PART I

VARIATIONAL METHODS

1 Preliminaries

The vessels, heavy laden, put to sea
With prosp'rous winds; a woman leads the way.
I know not, if by stress of weather driv'n,
Or was their fatal course dispos'd by Heav'n;
At last they landed, where from far your eyes
May view the turrets of new Carthage rise;
There bought a space of ground, which (Byrsa call'd,
From the bull's hide) they first inclos'd, and wall'd.

(Virgil, *The Aeneid*)

The above excerpt cryptically recounts the legend of Dido, the Queen of Carthage, from the ninth century BC. After being exiled from Tyre in Lebanon, Dido purportedly sailed to the shores of North Africa (now Tunisia) and requested that the local inhabitants give her and her party the land that could be enclosed by the hide of an ox. Not thinking that an oxhide could encompass a large portion of land, they granted her wish. She then proceeded to have the hide cut into narrow strips and extended end-to-end to form a semicircle bounding the shoreline and encompassing a nearby hill. This area became known as Carthage.

Certain branches of mathematics have arisen out of consideration of the theoretical consequences of known mathematical theorems. More often than not, however, new branches of mathematics have been developed to provide the means to address certain types of practical problems that initially are often very specific. For example, how did Dido know to arrange the oxhide in a circular shape in order to enclose the largest possible area with a given perimeter length?

One of the best examples of a branch of mathematics motivated by applications, and the one required to solve the problem of Dido, is the *calculus of variations*, which today encompasses numerous, far-reaching applications. Its application-driven roots live on in several fundamental variational principles that form the foundation of a number of important fields in the physical sciences and engineering. A glance through the table of contents of this text will hint at the variety of topics that can be treated using variational methods. This will become more evident in Section 1.3 as we motivate the need for variational calculus through consideration of three practical problems and by elucidation of a variety of physical phenomena and optimization principles through a unified variational framework throughout the remainder of the text.

1.1 A Bit of History

To fully appreciate our treatment of the calculus of variations, a bit of history is in order.[1] Doing so will assist us in placing the principles, methods, and applications to be encountered in a useful framework. In addition, our discussions will be replete with some of the greatest names in mathematics and physics, so it is hoped that a brief history will enrich our study of these topics and the people behind them.

Each era of human existence is marked to some extent by the scientific and technological problems of the day, and it is always the case that some of the most fertile minds are drawn to addressing the greatest challenges and grandest questions. After seminal contributions by the Greek philosophers in the sixth through fourth centuries BC and the Roman emphasis on infrastructure and engineering for the following 800 years, high-level intellectual pursuits were largely abandoned for a number of centuries while Western society was dominated by instability and decentralized power during the Middle Ages. Whereas this period was marked by the need to accurately launch projectiles into neighboring feudal castles, the subsequent periods of the Renaissance and Age of Enlightenment began a scientific revolution that we are still experiencing today. This period led to mathematical contributions, scientific progress, artistic expression, architectural prowess, and musical genius that have stood the test of time well into the modern – and now postmodern – era.

Coming out of the Middle Ages at the dawn of the sixteenth century, the Church provided needed stability, but it also wielded excessive power over the largely uneducated and illiterate populace throughout Europe. While the spiritual authority of the Church was being challenged by the reformers, such as Martin Luther, its scientific authority was being challenged by the increasingly rigorous methods of Nicolaus Copernicus and Galileo Galilei based on observation and experimentation. The Renaissance also saw a return to classical Greek philosophy and the Greek emphasis on realism in art and expression. Moreover, it was a time in which integration and harmonization of broad endeavors were encouraged, giving rise to the "Renaissance man." While Leonardo da Vinci was the quintessential Renaissance man of Italy, the center of gravity of the Renaissance, Christopher Wren exemplified the British variant. Best known for being the architect of St. Paul's Cathedral in London (and approximately fifty other churches after the Great Fire of London in 1666), he was also a noted astronomer, physicist, and mathematician, as well as the founder of the Royal Society of London.

Much of the most celebrated art, architecture, and music of the time was commissioned by or directed toward the Church, reflecting an overarching emphasis on unifying and integrating various human intellectual, spiritual, and artistic pursuits. There was a developing sense that an intrinsic order, beauty, and simplicity governs both the universe and the highest human endeavors. The Renaissance produced artists, musicians, scientists, and mathematicians emboldened by the potential contributions that they as individuals and as humans collectively could make in all human undertaking.

[1] Our brief history is admittedly European-centric, as this region provided the immediate historical context for the development of variational methods.

