2-u\,u''}{(1+(u')^2)^{3/2}}\right] \;=\; \frac{d}{dx}\,\frac{u}{\sqrt{1+(u')^2}} = 0.$$

We conclude that the quantity

$$\frac{u}{\sqrt{1+(u')^2}} = c, \tag{3.17}$$

is constant, and so the left hand side is a *first integral* for the differential equation. Solving for[‡]

$$\frac{du}{dx} = u' = \frac{\sqrt{u^2-c^2}}{c}$$

results in an autonomous first order ordinary differential equation, which we can immediately solve:

$$\int \frac{c\;du}{\sqrt{u^2-c^2}} = x+\delta,$$

where δ is a constant of integration. The most useful form of the left hand integral is in terms of the inverse to the hyperbolic cosine function $\cosh z = \frac{1}{2}(e^z + e^{-z})$, whereby

$$\cosh^{-1}\frac{u}{c} = x+\delta, \qquad \text{and hence} \qquad u = c\,\cosh\left(\frac{x+\delta}{c}\right). \tag{3.18}$$

In this manner, we have produced the general solution to the Euler–Lagrange equation (3.16). Any solution that also satisfies the boundary conditions provides a critical function for the surface area functional (3.15), and hence is a candidate for the minimizer.

The curve prescribed by the graph of a hyperbolic cosine function (3.18) is known as a *catenary*. It is *not* a parabola, even though to the untrained eye it looks similar. Owing to their minimizing properties, catenaries are quite common in engineering design — for instance a hanging chain has the shape of a catenary, while the arch in St. Louis is an inverted catenary.

So far, we have not taken into account the boundary conditions It turns out that there are three distinct possibilities, depending upon the configuration of the boundary points:

---

† Actually, as with many tricks, this is really an indication that something profound is going on. Noether's Theorem, a result of fundamental importance in modern physics that relates symmetries and conservation laws, [**7**, **13**], underlies the integration method.

‡ The square root is real since, by (3.17), $|\,u\,| \le |\,c\,|$.

(*a*) There is precisely one value of the two integration constants c, δ that satisfies the two boundary conditions. In this case, it can be proved that this catenary is the unique curve that minimizes the area of its associated surface of revolution.

(*b*) There are two different possible values of c, δ that satisfy the boundary conditions. In this case, one of these is the minimizer, and the other is a spurious solution — one that corresponds to a saddle point for the surface area functional.

(*c*) There are *no* values of c, δ that allow (3.18) to satisfy the two boundary conditions. This occurs when the two boundary points $\mathbf{a}, \mathbf{b}$ are relatively far apart. In this configuration, the physical soap film spanning the two circular wires breaks apart into two circular disks, and this defines the minimizer for the problem, i.e., unlike cases (*a*) and (*b*), there is *no* surface of revolution that has a smaller surface area than the two disks. However, the "function"† that minimizes this configuration consists of two vertical lines from the boundary points to the x axis, along with that segment on the axis lying between them. More precisely, we can approximate this function by a sequence of genuine functions that give progressively smaller and smaller values to the surface area functional (2.11), but the actual minimum is not attained among the class of (smooth) functions.

Thus, even in such a reasonably simple example, a number of the subtle complications can already be seen. Lack of space precludes a more detailed development of the subject, and we refer the interested reader to more specialized books on the calculus of variations, including [**4**, **7**, **10**].

The Brachistochrone Problem

The most famous classical variational principle is the so-called *brachistochrone problem*. The compound Greek word "brachistochrone" means "minimal time". An experimenter lets a bead slide down a wire that connects two fixed points. The goal is to shape the wire in such a way that, starting from rest, the bead slides from one end to the other in minimal time. Naïve guesses for the wire's optimal shape, including a straight line, a parabola, a circular arc, or even a catenary are wrong. One can do better through a careful analysis of the associated variational problem. The brachistochrone problem was originally posed by the Swiss mathematician Johann Bernoulli in 1696, and served as an inspiration for much of the subsequent development of the subject.

We take, without loss of generality, the starting point of the bead to be at the origin: $\mathbf{a} = (0, 0)$. The wire will bend downwards, and so, to avoid distracting minus signs in the subsequent formulae, we take the vertical y axis to point downwards. The shape of the wire will be given by the graph of a function $y = u(x) \geq 0$. The end point $\mathbf{b} = (b, \beta)$ is assumed to lie below and to the right, and so $b > 0$ and $\beta > 0$. The set-up is sketched in Figure 4.

To mathematically formulate the problem, the first step is to find the formula for the transit time of the bead sliding along the wire. Arguing as in our derivation of the optics

† Here "function" must be taken in a *very* broad sense, as this one does not even correspond to a generalized function!

