Chapter 21

The Calculus of Variations

We have already had ample encounters with Nature’s propensity to optimize. Minimization principles form one of the most powerful tools for formulating mathematical models governing the equilibrium configurations of physical systems. Moreover, the design of numerical integration schemes such as the powerful finite element method are also founded upon a minimization paradigm. This chapter is devoted to the mathematical analysis of 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. Solutions to minimization problems in the calculus of variations lead to boundary value problems for ordinary and partial differential equations. Numerical solutions are primarily based upon a nonlinear version of the finite element method. The methods developed to handle such problems prove to be fundamental in many areas of mathematics, physics, engineering, and other applications.

The history of the calculus of variations is tightly interwoven with the history of calculus, and has merited the attention of a remarkable range of mathematicians, beginning with Newton, then developed as a field of mathematics in its own right by the Bernoulli family. The first major developments appeared in the work of Euler, Lagrange and Laplace. In the nineteenth century, Hamilton, Dirichlet and Hilbert are but a few of the outstanding contributors. In modern times, the calculus of variations has continued to occupy center stage in research, including major theoretical advances, along with wide-ranging applications in physics, engineering and all branches of mathematics. In this chapter, we will only have time to scratch the surface of the vast area of classical and contemporary research.

Minimization problems amenable to the methods of 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, and many, many others. Many geometrical systems, 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 reduced the infinite-dimensional minimization problem to a finite-dimensional problem, to which we can apply the optimization techniques learned in Section 19.3.

We have already treated the simplest problems in the calculus of variations. As we learned in Chapters 11 and 15, minimization of a quadratic functional requires solving an associated boundary value problem for a linear differential equation. 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 gradient leads to a system of algebraic equations; the functional gradient leads to a boundary value problem for a nonlinear ordinary or partial differential equation. Thus, the passage from finite to infinite dimensional nonlinear systems mirrors the transition from linear algebraic systems to boundary value problems.


The best way to introduce the subject is to introduce some concrete examples of both mathematical and practical importance. These particular minimization problems played a key role in the historical development of the calculus of variations. And they still serve as an excellent motivation for learning its basic constructions.

**Minimal Curves and Geodesics**

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

\[
\mathbf{a} = (a, \alpha) \quad \text{and} \quad \mathbf{b} = (b, \beta) \quad \text{in the plane } \mathbb{R}^2.
\]  

Our goal is to find the curve of shortest length connecting them. “Obviously”, as you learn in childhood, the shortest path between two points is a straight line; see Figure 21.1. Mathematically, then, the minimizing curve we are after should be given as the graph of the particular affine function\(^\dagger\)

\[
y = c x + d = \frac{\beta - \alpha}{b - a} (x - a) + \alpha
\]  

passing through the two points. However, this commonly accepted “fact” that (21.2) is the solution to the minimization problem is, upon closer inspection, perhaps not so immediately obvious from a rigorous mathematical standpoint.

\(^\dagger\) We assume that \(a \neq b\), i.e., the points \(\mathbf{a}, \mathbf{b}\) do not lie on a common vertical line.
Let us see how we might properly formulate the minimal curve problem. Let us assume that the minimal curve is given as the graph of a smooth function \( y = u(x) \). Then, according to (A.27), the length of the curve is given by the standard arc length integral

\[
J[u] = \int_{a}^{b} \sqrt{1 + (u')^2} \, dx,
\]

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

\[
u(a) = \alpha, \quad u(b) = \beta,
\]

in order that its graph pass through the two prescribed points (21.1). The minimal curve problem requires us to find the function \( y = u(x) \) that minimizes the arc length functional (21.3) among all reasonable functions satisfying the prescribed boundary conditions. The student should pause to reflect on whether it is mathematically obvious that the affine function (21.2) is the one that minimizes the arc length integral (21.3) subject to the given boundary conditions. One of the motivating tasks of the calculus of variations, then, is to rigorously prove that our childhood intuition is indeed correct.

Indeed, the word "reasonable" is important. For the arc length functional to be defined, the function \( u(x) \) should be at least piecewise \( C^1 \), i.e., continuous with a piecewise continuous derivative. If we allow discontinuous functions, then the straight line (21.2) does not, in most cases, give the minimizer; see Exercise ■. Moreover, continuous functions which are not piecewise \( C^1 \) may not have a well-defined length. The more seriously one thinks about these issues, the less evident the solution becomes. But, rest assured that the "obvious" solution (21.2) does indeed turn out to be the true minimizer. However, a fully rigorous mathematical proof of this fact requires a proper development of the calculus of variations machinery.

A closely related problem arises in optics. The general principle, first formulated by the seventeenth century French mathematician Pierre de Fermat, is that when a light ray moves through an optical medium, e.g., a vacuum, it travels along a path that will minimize the travel time. As always, Nature seeks the most economical solution! Let \( c(x, y) \) denote the speed of light at each point in the medium\(^\dagger\). Speed is equal to the time derivative of distance traveled, namely, the arc length (21.3) 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 of the light ray is equal to

\[
T[u] = \int_{0}^{T} dt = \int_{a}^{b} \frac{dt}{dx} dx = \int_{a}^{b} \frac{\sqrt{1 + (u'(x))^2}}{c(x, u(x))} dx.
\]

Fermat’s Principle states that, to get from one point to another, the light ray follows the curve \( y = u(x) \) that minimizes this functional. If the medium is homogeneous, then

\(^\dagger\) For simplicity, we only consider the two-dimensional case here.
Figure 21.2. Geodesics on a Cylinder.

c(x, y) \equiv c is constant, and \( T[u] \) equals a multiple of the arc length functional, whose minimizers are the “obvious” straight lines. 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. All of the known laws of optics and lens design, governing focusing, refraction, etc., all follow as consequences of the minimization principle, \[ \text{optics} \].