Released from the patriarchal constraints on scientific thought rendered by the Church and emboldened with a strong sense of realism and drive toward harmonization of all aspects of life by the Renaissance artists and thinkers, the Age of Enlightenment was born. While the Renaissance up through the seventeenth century was marked primarily by artistic and musical innovation, a revolution in science and mathematics began that laid the foundation for the Age of Enlightenment, which spanned the eighteenth century and led to unprecedented mathematical and scientific progress. Both the Renaissance and Age of Enlightenment shared the ideals of realism in which objects, thought, and indeed the entire universe were viewed, rendered, and modeled as they really were.

The Age of Enlightenment brought a renewed emphasis on how to explain the existence and operation of the universe, and this required a new set of mathematical tools. An increased priority was placed on providing explanations and mathematical models to elucidate observed phenomena. Together these led to an explosion of development and discovery in science and mathematics, the two feeding off of each other's questions and contributions. Much of today's mathematical arsenal had its origins during the Age of Enlightenment, including complex variables, differential equations, and, most notably, differential calculus.

The stage was set for tremendous advances during the Age of Enlightenment by the work of Isaac Newton and the formation of scientific societies in the latter part of the seventeenth century. These societies provided a means for interaction among mathematicians and scientists and the broader communication of findings through printed journals, and Newton's work inspired a renewed search for universal physical principles and the mathematics to describe them. In particular, Newton's *Principia* (*Mathematical Principles of Natural Philosophy*) (1687) had several important influences beyond the actual articulation of his laws of motion and gravity:

1. The accurate prediction of the movement of the celestial bodies must have seemed almost magical to a world accustomed to its scientists primarily providing observations, with only occasional explanation.
2. The realization that the same physical principles that govern the falling of an apple also govern the movement of those celestial bodies throughout the universe accelerated the human pursuit of a unified theory that explains all observations in our universe.
3. A recognition of the power of mathematics to provide the unifying language for such models and laws.
4. Newton's methods formed the basis of the scientific method.

These four influences of Newton's *Principia* emboldened scientists and mathematicians for centuries to pursue a complete understanding of the universe. As concluded by Greer (2005),[2] "[Newton] demonstrated that, on the basis of experiments conducted in a tiny corner of the universe, and aided by the lever of mathematics, he had discovered the nature of gravitation *everywhere*."

In large part, Newton inaugurated the Age of Enlightenment with his physical laws articulated in the *Principia* and his – and Leibniz's – calculus. Calculus provided the mathematical tools, and his physical laws provided the framework and impetus

[2] *A Brief History of the Western World* by Thomas H. Greer.

for further developments in the ensuing years. The attitudes and successes of the Age of Enlightenment fostered a search for the ultimate truths that govern the physical world in which we live and the reasons for its existence. Permeating many aspects of life and thought was a renewed sense of order, idealism, and perfection. Scientists concluded that because the universe is ordered and behaves in a rational manner, it is amenable to explanation via universal laws expressed mathematically. It is into this world seeking order and beauty in its many forms that the contributions of Pierre Louis Moreau de Maupertuis, Leonhard Euler, and Joseph Louis Lagrange, among others, were made.

Coupled with Newton's influence on our understanding of the mechanics of moving bodies, and the ancillary philosophical contributions that his laws supplied, was Pierre de Fermat's principle asserting that light optimizes its path according to a minimum principle. These realizations inaugurated a search for other minimum principles that may govern the universe and gave rise to what we now call *optimization*. Today, we would classify Dido's problem as an optimization problem, in which we seek the "best" or "optimal" means of accomplishing a task (maximizing the land grab to form Carthage). Problems such as this occupied the efforts of some of the greatest minds of the late seventeenth and early eighteenth centuries. The differential calculus of Isaac Newton and Gottfried Wilhelm Leibniz succeeded in solving many such problems; however, some of them demanded a new calculus, later to be labeled *calculus of variations*. Some of the early optimization problems that motivated the development of the calculus of variations include:

- The shape of an object moving through a fluid to minimize resistance (Isaac Newton in 1687). See Section 10.2.
- The brachistochrone problem to determine the shape of a wire to minimize travel time of a sliding mass (Johann Bernoulli, Gottfried Wilhelm Leibniz, Isaac Newton, Jacob Bernoulli, and Guillaume de l'Hôpital in 1696). See Section 2.2.
- Minimal surface shapes (Leonhard Euler in 1744). See Section 2.5.2.