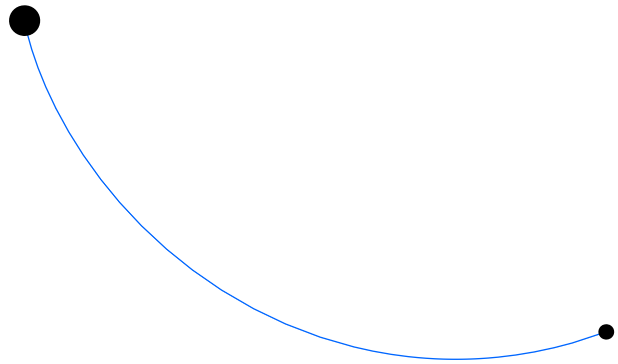

Figure 4. The Brachistochrone Problem.

functional (2.5), if $v(x)$ denotes the instantaneous speed of descent of the bead when it reaches position $(x, u(x))$, then the total travel time is

$$T[u] = \int_0^\ell \frac{ds}{v} = \int_0^b \frac{\sqrt{1+(u')^2}}{v}\, dx, \tag{3.19}$$

where $ds = \sqrt{1+(u')^2}\, dx$ is the usual arc length element, and ℓ is the overall length of the wire.

We shall use conservation of energy to determine a formula for the speed v as a function of the position along the wire. The kinetic energy of the bead is $\frac{1}{2} m v^2$, where m is its mass. On the other hand, due to our sign convention, the potential energy of the bead when it is at height $y = u(x)$ is $-m g u(x)$, where g the gravitational constant, and we take the initial height as the zero potential energy level. The bead is initially at rest, with 0 kinetic energy and 0 potential energy. Assuming that frictional forces are negligible, conservation of energy implies that the total energy must remain equal to 0, and hence

$$0 = \tfrac{1}{2} m v^2 - m g u.$$

We can solve this equation to determine the bead's speed as a function of its height:

$$v = \sqrt{2 g u}\,. \tag{3.20}$$

Substituting this expression into (3.19), we conclude that the shape $y = u(x)$ of the wire is obtained by minimizing the functional

$$T[u] = \int_0^b \sqrt{\frac{1+(u')^2}{2 g u}}\; dx, \tag{3.21}$$

subjet to the boundary conditions

$$u(0) = 0, \qquad u(b) = \beta. \tag{3.22}$$

The associated Lagrangian is

$$L(x,u,p) = \sqrt{\frac{1+p^2}{u}}\,,$$

where we omit an irrelevant factor of $\sqrt{2g}$ (or adopt physical units in which $g = \frac{1}{2}$). We compute

$$\frac{\partial L}{\partial u} = -\frac{\sqrt{1+p^2}}{2\,u^{3/2}}, \qquad \frac{\partial L}{\partial p} = \frac{p}{\sqrt{u\,(1+p^2)}}.$$

Therefore, the Euler–Lagrange equation for the brachistochrone functional is

$$-\frac{\sqrt{1+(u')^2}}{2\,u^{3/2}} - \frac{d}{dx}\,\frac{u'}{\sqrt{u\,(1+(u')^2)}} = -\frac{2\,u\,u'' + (u')^2 + 1}{2\,\sqrt{u\,(1+(u')^2)}} = 0. \tag{3.23}$$

Thus, the minimizing functions solve the nonlinear second order ordinary differential equation

$$2\,u\,u'' + (u')^2 + 1 = 0.$$

Rather than try to solve this differential equation directly, we note that the Lagrangian does not depend upon x, and therefore we can use the following result.

Theorem 3.2. *Suppose the Lagrangian $L(x,u,p) = L(u,p)$ does not depend on x. Then the* Hamiltonian function

$$H(u,u') = L(u,u') - u'\,\frac{\partial L}{\partial p}(u,u') \tag{3.24}$$

is a first integral *for the Euler–Lagrange equation, meaning that it is constant on each solution.*

Proof: Differentiating (3.24), we find

$$\frac{d}{dx}\,H(u,u') = \frac{d}{dx}\left[\,L(u,u') - u'\,\frac{\partial L}{\partial p}(u,u')\right] = u'\left(\frac{\partial L}{\partial u}(u,u') - \frac{d}{dx}\,\frac{\partial L}{\partial p}(u,u')\right) = 0\,,$$

which vanishes as a consequence of the Euler–Lagrange equation (3.11). This implies that

$$H(u,u') = k, \tag{3.25}$$

where k is a constant, whose value can depend upon the solution $u(x)$. Equation (3.25) has the form of an implicitly defined first order ordinary differential equation which can, in fact, be integrated. Indeed, solving for

$$u' = h(u,k)$$

produces an autonomous first order differential equation, whose general solution can be obtained by integration. *Q.E.D.*

Remark: This result is a special case of Noether's powerful Theorem, [**13**; Chapter 4], that relates symmetries of variational problems — in this case translations in the x coordinate — with first integrals, a.k.a. conservation laws.

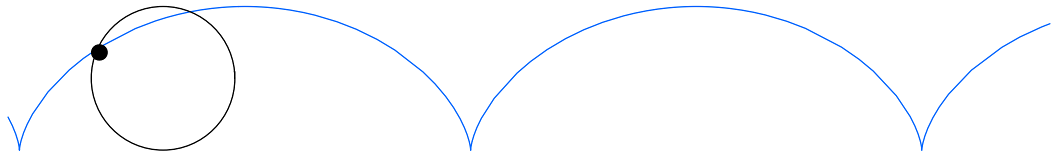

Figure 5. A Cycloid.