Another problem of a similar ilk is to construct the geodesics on a curved surface, meaning the curves of minimal length. In other words, given two points \( a, b \) 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 the geodesic curves turn out to be straight lines parallel to the center line, circles orthogonal to the center line, and spiral helices; see Figure 21.2 for an illustration. Similarly, the geodesics on a sphere are arcs of great circles; these include the circumpolar paths followed by airplanes around the globe. However, both of these claims are in need of rigorous justification.

In order to mathematically formulate the geodesic 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

\[
\begin{align*}
\mathbf{a} &= (a, \alpha, F(a, \alpha)), & \text{and} & & \mathbf{b} &= (b, \beta, F(b, \beta)), & \text{on the surface} & & S.
\end{align*}
\]

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

\[
\begin{align*}
y &= u(x), & z &= F(x, u(x)).
\end{align*}
\]

In particular, this requires \( a \neq b \). The length of the curve is given by the usual arc length integral (B.17), and so we must minimize the functional

\[
J[u] = \int_a^b \sqrt{1 + \left( \frac{dy}{dx} \right)^2 + \left( \frac{dz}{dx} \right)^2} \, dx
\]

\[
= \int_a^b \sqrt{1 + \left( \frac{du}{dx} \right)^2 + \left( \frac{\partial F}{\partial x}(x, u(x)) + \frac{\partial F}{\partial u}(x, u(x)) \frac{du}{dx} \right)^2} \, dx,
\]

\[ \text{Cylinders are not graphs, but can be placed within this framework by passing to cylindrical coordinates. Similarly, spherical surfaces are best treated in spherical coordinates.} \]
subject to the boundary conditions
\[ u(a) = \alpha, \quad u(b) = \beta. \]
For example, the geodesics on the paraboloid
\[ z = \frac{1}{2} x^2 + \frac{1}{2} y^2 \]
can be found by minimizing the functional
\[ J[u] = \int_a^b \sqrt{1 + (u')^2 + (x + uu')^2} \, dx \]
subject to prescribed boundary conditions.

**Minimal Surfaces**

The minimal surface problem is a natural generalization of the minimal curve problem. In its simplest manifestation, we are given a simple closed curve \( C \subset \mathbb{R}^3 \). The problem is to find the surface \( S \) of least total area among all those whose boundary \( \partial S = C \) coincides with the given curve. Therefore, 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.

Physically, if we take a wire in the shape of the curve \( C \) and dip it into soapy water, then the surface tension forces in the resulting soap film will force it to minimize surface area, and hence be a minimal surface. For example, if the curve is a closed plane curve, e.g., a circle, then the minimal surface will just be the planar region enclosed by the curve. But, if the curve \( C \) twists into the third dimension, then the shape of the minimizer is by no means evident. Soap films and bubbles have been the source of much fascination, physical, aesthetic and mathematical, over the centuries. The least area problem is also known as *Plateau’s Problem*, after the nineteenth century French physicist Joseph Plateau, who conducted systematic experiments. A satisfactory solution to the simplest version of the minimal surface problem was only achieved in the mid twentieth century, [109, 112]. Problems arising in engineering design, architecture, and biology, such as foams, membranes and drug delivery methods, make this problem of continued contemporary importance and an active area of research.

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 minsurf. The space curve \( C \subset \mathbb{R}^3 \) is then given by \( z = g(x,y) \)

---

\( ^1 \) More correctly, the soap film will realize a local but not necessarily global minimum for the surface area functional.
for \((x, y) \in \partial \Omega\). For reasonable curves \(C\), we expect that the minimal surface \(S\) will be described as the graph of a function \(z = u(x, y)\) parametrized by \((x, y) \in \Omega\). The surface area of such a graph is given by the double integral

\[
J[u] = \int \int_{\Omega} \sqrt{1 + \left( \frac{\partial u}{\partial x} \right)^2 + \left( \frac{\partial u}{\partial y} \right)^2} \, dx \, dy;
\]

(21.8)

see (B.39). To find the minimal surface, then, we seek the function \(z = u(x, y)\) that minimizes the surface area integral (21.8) when subject to the Dirichlet boundary conditions

\[
u(x, y) = g(x, y) \quad \text{for} \quad (x, y) \in \partial \Omega
\]

(21.9)

that prescribe the boundary curve \(C\). As we shall see, the solutions to this minimization problem satisfy a certain nonlinear second order partial differential equation, given in (21.50) below.

A simple version of the minimal surface problem, that still contains many interesting features, is to find minimal surfaces of revolution. Recall that 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 \(a = (a, \alpha), b = (b, \beta) \in \mathbb{R}^2\), our 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. According to Exercise \(\Box\), the area of such a surface of revolution is given by

\[
J[u] = \int_{a}^{b} 2\pi |u| \sqrt{1 + (u')^2} \, dx.
\]

(21.10)

We seek a minimizer of this integral among all functions \(u(x)\) that satisfy the boundary conditions \(u(a) = \alpha, \ u(b) = \beta\).

\[
u(a) = \alpha, \quad u(b) = \beta.
\]

The minimal surface of revolution can be physically realized by stretching a soap film between two wire circles, of radius \(\alpha\) and \(\beta\), placed a distance \(b - a\) apart. Symmetry considerations will require the minimizing surface to be rotationally symmetric. Interestingly, the revolutionary surface area functional (21.10) is exactly the same as the optical functional (21.5) when the light speed at a point is inversely proportional to its distance from the horizontal axis, namely \(c(x, y) = 1/2 \pi |y|\).

### 21.2. The Simplest Variational Problem.

Even the preceding, rather limited collection of examples of variational problems should already convince the reader of the 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 the classical approach, leaving more modern direct methods — the function space equivalent of the gradient descent method — to a more in-depth treatment of the subject, \([cvar]\).

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
the function \( y = u(x) \in C^1[a, b] \) that minimizes the **objective functional**
\[
J[u] = \int_a^b L(x, u, u') \, dx
\]
subject to certain prescribed boundary conditions. The integrand \( L(x, u, p) \) is known as the
**Lagrangian** for the variational problem, in honor of Joseph–Louis Lagrange, who was one of
the founders of the subject. We usually assume that \( L(x, u, p) \) 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 (21.3) has Lagrangian function \( L(x, u, p) = \sqrt{1 + p^2} \),
whereas in the surface of revolution problem (21.10), we have \( L(x, u, p) = 2\pi |u| \sqrt{1 + p^2} \).
(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 — that arose in Chapter 11 are also of interest here. In
the interests of brevity, we shall concentrate on the Dirichlet boundary conditions
\[
u(a) = \alpha, \quad u(b) = \beta,
\]
as these are the most common in physical problems, although some of the exercises will
investigate other types.

**The First Variation and the Euler–Lagrange Equation**

According to Section 19.3, the (local) minimizers of a (sufficiently nice) function defined
on a finite-dimensional vector space are initially characterized as critical points, where the gradient of the function vanishes. An analogous construction applies in the
infinite-dimensional context treated by the calculus of variations. Every minimizer \( u_* \) of
a sufficiently nice functional \( J[u] \) is a “critical function”, meaning that the functional gradient \( \nabla J[u_*] = 0 \) vanishes at that function. Indeed, the justification of this result
that was outlined in Section 19.3 continues to apply here; see, in particular, the proof
of Theorem 19.42. Of course, not every critical point turns out to be a minimum. In
nondegenerate situations, the classification of critical points into local minima, maxima,
or saddle points, relies on the second derivative test. The functional version of the second
derivative test — the second variation — is the topic of Section 21.3.

Thus, our first order of business is to learn how to compute the gradient of a functional
that is defined on an infinite-dimensional function space. Adapting the general
Definition 19.38 of the gradient of a function defined on an inner product space, the gradient \( \nabla J[u] \) of the functional (21.11) should be defined by the same basic formula
\[
\langle \nabla J[u] ; v \rangle = \frac{d}{dt} J[u + tv] \bigg|_{t=0}.
\]
Here \( v(x) \) is a function — 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”. The gradient operator on functionals is often referred to as the \textit{variational derivative}. The inner product used in (21.13) is taken (again for simplicity) to be the standard $L^2$ inner product
\begin{equation}
\langle f ; g \rangle = \int_a^b f(x) g(x) \, dx
\end{equation}
on function space.

Now, starting with (21.11), we have
\begin{equation}
J[u + t \, v] = \int_a^b L(x, u + t \, v, u' + t \, v') \, dx.
\end{equation}
We need to compute the derivative of the integral with respect to $t$. Assuming smoothness of the integrand allows us to bring the derivative inside the integral and so, by the chain rule,
\begin{align*}
\frac{d}{dt} J[u + t \, v] &= \int_a^b \frac{d}{dt} L(x, u + t \, v, u' + t \, v') \, dx \\
&= \int_a^b \left[ v \frac{\partial L}{\partial u} (x, u + t \, v, u' + t \, v') + v' \frac{\partial L}{\partial p} (x, u + t \, v, u' + t \, v') \right] \, dx.
\end{align*}
Therefore, setting $t = 0$ to evaluate (21.13), we find
\begin{equation}
\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.
\end{equation}
The resulting integral often referred to as the \textit{first variation} of the functional $J[u]$. The condition $\langle \nabla J[u] ; v \rangle = 0$ for a minimizer is known as the \textit{weak form} of the variational principle.

To obtain the strong form, the right hand side of (21.16) needs to be written as an inner product,
\begin{equation}
\langle \nabla J[u] ; v \rangle = \int_a^b \nabla J[u] \, v \, dx = \int_a^b h \, v \, dx
\end{equation}
between some function $h(x) = \nabla J[u]$ and the variation $v$. The first term has this form, but the derivative $v'$ appearing in the second term is problematic. However, as the reader of Chapter 11 already knows, the secret behind removing derivatives in an integral formula is integration by parts. If we set
\begin{equation}
\frac{\partial L}{\partial p} (x, u(x), u'(x)) \equiv r(x),
\end{equation}
we can re-express the offending term as
\begin{equation}
\int_a^b r(x) \, v'(x) \, dx = \left[ r(b) \, v(b) - r(a) \, v(a) \right] - \int_a^b r'(x) \, v(x) \, dx,
\end{equation}
where — again by the chain rule —

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

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

\[ u(a) = \alpha, \quad u(b) = \beta. \]

Therefore, we must make sure that the varied function \( \hat{u}(x) = u(x) + t v(x) \) remains within this space of functions, and so it must satisfy the same boundary conditions \( \hat{u}(a) = \alpha, \hat{u}(b) = \beta \). But \( u(x) \) already satisfies the boundary conditions, and so the variation \( v(x) \) must satisfy the corresponding homogeneous boundary conditions

\[ v(a) = 0, \quad v(b) = 0. \]

As a result, both boundary terms in our integration by parts formula (21.17) vanish, and we can write (21.16) as

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

We conclude that

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

This is our explicit formula for the functional gradient or variational derivative of the functional (21.11) with Lagrangian \( L(x, u, p) \). Note that the gradient \( \nabla J[u] \) of a functional is a function.