In our case, the Hamiltonian function (3.24) is

$$H(x,u,p) = L - p\,\frac{\partial L}{\partial p} = \frac{1}{\sqrt{u(1+p^2)}}$$

defines a first integral. Thus,

$$H(x,u,u') = \frac{1}{\sqrt{u(1+(u')^2)}} = k, \qquad \text{which we rewrite as} \qquad u(1+(u')^2) = c,$$

where $c = 1/k^2$ is a constant. (This can be checked by directly calculating $dH/dx \equiv 0$.) Solving for the derivative u' results in the first order autonomous ordinary differential equation

$$\frac{du}{dx} = \sqrt{\frac{c-u}{u}}\,.$$

This equation can be explicitly solved by separation of variables, and so

$$\int \sqrt{\frac{u}{c-u}}\;du = x + \delta$$

for some constant δ. The left hand integration relies on the trigonometric substitution

$$u = \tfrac{1}{2}\,c\,(1-\cos\theta),$$

whereby

$$x + \delta = \frac{c}{2}\int\sqrt{\frac{1-\cos\theta}{1+\cos\theta}}\;\sin\theta\,d\theta = \frac{c}{2}\int(1-\cos\theta)\,d\theta = \tfrac{1}{2}\,c\,(\theta - \sin\theta).$$

The left hand boundary condition implies $\delta = 0$, and so the solution to the Euler–Lagrange equation are curves parametrized by

$$x = r\,(\theta - \sin\theta), \qquad u = r\,(1-\cos\theta). \tag{3.26}$$

With a little more work, it can be proved that the parameter $r = \frac{1}{2}\,c$ is uniquely prescribed by the right hand boundary condition, and moreover, the resulting curve supplies the global minimizer of the brachistochrone functional, [**7**]. The miniming curve is known as a *cycloid*, which can be visualized as the curve traced by a point sitting on the edge of a rolling wheel of radius r, as plotted in Figure 5. Interestingly, in certain configurations, namely if $\beta < 2\,b/\pi$, the cycloid that solves the brachistochrone problem dips below the right hand endpoint $\mathbf{b} = (b,\beta)$, and so the bead is moving upwards when it reaches the end of the wire.

4. The Second Variation.

The solutions to the Euler–Lagrange boundary value problem are the critical functions for the variational principle, meaning that they cause the functional gradient to vanish. For finite-dimensional optimization problems, being a critical point is only a necessary condition for minimality. One must impose additional conditions, based on the second derivative of the objective function at the critical point, in order to guarantee that it is a minimum and not a maximum or saddle point. Similarly, in the calculus of variations, the solutions to the Euler–Lagrange equation may also include (local) maxima, as well as other non-extremal critical functions. To distinguish between the possibilities, we need to formulate a second derivative test for the objective functional. In the calculus of variations, the second derivative of a functional is known as its *second variation*, and the goal of this section is to construct and analyze it in its simplest manifestation.

For a finite-dimensional objective function $F(u_1, \ldots, u_n)$, the second derivative test was based on the positive definiteness of its Hessian matrix. The justification was based on the second order Taylor expansion of the objective function at the critical point. In an analogous fashion, we expand an objective functional $J[u]$ near the critical function. Consider the scalar function

$$h(\varepsilon) = J[u + \varepsilon v],$$

where the function $v(x)$ represents a variation. The second order Taylor expansion of $h(\varepsilon)$ takes the form

$$h(\varepsilon) = J[u + \varepsilon v] = J[u] + \varepsilon\, K[u; v] + \tfrac{1}{2}\,\varepsilon^2\, Q[u; v] + \ \cdots\ .$$

The first order terms are linear in the variation v, and, according to our earlier calculation, given by the inner product

$$h'(0) = K[u; v] = \langle\, \nabla J[u]\, ; v\,\rangle$$

between the variation and the functional gradient. In particular, if u is a critical function, then the first order terms vanish,

$$K[u; v] = \langle\, \nabla J[u]\, ; v\,\rangle = 0,$$

for all allowable variations v, meaning those that satisfy the homogeneous boundary conditions. Therefore, the nature of the critical function u — minimum, maximum, or neither — will, in most cases, be determined by the second derivative terms

$$h''(0) = Q[u; v].$$

Now, if u is a minimizer, then $Q[u; v] \geq 0$. Conversely, if $Q[u; v] > 0$ for all $v \not\equiv 0$, i.e., the second variation is positive definite, then the critical function u will be a strict local minimizer. This forms the crux of the second derivative test.

Let us explicitly evaluate the second variational for the simplest functional (3.1). Consider the scalar function

$$h(\varepsilon) = J[u + \varepsilon v] = \int_a^b L(x, u + \varepsilon v, u' + \varepsilon v')\, dx,$$

whose first derivative $h'(0)$ was already determined in (3.6); here we require the second variation

$$Q[u;v] = h''(0) = \int_a^b \left[\, A v^2 + 2\,B\,v\,v' + C\,(v')^2 \,\right] dx, \tag{4.1}$$

where the coefficient functions

$$A(x) = \frac{\partial^2 L}{\partial u^2}\,(x,u,u')\,, \qquad B(x) = \frac{\partial^2 L}{\partial u\,\partial p}\,(x,u,u')\,, \qquad C(x) = \frac{\partial^2 L}{\partial p^2}\,(x,u,u')\,, \tag{4.2}$$

are found by evaluating certain second order derivatives of the Lagrangian at the critical function $u(x)$. In contrast to the first variation, integration by parts will not eliminate all of the derivatives on v in the quadratic functional (4.1), which causes significant complications in the ensuing analysis.

The second derivative test for a minimizer relies on the positivity of the second variation. So, in order to formulate conditions that the critical function be a minimizer for the functional, we need to establish criteria guaranteeing the positive definiteness of such a quadratic functional, meaning that $Q[u;v] > 0$ for all non-zero allowable variations $v(x) \not\equiv 0$. Clearly, if the integrand is positive definite at each point, so

$$A(x)\,v^2 + 2\,B(x)\,v\,v' + C(x)\,(v')^2 > 0 \quad \text{whenever} \quad a < x < b, \quad \text{and} \quad v(x) \not\equiv 0, \tag{4.3}$$

then $Q[u;v] > 0$ is also positive definite.

Example 4.1. For the arc length minimization functional (2.3), the Lagrangian is $L(x,u,p) = \sqrt{1+p^2}$. To analyze the second variation, we first compute

$$\frac{\partial^2 L}{\partial u^2} = 0, \qquad \frac{\partial^2 L}{\partial u\,\partial p} = 0, \qquad \frac{\partial^2 L}{\partial p^2} = \frac{1}{(1+p^2)^{3/2}}\,.$$

For the critical straight line function

$$u(x) = \frac{\beta-\alpha}{b-a}\,(x-a) + \alpha, \qquad \text{with} \qquad p = u'(x) = \frac{\beta-\alpha}{b-a}\,,$$

we find

$$A(x) = \frac{\partial^2 L}{\partial u^2} = 0, \quad B(x) = \frac{\partial^2 L}{\partial u\,\partial p} = 0, \quad C(x) = \frac{\partial^2 L}{\partial p^2} = \frac{(b-a)^3}{\left[\,(b-a)^2 + (\beta-\alpha)^2\,\right]^{3/2}} \equiv c.$$

Therefore, the second variation functional (4.1) is

$$Q[u;v] = \int_a^b c\,(v')^2\,dx,$$

where $c > 0$ is a positive constant. Thus, $Q[u;v] = 0$ vanishes if and only if $v(x)$ is a constant function. But the variation $v(x)$ is required to satisfy the homogeneous boundary conditions $v(a) = v(b) = 0$, and hence $Q[u;v] > 0$ for all allowable nonzero variations. Therefore, we conclude that the straight line is, indeed, a (local) minimizer for the arc length functional. We have at last rigorously justified our intuition that the shortest distance between two points is a straight line!

However, as the following example demonstrates, the pointwise positivity condition (4.3) is overly restrictive.

Example 4.2. Consider the quadratic functional

$$Q[v] = \int_0^1 \left[(v')^2 - v^2 \right] dx. \tag{4.4}$$

We claim that $Q[v] > 0$ for all nonzero $v \not\equiv 0$ subject to homogeneous Dirichlet boundary conditions $v(0) = 0 = v(1)$. This result is not trivial! Indeed, the boundary conditions play an essential role, since choosing $v(x) \equiv c \neq 0$ to be any constant function will produce a negative value for the functional: $Q[v] = -\,c^2$.

To prove the claim, consider the quadratic functional

$$\widetilde{Q}[v] = \int_0^1 (v' + v \tan x)^2 \, dx \;\geq\; 0,$$

which is clearly non-negative, since the integrand is everywhere ≥ 0. Moreover, by continuity, the integral vanishes if and only if v satisfies the first order linear ordinary differential equation

$$v' + v \tan x = 0, \qquad \text{for all} \qquad 0 \leq x \leq 1.$$

The only solution that also satisfies boundary condition $v(0) = 0$ is the trivial one $v \equiv 0$. We conclude that $\widetilde{Q}[v] = 0$ if and only if $v \equiv 0$, and hence $\widetilde{Q}[v]$ is a positive definite quadratic functional on the space of allowable variations.

Let us expand the latter functional,

$$\begin{aligned} \widetilde{Q}[v] &= \int_0^1 \left[(v')^2 + 2\,v\,v' \tan x + v^2 \tan^2 x \right] dx \\ &= \int_0^1 \left[(v')^2 - v^2 \,(\tan x)' + v^2 \tan^2 x \right] dx = \int_0^1 \left[(v')^2 - v^2 \right] dx = Q[v]. \end{aligned}$$

In the second equality, we integrated the middle term by parts, using $(v^2)' = 2\,v\,v'$, and noting that the boundary terms vanish owing to our imposed boundary conditions. Since $\widetilde{Q}[v]$ is positive definite, so is $Q[v]$, justifying the previous claim.

To appreciate how subtle this result is, consider the almost identical quadratic functional

$$\widehat{Q}[v] = \int_0^4 \left[(v')^2 - v^2 \right] dx, \tag{4.5}$$

the only difference being the upper limit of the integral. A quick computation shows that the function $v(x) = x(4-x)$ satisfies the boundary conditions $v(0) = 0 = v(4)$, but

$$\widehat{Q}[v] = \int_0^4 \left[(4-2\,x)^2 - x^2 (4-x)^2 \right] dx = -\,\frac{64}{5} < 0.$$

Therefore, $\widehat{Q}[v]$ is *not* positive definite. Our preceding analysis does not apply because the function $\tan x$ becomes singular at $x = \frac{1}{2}\pi$, and so the auxiliary integral $\displaystyle\int_0^4 (v' + v \tan x)^2 \, dx$ does not converge.