The critical functions \( u(x) \) — which include all local minimizers — are, by definition, where the functional gradient vanishes: \( \nabla J[u] = 0 \). Thus, \( u(x) \) must satisfy

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

In view of (21.18), we see that (21.21) is, in fact, a second order ordinary differential equation,

\[ E(x, u, u', u'') = \frac{\partial L}{\partial u} (x, u, u') - \frac{\partial^2 L}{\partial x \partial p} (x, u, u') - u' \frac{\partial^2 L}{\partial u \partial p} (x, u, u') - u'' \frac{\partial^2 L}{\partial p^2} (x, u, u') = 0, \]

known as the Euler–Lagrange equation associated with the variational problem (21.11). 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 desired minimizer without further ado.
Theorem 21.1. Suppose the Lagrangian function is at least twice continuously differentiable: \( L(x, u, p) \in C^2 \). Then any \( C^2 \) minimizer \( u(x) \) to the corresponding functional \( J[u] = \int_a^b L(x, u, u') \, dx \) must satisfy the associated Euler–Lagrange equation (21.21).

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 warning: there do exist seemingly reasonable functionals whose minimizers are not, in fact, \( C^2 \), and hence do not solve the Euler–Lagrange equation; see [14] for examples. Fortunately, in the problems we usually consider, such pathologies do not appear.

Curves of Shortest Length

Consider the problem of finding the curve of shortest length connecting two points \( a = (a, \alpha), b = (b, \beta) \in \mathbb{R}^2 \) in the plane. As we saw in Section 21.2, this requires minimizing the arc length integral
\[
J[u] = \int_a^b \sqrt{1 + (u')^2} \, dx \quad \text{with Lagrangian} \quad L(x, u, p) = \sqrt{1 + p^2}.
\]

Since
\[
\frac{\partial L}{\partial u} = 0, \quad \frac{\partial L}{\partial p} = \frac{p}{\sqrt{1 + p^2}},
\]
the Euler–Lagrange equation (21.21) 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, the Euler–Lagrange equation reduces to the simplest second order ordinary differential equation
\[
u'' = 0. \quad (21.22)
\]
All solutions to the Euler–Lagrange equation are affine functions, \( u = cx + d \), whose graphs are straight lines. Since our solution must also satisfy the boundary conditions \( \alpha = u(a), \beta = u(b) \), the only critical function, and hence the sole candidate to be a minimizer, is the unique straight line
\[
y = \frac{\beta - \alpha}{b - a} (x - a) + \alpha \quad (21.23)
\]
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 gives it the status of a candidate for minimizing the variational problem. By the same token, a critical function is also a candidate for maximizing the variational problem, too. The nature of the critical functions can only be distinguished by a second derivative test, which requires further work. Of course, for the present problem, we “know” that a straight line cannot maximize distance, and must be
the minimizer. Nevertheless, the reader should have a little nagging doubt that we have completely solved the minimum distance problem . . .

**Minimal Surface of Revolution**

Consider next the problem of finding the curve connecting two points having 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 (21.10), the required curve will minimize the functional

$$J[u] = \int_{a}^{b} u \sqrt{1 + (u')^2} \, dx,$$

with Lagrangian $L(x, u, p) = u \sqrt{1 + p^2}$, (21.24)

where we have dropped an irrelevant factor of $2\pi$ and used our positivity assumption to omit the absolute value on $u$ in the integrand. Since

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

the Euler–Lagrange equation (21.21) is

$$\sqrt{1+(u')^2} - \frac{d}{dx} \frac{uu'}{\sqrt{1+(u')^2}} = \frac{1+(u')^2 - uu''}{(1+(u')^2)^{3/2}} = 0.$$ (21.25)

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 by $u'$, then we can rewrite the result as an exact derivative

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

We conclude that

$$\frac{u}{\sqrt{1+(u')^2}} = c,$$ (21.26)

where $c$ is a constant of integration. The left hand side of (21.26), being constant on the entire solution, is a first integral for the differential equation, cf. Definition 20.25. The resulting equation is an implicit form of an autonomous first order differential equation. Solving for

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

leads to an autonomous first order ordinary differential equation, which we can immediately solve:

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

\[\uparrow\] 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, [64, 117], underlies the integration method. See also Exercise 2/25/04 935 © 2004 Peter J. Olver
where $\delta$ is a constant of integration. According to Exercise 1, the most useful form of the integral is in terms of the inverse to the hyperbolic function $cosh \, z = \frac{1}{2}(e^z + e^{-z})$, whereby

$$cosh^{-1} \frac{u}{c} = x + \delta, \text{ and hence } u = c \, cosh \left( \frac{x + \delta}{c} \right). \tag{21.27}$$

In this manner, we have produced the general solution to the Euler–Lagrange equation (21.25). Any solution that also satisfies the boundary conditions provides a critical function for the surface area functional (21.24), and hence is a candidate for the minimizer.

The curve prescribed by the graph of a hyperbolic cosine function (21.27) is known as a catenary. It is not a parabola, even though to the untrained eye it looks similar. Interestingly, the catenary is the same profile as a hanging chain. Owing to their minimizing properties, catenaries are quite common in engineering design — for instance the cables in a suspension bridge such as the Golden Gate Bridge are catenaries, as is the arch in St. Louis.

So far, we have not taken into account the boundary conditions $u(a) = \alpha$ and $u(b) = \beta$. It turns out that there are three distinct possibilities, depending upon the configuration of the boundary points:

(a) There is precisely one value of the two integration constants $c, \delta$ 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 $c, \delta$ 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 functional.

(c) There are no values of $c, \delta$ that allow (21.27) to satisfy the two boundary conditions. This occurs when the two boundary points $a, b$ 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., there is no surface of revolution that has a smaller surface area than the two disks. (In the first two cases, this is not valid; the minimizing catenary has a smaller surface area than the two disks.) However, the “function”‡ that minimizes this configuration consists of two vertical lines from $a$ and $b$ to the $x$ axis along with the portion of the axis lying between them. We can approximate this function by a sequence of genuine functions that give progressively smaller and smaller values to the surface area functional (21.10), 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 arising in the calculus of variations can already be seen. Lack of space precludes a more detailed development of these ideas here, and we refer the interested reader to more specialized books devoted to the calculus of variations, including [39, 64].

‡ Here “function” must be taken in a very broad sense, as this situation does not even correspond to a generalized function!
The Brachistochrone Problem

The most famous classical variational problem is the so-called brachistochrone problem. The word “brachistochrone” means “minimal time” in Latin. An experimenter lets a bead slide down a wire that connects two prescribed 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 shape, including a straight line, a parabola, and a circular arc, are wrong. One can do better through a careful analysis of the associated variational problem. The brachistochrone problem was originally posed by Johann Bernoulli in 1696, and served as an inspiration for much of the subsequent development of the subject.

We take the starting point of the bead at the origin: \( a = (0,0) \). The wire will bend downwards, and to avoid annoying minus signs in the subsequent formulae, we take the vertical \( y \) axis to point downwards, so the wire has the shape given by the graph of \( y = u(x) > 0 \). The end point \( b = (b, \beta) \) is assumed to lie below and to the right, and so \( b > 0 \) and \( \beta > 0 \); the set-up is sketched in Figure 21.3. 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 functional (21.5), if \( v \) denotes the speed of descent of the bead, then the total travel time is

\[
T[u] = \int_0^b \frac{\sqrt{1 + (u')^2}}{v} \, dx.
\]  

(21.28)

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 and \( v \geq 0 \) its speed of descent. On the other hand, due to our sign convention, the potential energy of the bead when it is at height \( y \) is \( -mg y \), where \( m \) is its mass and \( g \) the gravitational force, and we take the initial height \( y = 0 \) 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

\[
0 = \frac{1}{2} m v^2 - m g y.
\]

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

\[
v = \sqrt{2 g y}.
\]  

(21.29)
Substituting this expression into (21.28), we conclude that the shape \( y = u(x) \) of the wire is obtained by minimizing the functional

\[
T[u] = \int_0^b \sqrt{\frac{1 + (u')^2}{2gu}}
\]

The associated Lagrangian is

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

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

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

Therefore, the Euler–Lagrange equation for the brachistochrone functional (21.30) is

\[
-\frac{\sqrt{1 + (u')^2}}{2 u^{3/2}} - \frac{d}{dx} \frac{u'}{\sqrt{u(1 + (u')^2)}} = -\frac{2 uu'' + (u')^2 + 1}{2 \sqrt{u(1 + (u')^2)}} = 0,
\]

and is equivalent to the nonlinear second order ordinary differential equation

\[
2 uu'' + (u')^2 + 1 = 0.
\]

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 result of Exercise \( \blacksquare \) that states that the Hamiltonian

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

is a first integral, and hence

\[
\frac{1}{\sqrt{u(1 + (u')^2)}} = k, \quad \text{which we rewrite as} \quad u(1 + (u')^2) = c,
\]

where \( c = 1/k^2 \) is a constant. 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, integrating from the initial point \( x = u = 0, \)

\[
\int_0^u \frac{u}{c - u} \, du = x.
\]

The integration can be done by use of a trigonometric substitution, namely

\[
u = \frac{1}{2} c (1 - \cos r), \quad \text{whereby} \quad x = c \int_0^r (1 - \cos r), \, dr = \frac{1}{2} c (r - \sin r). \quad (21.31)
\]
The resulting pair of equations (21.31) serve to parametrize a curve \((x(r), u(r))\) known as a \textit{cycloid}. According to Exercise 21.1, a cycloid can be visualized as the curve that is traced by a point sitting on the edge of a rolling wheel. Thus, all solutions to the Euler–Lagrange equation are the cycloids, described in parametric form by (21.31). Any cycloid which satisfies the boundary conditions supplies us with a critical function, and hence a candidate for the solution to the brachistochrone minimization problem.

With a little more work, it can be proved that there is precisely one value of the integration constant \(c\) that satisfies the two boundary conditions, and, moreover, that this particular cycloid minimizes the brachistochrone functional. An example of a cycloid is plotted in Figure 21.4. Interestingly, in certain configurations, namely if \(\beta < 2b/\pi\), the cycloid that solves the brachistochrone problem dips below the lower endpoint \(b\).

\section{21.3. 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. In the finite-dimensional theory, 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 different possible solutions, we need to formulate a second derivative test for the objective functional on an infinite-dimensional function space. In the calculus of variations, the second derivative of a functional is known as its \textit{second variation}, the Euler–Lagrange expression being also known as the \textit{first variation}.