The complete analysis of positive definiteness of quadratic functionals is quite subtle. Indeed, the strange appearance of $\tan x$ in the preceding example turns out to be an important clue! In the interests of brevity, let us just state without proof a fundamental theorem, and refer the interested reader to [**7**] for full details.

Theorem 4.3. *Let $A(x), B(x), C(x) \in \mathrm{C}^0[a,b]$ be continuous functions. The quadratic functional*

$$Q[v] = \int_a^b \left[A v^2 + 2 B v v' + C (v')^2 \right] dx$$

is positive definite, so $Q[v] > 0$ for all $v \not\equiv 0$ satisfying the homogeneous Dirichlet boundary conditions $v(a) = v(b) = 0$, provided

(a) *$C(x) > 0$ for all $a \le x \le b$, and*

(b) *For any $a < c \le b$, the only solution to its linear Euler–Lagrange boundary value problem*

$$-(C v')' + (A - B')v = 0, \qquad v(a) = 0 = v(c), \tag{4.6}$$

is the trivial function $v(x) \equiv 0$.

Remark: A value c for which (4.6) has a nontrivial solution is known as a *conjugate point* to a. Thus, condition (*b*) can be restated that the variational problem has no conjugate points in the interval $(a, b]$.

Example 4.4. The quadratic functional

$$Q[v] = \int_0^b \left[(v')^2 - v^2 \right] dx \tag{4.7}$$

has Euler–Lagrange equation

$$-v'' - v = 0.$$

The solutions $v(x) = k \sin x$ satisfy the boundary condition $v(0) = 0$. The first conjugate point occurs at $c = \pi$ where $v(\pi) = 0$. Therefore, Theorem 4.3 implies that the quadratic functional (4.7) is positive definite *provided* the upper integration limit $b < \pi$. This explains why the original quadratic functional (4.4) is positive definite, since there are no conjugate points on the interval $[0,1]$, while the modified version (4.5) is *not*, because the first conjugate point π lies on the interval $(0,4]$.

In the case when the quadratic functional arises as the second variation of a functional (3.1), the coefficient functions A, B, C are given in terms of the Lagrangian $L(x,u,p)$ by formulae (4.2). In this case, the first condition in Theorem 4.3 requires

$$\frac{\partial^2 L}{\partial p^2}(x,u,u') > 0 \tag{4.8}$$

for the minimizer $u(x)$. This is known as the *Legendre condition*. The second, *conjugate point condition* requires that the so-called *linear variational equation*

$$-\frac{d}{dx}\left(\frac{\partial^2 L}{\partial p^2}(x,u,u') \frac{dv}{dx} \right) + \left(\frac{\partial^2 L}{\partial u^2}(x,u,u') - \frac{d}{dx} \frac{\partial^2 L}{\partial u\, \partial p}(x,u,u') \right) v = 0 \tag{4.9}$$

has no nontrivial solutions $v(x) \not\equiv 0$ that satisfy $v(a) = 0$ and $v(c) = 0$ for $a < c \leq b$. In this way, we have arrived at a rigorous form of the second derivative test for the simplest functional in the calculus of variations.

Theorem 4.5. *If the function $u(x)$ satisfies the Euler–Lagrange equation* (3.11), *and, in addition, the Legendre condition* (4.8) *holds and there are no conjugate points on the interval, then $u(x)$ is a strict local minimum for the functional.*

5. Multi-dimensional Variational Problems.

The calculus of variations encompasses a very broad range of mathematical applications. The methods of variational analysis can be applied to an enormous variety of physical systems, whose equilibrium configurations inevitably minimize a suitable functional, which, typically, represents the potential energy of the system. Minimizing configurations appear as critical functions at which the functional gradient vanishes. Following similar computational procedures as in the one-dimensional calculus of variations, we find that the critical functions are characterized as solutions to a system of partial differential equations, known as the Euler–Lagrange equations associated with the variational principle. Each solution to the boundary value problem specified by the Euler–Lagrange equations subject to appropriate boundary conditions is, thus, a candidate minimizer for the variational problem. In many applications, the Euler–Lagrange boundary value problem suffices to single out the physically relevant solutions, and one need not press on to the considerably more difficult second variation.

Implementation of the variational calculus for functionals in higher dimensions will be illustrated by looking at a specific example — a first order variational problem involving a single scalar function of two variables. Once this is fully understood, generalizations and extensions to higher dimensions and higher order Lagrangians are readily apparent. Thus, we consider an objective functional

$$J[u] = \iint_\Omega L(x, y, u, u_x, u_y)\; dx\, dy, \tag{5.1}$$

having the form of a double integral over a prescribed domain $\Omega \subset \mathbb{R}^2$. The *Lagrangian* $L(x, y, u, p, q)$ is assumed to be a sufficiently smooth function of its five arguments. Our goal is to find the function(s) $u = f(x, y)$ that minimize the value of $J[u]$ when subject to a set of prescribed boundary conditions on $\partial\Omega$, the most important being our usual Dirichlet, Neumann, or mixed boundary conditions. For simplicity, we concentrate on the Dirichlet boundary value problem, and require that the minimizer satisfy

$$u(x, y) = g(x, y) \qquad \text{for} \qquad (x, y) \in \partial\Omega. \tag{5.2}$$

The First Variation

The basic necessary condition for an extremum (minimum or maximum) is obtained in precisely the same manner as in the one-dimensional framework. Consider the scalar function

$$h(\varepsilon) \equiv J[u+\varepsilon v] = \iint_\Omega L(x,y,u+\varepsilon v, u_x+\varepsilon v_x, u_y+\varepsilon v_y)\, dx\, dy$$

depending on $\varepsilon \in \mathbb{R}$. The *variation* $v(x,y)$ is assumed to satisfy homogeneous Dirichlet boundary conditions

$$v(x,y) = 0 \qquad \text{for} \qquad (x,y) \in \partial\Omega, \tag{5.3}$$

to ensure that $u+\varepsilon v$ satisfies the same boundary conditions (5.2) as u itself. Under these conditions, if u is a minimizer, then the scalar function $h(\varepsilon)$ will have a minimum at $\varepsilon = 0$, and hence

$$h'(0) = 0.$$

When computing $h'(\varepsilon)$, we assume that the functions involved are sufficiently smooth so as to allow us to bring the derivative inside the integral, and then apply the chain rule. At $\varepsilon = 0$, the result is

$$h'(0) = \frac{d}{d\varepsilon} J[u+\varepsilon v] \,\bigg|_{\varepsilon=0} = \iint_\Omega \left(v\,\frac{\partial L}{\partial u} + v_x\,\frac{\partial L}{\partial p} + v_y\,\frac{\partial L}{\partial q}\right) dx\, dy, \tag{5.4}$$

where the derivatives of L are all evaluated at x, y, u, u_x, u_y. To identify the functional gradient, we need to rewrite this integral in the form of an inner product:

$$h'(0) = \langle\, \nabla J[u]\,;v\,\rangle = \iint_\Omega h(x,y)\, v(x,y)\, dx\, dy, \qquad \text{where} \qquad h = \nabla J[u].$$

To convert (5.4) into this form, we need to remove the offending derivatives from v. In two dimensions, the requisite integration by parts formula is based on Green's Theorem:

$$\iint_\Omega \left(\frac{\partial v}{\partial x}\, w_1 + \frac{\partial v}{\partial y}\, w_2 \right) dx\, dy = \oint_{\partial\Omega} v\,(-w_2\, dx + w_1\, dy) - \iint_\Omega v \left(\frac{\partial w_1}{\partial x} + \frac{\partial w_2}{\partial y} \right) dx\, dy, \tag{5.5}$$

in which w_1, w_2 are arbitrary smooth functions. Setting $w_1 = \dfrac{\partial L}{\partial p}$, $w_2 = \dfrac{\partial L}{\partial q}$, we find

$$\iint_\Omega \left(v_x\,\frac{\partial L}{\partial p} + v_y\,\frac{\partial L}{\partial q}\right) dx\, dy = -\iint_\Omega v \left[\frac{\partial}{\partial x}\left(\frac{\partial L}{\partial p}\right) + \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial q}\right)\right] dx\, dy,$$

where the boundary integral vanishes owing to the boundary conditions (5.3) that we impose on the allowed variations. Substituting this result back into (5.4), we conclude that

$$h'(0) = \iint_\Omega v \left[\frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial p}\right) - \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial q}\right)\right] dx\, dy = \langle\, \nabla J[u]\,;v\,\rangle, \tag{5.6}$$

where

$$\nabla J[u] = \frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial p}\right) - \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial q}\right)$$

is the desired first variation or functional gradient. Since the gradient vanishes at a critical function, we conclude that the minimizer $u(x,y)$ must satisfy the *Euler–Lagrange equation*

$$\frac{\partial L}{\partial u}(x,y,u,u_x,u_y) - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial p}(x,y,u,u_x,u_y)\right) - \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial q}(x,y,u,u_x,u_y)\right) = 0. \quad (5.7)$$

Once we explicitly evaluate the derivatives, the net result is a second order partial differential equation

$$L_u - L_{xp} - L_{yq} - u_x L_{up} - u_y L_{uq} - u_{xx} L_{pp} - 2\, u_{xy} L_{pq} - u_{yy} L_{qq} = 0, \qquad (5.8)$$

where we use subscripts to indicate derivatives of both u and L, the latter being evaluated at x, y, u, u_x, u_y.

Example 5.1. As a first elementary example, consider the Dirichlet minimization problem

$$J[u] = \iint_\Omega \tfrac{1}{2}\left(u_x^2 + u_y^2\right) dx\, dy. \qquad (5.9)$$

In this case, the associated Lagrangian is

$$L = \tfrac{1}{2}(p^2+q^2), \qquad \text{with} \qquad \frac{\partial L}{\partial u} = 0, \qquad \frac{\partial L}{\partial p} = p = u_x, \qquad \frac{\partial L}{\partial q} = q = u_y.$$

Therefore, the Euler–Lagrange equation (5.7) becomes

$$-\frac{\partial}{\partial x}(u_x) - \frac{\partial}{\partial y}(u_y) = -\,u_{xx} - u_{yy} = -\,\Delta u = 0,$$

which is the two-dimensional Laplace equation. Subject to the selected boundary conditions, the solutions, i.e., the harmonic functions, are critical functions for the Dirichlet variational principle.