In the finite-dimensional version, the second derivative test was based on the positive definiteness of the Hessian matrix. The justification relied on a second order Taylor expansion of the objective function at the critical point. Thus, in an analogous fashion, we expand the objective functional \(J[u]\) near the critical function. Consider the scalar function \(g(t) = J[u + t v]\), where the function \(v(x)\) represents a variation. The second order Taylor expansion of \(g(t)\) takes the form

\[ g(t) = J[u + t v] = J[u] + t K[u; v] + \frac{1}{2} t^2 Q[u; v] + \cdots . \]

The first order terms are linear in the variation \(v\), and given by an inner product

\[ g'(0) = K[u; v] = \langle \nabla J[u]; v \rangle \]
between the variation and the functional gradient. In particular, if \( u = u^* \) 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 — is, in most cases, determined by the second derivative terms

\[
g''(0) = Q[u^*; v].
\]

As in the finite-dimensional Theorem 19.45, if \( u \) is a minimizer, then \( Q[u; v] \geq 0 \). Conversely, if \( Q[u; v] > 0 \) for \( v \neq 0 \), i.e., the second derivative terms satisfy a condition of positive definiteness, then \( u \) will be a strict local minimizer. This forms the crux of the second derivative test.

Let us explicitly evaluate the second derivative terms for the simplest variational problem (21.11). We need to expand the scalar function

\[
g(t) = J[u + t v] = \int_a^b L(x, u + t v, u' + tv') \, dx
\]

in a Taylor series around \( t = 0 \). The linear terms in \( t \) were already found in (21.16), and so we need to compute the quadratic terms:

\[
Q[u; v] = g''(0) = \int_a^b \left[ A v^2 + 2 B v v' + C (v')^2 \right] \, dx,
\]

(21.32)

where the coefficient functions

\[
A(x) = \frac{\partial^2 L}{\partial u^2} (x, u, u'), \quad B(x) = \frac{\partial^2 L}{\partial u \partial p} (x, u, u'), \quad C(x) = \frac{\partial^2 L}{\partial p^2} (x, u, u'),
\]

(21.33)

are found by evaluating certain second order derivatives of the Lagrangian at the critical function \( u(x) \). The quadratic functional (21.32) is known as the second variation of the original functional \( J[u] \), and plays the role of the Hessian matrix for functionals. In contrast to the first variation, it is not possible to eliminate all of the derivatives on \( v \) in the quadratic functional (21.32) through integration by parts. This causes significant complications for the analysis.

To formulate conditions that the critical function be a minimizer for the functional, we need to determine when such a quadratic functional is positive definite, meaning that \( Q[u; v] > 0 \) for all non-zero allowable variations \( v(x) \neq 0 \). Clearly, if the integrand is positive definite at each point, so

\[
A(x) v^2 + 2 B(x) v v' + C(x) (v')^2 > 0 \quad \text{for all} \quad a \leq x \leq b, \quad (v, v') \neq 0,
\]

(21.34)

then the second variation \( Q[u; v] \) is also positive definite.
**Example 21.2.** For the arc length minimization functional (21.3), the Lagrangian is

\[ L(x, u, p) = \sqrt{1 + p^2}. \]

To analyze the second variation, we first compute

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

For the critical straight line function \( u = u^* \) given in (21.23), we evaluate at \( p = u' = (\beta - \alpha)/(b - a) \), and so

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

Therefore, the second variation functional (21.32) is

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

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

In general, as the following example points out, the pointwise positivity condition (21.34) is overly restrictive.

**Example 21.3.** Consider the quadratic functional

\[ Q[v] = \int_0^1 [(v')^2 - v^2] \, dx. \tag{21.35} \]

The claim is that \( Q[v] > 0 \) is positive definite for all nonzero \( v \neq 0 \) subject to homogeneous Dirichlet boundary conditions \( v(0) = v(1) = 0 \). This result is not trivial! Indeed, the boundary conditions play an essential role, since choosing \( v(x) \equiv c \) 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

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

which is clearly positive semi-definite since the integrand is everywhere \( \geq 0 \); moreover, the integral vanishes if and only if \( v \) satisfies the first order linear ordinary differential equation

\[ v' + v \tan x = 0, \quad \text{for all} \quad 0 \leq x \leq 1. \]

The only solution that also satisfies boundary condition \( v(0) = v(1) = 0 \) is the trivial one \( v \equiv 0 \). We conclude that \( \tilde{Q}[v] = 0 \) if and only if \( v \equiv 0 \), and hence \( \tilde{Q} > 0 \) is a positive definite quadratic functional.
Let us expand the latter functional,
\[ 
\tilde{Q}[v] = \int_0^1 \left[ (v')^2 + 2vv' \tan x + v^2 \tan^2 x \right] dx 
\]
\[ = \int_0^1 \left[ (v')^2 - v^2 (\tan x)' + v^2 \tan^2 x \right] dx = \int_0^1 \left[ (v')^2 - v^2 \right] dx = Q[v]. 
\]

In the second equality, we integrated the middle term by parts, using \((v')' = 2vv'\), and noting that the boundary terms vanish. Since \(\tilde{Q}[v]\) is positive definite, so is \(Q[v]\), justifying the previous claim.

To see how subtle this result is, consider the almost identical quadratic functional
\[ 
\hat{Q}[v] = \int_0^4 \left[ (v')^2 - v^2 \right] dx. \quad (21.36) 
\]

The only difference is in the upper limit to the integral. A quick computation shows that the function \(v(x) = x(4 - x)\) satisfies the homogeneous Dirichlet boundary conditions \(v(0) = 0 = v(4)\), but
\[ 
\hat{Q}[v] = \int_0^4 \left[ (4 - 2x)^2 - x^2 (4 - x)^2 \right] dx = -\frac{128}{5} < 0. 
\]

Therefore, \(\hat{Q}[v]\) 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 \(\int_0^1 (v' + v\tan x)^2 dx\) does not converge.