However, the calculus of variations approach, as developed so far, leads to a much weaker result since it only singles out the harmonic functions as *candidates* for minimizing the Dirichlet integral; they could just as easily be maximizing functions or saddle points. When dealing with a quadratic variational problem, the direct algebraic approach is, when applicable, the more powerful, since it assures us that the solutions to the Laplace equation really do minimize the integral among the space of functions satisfying the appropriate boundary conditions. However, the direct method is restricted to quadratic variational problems, whose Euler–Lagrange equations are linear partial differential equations. In nonlinear cases, one really does need to utilize the full power of the variational machinery.

Example 5.2. Let us derive the Euler–Lagrange equation for the minimal surface problem. From (2.9), the surface area integral

$$J[u] = \iint_\Omega \sqrt{1+u_x^2+u_y^2}\; dx\, dy \qquad \text{has Lagrangian} \qquad L = \sqrt{1+p^2+q^2}\,.$$

Note that

$$\frac{\partial L}{\partial u} = 0, \qquad \frac{\partial L}{\partial p} = \frac{p}{\sqrt{1+p^2+q^2}}\,, \qquad \frac{\partial L}{\partial q} = \frac{q}{\sqrt{1+p^2+q^2}}\,.$$

Therefore, replacing $p \to u_x$ and $q \to u_y$ and then evaluating the derivatives, the Euler–Lagrange equation (5.7) becomes

$$- \frac{\partial}{\partial x} \frac{u_x}{\sqrt{1 + u_x^2 + u_y^2}} - \frac{\partial}{\partial y} \frac{u_y}{\sqrt{1 + u_x^2 + u_y^2}} = \frac{-(1 + u_y^2) u_{xx} + 2 u_x u_y u_{xy} - (1 + u_x^2) u_{yy}}{(1 + u_x^2 + u_y^2)^{3/2}} = 0.$$

Thus, a surface described by the graph of a function $u = f(x,y)$ is a critical function, and hence a candidate for minimizing surface area, provided it satisfies the *minimal surface equation*

$$(1 + u_y^2)\, u_{xx} - 2\, u_x\, u_y\, u_{xy} + (1 + u_x^2)\, u_{yy} = 0. \tag{5.10}$$

We are confronted with a complicated, nonlinear, second order partial differential equation, which has been the focus of some of the most sophisticated and deep analysis over the past two centuries, with significant progress on understanding its solution only within the past 70 years. In this book, we have not developed the sophisticated analytical, geometrical, and numerical techniques that are required to have anything of substance to say about its solutions. We refer the interested reader to the advanced texts [**11**, **12**] for further developments in this fascinating area of mathematical analysis.

Example 5.3. The deformations of an elastic body $\Omega \subset \mathbb{R}^n$ are described by the *displacement* field, $\mathbf{u}\colon \Omega \to \mathbb{R}^n$. Each material point $\mathbf{x} \in \Omega$ in the undeformed body will move to a new position $\mathbf{y} = \mathbf{x} + \mathbf{u}(\mathbf{x})$ in the deformed body

$$\widetilde{\Omega} = \{\, \mathbf{y} = \mathbf{x} + \mathbf{u}(\mathbf{x}) \mid \ \mathbf{x} \in \Omega \,\}.$$

The one-dimensional case governs bars, beams and rods, two-dimensional bodies include thin plates and shells, while $n = 3$ for fully three-dimensional solid bodies. See [**1**, **8**] for details and physical derivations.

For small deformations, we can use a linear theory to approximate the much more complicated equations of nonlinear elasticity. The simplest case is that of a homogeneous and isotropic planar body $\Omega \subset \mathbb{R}^2$. The equilibrium mechanics are described by the deformation function $\mathbf{u}(\mathbf{x}) = (\, u(x,y), v(x,y) \,)$. A detailed physical analysis of the constitutive assumptions leads to a minimization principle based on the following functional:

$$\begin{aligned} J[u,v] &= \iint_\Omega \left[\, \tfrac{1}{2}\,\mu\, \| \nabla \mathbf{u} \|^2 + \tfrac{1}{2}\,(\lambda + \mu)(\nabla \cdot \mathbf{u})^2 \,\right] dx\, dy \\ &= \iint_\Omega \left[\, \left(\tfrac{1}{2}\,\lambda + \mu \right) (u_x^2 + v_y^2) + \tfrac{1}{2}\,\mu\,(u_y^2 + v_x^2) + (\lambda + \mu)\, u_x\, v_y \,\right] dx\, dy. \end{aligned} \tag{5.11}$$

The parameters λ, μ are known as the *Lamé moduli* of the material, and govern its intrinsic elastic properties. They are measured by performing suitable experiments on a sample of the material. Physically, (5.11) represents the stored (or potential) energy in the body under the prescribed displacement. Nature, as always, seeks the displacement that will minimize the total energy.