The complete analysis of positive definiteness of quadratic functionals is quite subtle. The strange appearance of \(\tan x\) in this particular 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 [64] for full details.

**Theorem 21.4.** Let \(A(x), B(x), C(x) \in C^0[a, b]\) be continuous functions. The quadratic functional
\[ 
Q[v] = \int_a^b \left[ Av^2 + 2Bvv' + C(v')^2 \right] dx 
\]
is positive definite, so \(Q[v] > 0\) for all \(v \neq 0\) satisfying the homogeneous Dirichlet boundary conditions, provided

(a) \(C(x) > 0\) for all \(a \leq x \leq b\), and

(b) For any \(a < c \leq b\), the only solution to the associated linear Euler–Lagrange boundary value problem
\[ 
-(C w')' + (A - B') w = 0, \quad w(a) = w(c), \quad w'(a) = w'(c), 
\]
is the trivial function \(w(x) \equiv 0\).
Remark: A value $c$ for which (21.37) 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 21.5. The quadratic functional

$$Q[v] = \int_0^b \left[ (v')^2 - v^2 \right] dx \quad (21.38)$$

has Euler–Lagrange equation

$$-w'' - w = 0.$$  

The solutions $w(x) = k \sin x$ satisfy the boundary condition $w(0) = 0$. The first conjugate point occurs at $c = \pi$ where $w(\pi) = 0$. Therefore, Theorem 21.4 implies that the quadratic functional (21.38) is positive definite provided the upper integration limit $b < \pi$. This explains why the first quadratic functional (21.35) is positive definite, since there are no conjugate points on the interval $[0, 1]$, while the second (21.36) is not because the first conjugate point $\pi$ lies on the interval $[0, 4]$.

In the case when the quadratic functional arises as the second variation of a functional (21.11), then the coefficient functions $A, B, C$ are given in terms of the Lagrangian $L(x, u, p)$ by formulae (21.33). In this case, the first condition in Theorem 21.4 requires

$$\frac{\partial^2 L}{\partial p^2} (x, u, u') > 0 \quad (21.39)$$

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

$$- \frac{d}{dx} \left( \frac{\partial^2 L}{\partial p^2} (x, u, u') \frac{dw}{dx} \right) + \left( \frac{\partial^2 L}{\partial u^2} (x, u, u') - \frac{d}{dx} \frac{\partial^2 L}{\partial u \partial p} (x, u, u') \right) w = 0 \quad (21.40)$$

has no nontrivial solutions $w(x) \neq 0$ that satisfy $w(a) = 0$ and $w(c) = 0$ for $a < c \leq b$.


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, in which the equilibrium configurations minimize a suitable functional — typically, the potential energy of the system. The minimizing configurations are among the critical points of the functional where its functional gradient vanishes. Following similar computational procedures as in the simple one-dimensional version, we find that the critical functions are characterized as solutions to a system of partial differential equations, called the Euler–Lagrange equations associated with the variational principle. Each solution to the boundary value problem specified by the Euler–Lagrange equations is, thus, a candidate minimizer for the variational problem. In many applications, the Euler–Lagrange equations suffice to single out the desired physical solutions, and one does not continue 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. Thus, we consider a functional in the form

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

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 given functional among all sufficiently smooth functions that satisfy a set of prescribed boundary conditions on \( \partial \Omega \). The most important are our usual Dirichlet, Neumann and mixed boundary conditions. For simplicity, we concentrate on the Dirichlet boundary value problem

\[ u(x, y) = g(x, y) \quad \text{for} \quad (x, y) \in \partial \Omega. \]  

**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 function

\[ g(t) \equiv J[u + t v] = \iint_{\Omega} L(x, y, u + t v_x, u + t v_y, u_x + t v_x, u_y + t v_y) \, dx \, dy \]

for \( t \in \mathbb{R} \). The variation \( v(x, y) \) is assumed to satisfy homogeneous Dirichlet boundary conditions

\[ v(x, y) = 0 \quad \text{for} \quad (x, y) \in \partial \Omega, \]  

(21.43)

to ensure that \( u + t v \) satisfies the same boundary conditions (21.42) as \( u \) itself. Under these conditions, if \( u \) is a minimizer, then the scalar function \( g(t) \) will have a minimum at \( t = 0 \), and hence \( g'(0) = 0 \). When computing \( g'(t) \), we assume that the functions involved are sufficiently smooth so as to allow us to bring the derivative \( d/\,dt \) inside the integral and then apply the chain rule. At \( t = 0 \), the result is

\[ g'(0) = \left. \frac{d}{dt} J[u + t v] \right|_{t=0} = \iint_{\Omega} \left( v \frac{\partial L}{\partial u} + v_x \frac{\partial L}{\partial p} + v_y \frac{\partial L}{\partial q} \right) \, dx \, dy, \]

(21.44)

where the derivatives of \( L \) are 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

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

As before, we need to remove the offending derivatives from \( v \). In two dimensions, the requisite integration by parts formula

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

(21.45)
in which $w_1, w_2$ are arbitrary smooth functions, appears in (15.79). Setting $w_1 = \partial L/\partial p, w_2 = \partial L/\partial q$, 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 when $v(x, y)$ satisfies the homogeneous Dirichlet boundary conditions (21.43) that we impose on the allowable variations. Substituting this result back into (21.44), we conclude that

$$
g'(0) = \iint_{\Omega} v \left[ \frac{\partial L}{\partial u} - \frac{\partial}{\partial x} \left( \frac{\partial L}{\partial p} \right) - \frac{\partial}{\partial y} \left( \frac{\partial L}{\partial q} \right) \right] dx \, dy = 0. \quad (21.46)
$$

The quantity in brackets is the desired first variation or functional gradient:

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

which must vanish at a critical function. We conclude that the minimizer $u(x, y)$ must satisfy the Euler–Lagrange equation

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

(21.47)

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

$$
L_u - L_x p - L_y q - u_x L_{up} - u_y L_{uq} - u_{xx} L_{pp} - u_{xy} L_{pq} - u_{yy} L_{qq}, \quad (21.48)
$$

where we use subscripts to indicate derivatives of both $u$ and $L$, the latter being evaluated at $x, y, u, u_x, u_y$. Solutions to the Euler–Lagrange equation are critical functions for the variational problem, and hence include any 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 very complicated, and still awaits a completely satisfactory resolution!