To compute the Euler–Lagrange equations, we consider the functional variation

$$h(\varepsilon) = J[\, u + \varepsilon\, f, v + \varepsilon\, g \,],$$

in which the individual variations f, g are arbitrary functions subject only to the given homogeneous boundary conditions. If u, v minimize J, then $h(\varepsilon)$ has a minimum at $\varepsilon = 0$, and so we are led to compute

$$h'(0) = \langle\, \nabla J\,;\mathbf{f}\,\rangle = \iint_\Omega \left(f\, \nabla_u J + g\, \nabla_v J \right) dx\, dy,$$

which we write as an inner product (using the standard L^2 inner product between vector fields) between the variation $\mathbf{f}$ and the functional gradient $\nabla J = (\,\nabla_u J, \nabla_v J\,)$. For the particular functional (5.11), we find

$$h'(0) = \iint_\Omega \left[\, (\,\lambda + 2\,\mu\,)\,(u_x\, f_x + v_y\, g_y) + \mu\,(u_y\, f_y + v_x\, g_x) + (\lambda + \mu)\,(u_x\, g_y + v_y\, f_x) \,\right] dx\, dy.$$

We use the integration by parts formula (5.5) to remove the derivatives from the variations f, g. Discarding the boundary integrals, which are used to prescribe the allowable boundary conditions, we find

$$h'(0) = -\iint_\Omega \left(\begin{array}{l} \left[\, (\lambda + 2\,\mu)\, u_{xx} + \mu\, u_{yy} + (\lambda + \mu)\, v_{xy} \,\right] f \\ \qquad + \left[\, (\lambda + \mu)\, u_{xy} + \mu\, v_{xx} + (\lambda + 2\,\mu)\, v_{yy} \,\right] g \end{array} \right) dx\, dy.$$

The two terms in brackets give the two components of the functional gradient. Setting them equal to zero, we derive the second order linear system of Euler–Lagrange equations

$$\begin{aligned} (\lambda + 2\,\mu)\, u_{xx} + \mu\, u_{yy} + (\lambda + \mu)\, v_{xy} &= 0, \\ (\lambda + \mu)\, u_{xy} + \mu\, v_{xx} + (\lambda + 2\,\mu)\, v_{yy} &= 0, \end{aligned} \tag{5.12}$$

known as *Navier's equations*, which can be compactly written as

$$\mu\, \Delta \mathbf{u} + (\mu + \lambda)\, \nabla(\nabla \cdot \mathbf{u}) = \mathbf{0} \tag{5.13}$$

for the displacement vector $\mathbf{u} = (\, u, v\,)$. The solutions to are the critical displacements that, under appropriate boundary conditions, minimize the potential energy functional.

Since we are dealing with a quadratic functional, a more detailed algebraic analysis will demonstrate that the solutions to Navier's equations are the minimizers for the variational principle (5.11). Although only valid in a limited range of physical and kinematical conditions, the solutions to the planar Navier's equations and its three-dimensional counterpart are successfully used to model a wide class of elastic materials.

In general, the solutions to the Euler–Lagrange boundary value problem are critical functions for the variational problem, and hence include all (smooth) local and global minimizers. Determination of which solutions are genuine minima requires a further analysis of the positivity properties of the second variation, which is beyond the scope of our introductory treatment. Indeed, a complete analysis of the positive definiteness of the second variation of multi-dimensional variational problems is quite complicated, and still awaits a completely satisfactory resolution!

References

[**1**] Antman, S.S., *Nonlinear Problems of Elasticity*, Appl. Math. Sci., vol. 107, Springer–Verlag, New York, 1995.

[**2**] Ball, J.M., and Mizel, V.J., One-dimensional variational problem whose minimizers do not satisfy the Euler-Lagrange equation, *Arch. Rat. Mech. Anal.* **90** (1985), 325–388.

[**3**] Born, M., and Wolf, E., *Principles of Optics*, Fourth Edition, Pergamon Press, New York, 1970.

[**4**] Courant, R., and Hilbert, D., *Methods of Mathematical Physics*, vol. I, Interscience Publ., New York, 1953.

[**5**] Dacorogna, B., *Introduction to the Calculus of Variations*, Imperial College Press, London, 2004.

[**6**] do Carmo, M.P., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, N.J., 1976.

[**7**] Gel'fand, I.M., and Fomin, S.V., *Calculus of Variations*, Prentice–Hall, Inc., Englewood Cliffs, N.J., 1963.

[**8**] Gurtin, M.E., *An Introduction to Continuum Mechanics*, Academic Press, New York, 1981.

[**9**] Hildebrandt, S., and Tromba, A., *Mathematics and Optimal Form*, Scientific American Books, New York, 1985.

[**10**] Kot, M., *A First Course in the Calculus of Variations*, American Mathematical Society, Providence, R.I., 2014.

[**11**] Morgan, F., *Geometric Measure Theory*: *a Beginner's Guide*, Academic Press, New York, 2000.

[**12**] Nitsche, J.C.C., *Lectures on Minimal Surfaces*, Cambridge University Press, Cambridge, 1988.

[**13**] Olver, P.J., *Applications of Lie Groups to Differential Equations*, 2nd ed., Graduate Texts in Mathematics, vol. 107, Springer–Verlag, New York, 1993.

[**14**] Olver, P.J., *Introduction to Partial Differential Equations*, Undergraduate Texts in Mathematics, Springer–Verlag, New York, to appear.