**Example 21.6.** As a first elementary example, consider the Dirichlet minimization problem

$$
J[u] = \iint_{\Omega} \frac{1}{2} \left( u_x^2 + u_y^2 \right) dx \, dy \quad (21.49)
$$

that we first encountered in our analysis of the solutions to the Laplace equation (15.91). In this case, the associated Lagrangian is

$$
L = \frac{1}{2}(p^2 + q^2), \quad \text{with} \quad \frac{\partial L}{\partial u} = 0, \quad \frac{\partial L}{\partial p} = p = u_x, \quad \frac{\partial L}{\partial q} = q = u_y.
$$
Therefore, the Euler–Lagrange equation (21.47) 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 boundary conditions, the solutions, i.e., the harmonic functions, are the critical functions for the Dirichlet variational principle. This reconfirms the Dirichlet characterization of harmonic functions as minimizers of the variational principle, as stated in Theorem 15.13. 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. In the quadratic case, 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 21.7. Let us derive the Euler–Lagrange equation for the minimal surface problem. From (21.8), the surface area integral

\[ J[u] = \int_{\Omega} \sqrt{1 + u_x^2 + u_y^2} \, dx \, dy \]

has Lagrangian \[ L = \sqrt{1 + p^2 + q^2} \].

Note that

\[ \frac{\partial L}{\partial u} = 0, \quad \frac{\partial L}{\partial p} = \frac{p}{\sqrt{1 + p^2 + q^2}}, \quad \frac{\partial L}{\partial q} = \frac{q}{\sqrt{1 + p^2 + q^2}}. \]

Therefore, replacing \( p \rightarrow u_x \) and \( q \rightarrow u_y \) and then evaluating the derivatives, the Euler–Lagrange equation (21.47) becomes

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

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

\[(1 + u_y^2) u_{xx} - 2 u_x u_y u_{xy} + (1 + u_x^2) u_{yy} = 0. \] (21.50)

Thus, 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 preceding two centuries, with significant progress on understanding its solution only within the past 70 years. We have not developed the sophisticated analytical and numerical techniques that are required to have anything of substance to say about its solutions here, and will refer the interested reader to the advanced texts [109, 112].
Example 21.8. The small deformations of an elastic body $\Omega \subset \mathbb{R}^n$ are described by the *displacement* field, $u: \Omega \to \mathbb{R}^n$. Each material point $x \in \Omega$ in the undeformed body will move to a new position $x + u(x)$ in the deformed body $\Omega = \{ x + u(x) | x \in \Omega \}$. 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 [8, 69] 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 an isotropic, homogeneous planar body $\Omega \subset \mathbb{R}^2$, i.e., a thin plate. The equilibrium mechanics are described by the deformation function $u(x) = (u(x,y), v(x,y))^T$. A detailed physical analysis of the constitutive assumptions leads to the following minimization principle

\[
J[u,v] = \int \int_\Omega \left[ \frac{1}{2} \mu \| \nabla u \|^2 + \frac{1}{2} (\lambda + \mu)(\nabla \cdot u)^2 \right] dx \, dy \tag{21.51}
\]

The parameters $\lambda, \mu$ 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, (21.51) 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 $g(t) = J[u + tf, v + tg]$, in which the individual variations $f, g$ are arbitrary functions subject only to the given homogeneous boundary conditions. If $u, v$ minimize $J$, then $g(t)$ has a minimum at $t = 0$, and so we are led to compute

\[
g'(0) = \langle \nabla J; f \rangle = \int \int_\Omega (f \nabla_u J + g \nabla_v J) \, dx \, dy,
\]

which we write as an inner product (using the standard $L^2$ inner product between vector fields) between the variation $f$ and the functional gradient $\nabla J = (\nabla_u J, \nabla_v J)^T$. For the particular functional (21.51), we find

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

We use the integration by parts formula (21.45) to remove the derivatives from the variations $f, g$. Discarding the boundary integrals, which are used to prescribe the allowable boundary conditions, we find

\[
g'(0) = - \int \int_\Omega \left( \left[ (\lambda + 2\mu) u_{xx} + \mu u_{yy} + (\lambda + \mu) v_{xy} \right] f + \left[ (\lambda + \mu) u_{xy} + \mu v_{xx} + (\lambda + 2\mu) v_{yy} \right] g \right) \, dx \, dy.
\]

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

\[(\lambda + 2\mu) u_{xx} + \mu u_{yy} + (\lambda + \mu) v_{xy} = 0, \quad (\lambda + \mu) u_{xy} + \mu v_{xx} + (\lambda + 2\mu) v_{yy} = 0, \tag{21.52}\]
known as Navier’s equations, which can be compactly written as

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

(21.53)

for the displacement vector \( \mathbf{u} = (u, v)^T \). 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 (21.51). 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.