Comprehensive treatment of optimization problems had to await the arrival of the digital computer in the latter half of the twentieth century. In the intervening two hundred years, however, the calculus of variations revolutionized mechanics and physics through the development and application of de Maupertuis's least action principle, later refined to become Hamilton's principle, which will occupy our considerations in Part II of this text. It is remarkable that this mathematical framework, marked by an enduring elegance, is capable of encapsulating many physical and optimization principles of which its inventors could have never conceived. This is the "miracle of the appropriateness of the language of mathematics to the formulation of the laws of physics" that Physics Nobel Laureate Eugene Wigner spoke of in 1960. The "miracle" of variational methods is stronger than ever today, in particular as formal optimization techniques permeate more and more areas of science, engineering, and economics. As such, the calculus of variations has come full circle; originally conceived to formally address a limited set of early optimization problems, variational methods now provide the mathematical foundation and framework to treat modern, large-scale optimization problems. Every day we optimize our decisions, activities, processes, and designs; therefore, variational methods are proving to be an indispensable tool in providing a formal basis for optimization along with providing the foundation for an impressive array of physical principles that encompass classical mechanics and modern physics.

1.2 Introduction

With respect to its strong ties to applications, this treatment of the calculus of variations will emphasize three themes. First, formulating many problems in the physical sciences and engineering from first principles[3] leads naturally to variational forms. Second, these variational forms often provide additional physical insight that is not readily apparent from the equivalent, typically more familiar, differential formulation. Third, the variational approach to optimization and control supplies a general and formal framework within which to apply such principles to a broad spectrum of diverse fields. Variational methods furnish the mathematical tools to both encapsulate a wide variety of physical systems and processes under a unified principle, known as Hamilton's principle, and to provide a framework for their optimization and control. This is because the fundamental principles underlying so much of classical and modern physics, as well as optimization and control, are based on extremum principles to which the calculus of variations supplies an unrivaled mathematical framework that is tailor-made for their treatment.

This text can be divided into three distinct parts: Part I addresses the variational methods themselves and Parts II and III treat the applications of such methods. In the remainder of Chapter 1, three examples are formulated to instill the need for the calculus of variations to the reader who is new to the subject. The first two examples are based on physical principles, and the third is an optimization problem. In addition, a review of material from differential calculus related to finding extrema of functions with and without constraints is provided. This material serves as the basis for developing similar techniques for functionals in the calculus of variations. There is also a review of integration by parts and a proof of the fundamental lemma of the calculus of variations along with a review of self-adjoint differential operators. Chapter 2 encompasses all of the mathematical techniques necessary to solve these examples and treat numerous other problems that arise in the calculus of variations. The applications-minded reader will be pleased to note that the relatively compact Chapter 2 contains all of the necessary material required for a scientist or engineer to become proficient in applying these variational methods to practical and research areas that utilize variational calculus. Part I closes with a brief account of approximate and numerical methods as applied to variational methods in Chapter 3. In Part II, encompassing Chapters 4 through 9, we take a rather long journey through the application of variational methods to physical principles. Chapter 4 provides the physical centerpiece, Hamilton's principle, that forms the foundation for the physical applications treated throughout Part II. The remaining chapters in this part address classical mechanics of nondeformable and deformable bodies, stability of dynamical systems, optics and electromagnetics, modern physics, and fluid mechanics. At the risk of overwhelming the reader, the breadth of these

[3] When we use the phrase "first principles" in physical contexts, as is the case here, we mean starting directly from established physical laws, that is, without assumptions, approximations, or empirical modeling.

topics is intended to illustrate the range and variety of fields that yield to variational methods. Part III of the text is devoted to applying the calculus of variations to a variety of applications in optimization. These applications include optimization and control, image processing, data analysis, and grid generation for numerical methods. Whereas Chapter 2 contains a cohesive treatment of the fundamental mathematical methods required, which is of use to all readers, the subsequent applications-oriented material is designed such that each chapter largely stands on its own and does not depend on material other than that in Chapter 2 (and Chapter 4 for Chapters 5 through 9, and Chapter 10 for Chapters 11 and 12). Interestingly, the ordering of the topics of this text, which is intended to reflect a logical progression through the material, also roughly follows their chronological development historically.

1.3 Motivation

Before developing the mathematical techniques required to solve variational problems, it is instructive to formulate a few such problems. In doing so, we will illustrate that formulating these problems from first principles leads naturally to variational forms, thereby motivating the need for an ability to solve variational problems. We consider two physical scenarios and one optimization problem that will illustrate what a functional is and that will highlight some of the issues requiring attention in the remainder of the text.

1.3.1 Optics

Optics is an excellent example in which to illustrate the difference between differential and variational calculus. We utilize *differential calculus* when finding *extrema*, that is, minimums and maximums, *of functions*. In this case we differentiate the function, set it equal to zero, and seek the values of the independent variable(s) where the function is an extremum. As you will see, *variational calculus* is utilized when finding *extrema* of *functionals*, where a functional is a definite integral involving an unknown function and its derivatives. In the case of calculus of variations, we seek the function for which the functional is an extremum.

It will be shown that when we apply Fermat's principle of optics, which states that the path of light between two points is the one that requires the minimum travel time,[4] results in a problem in *differential calculus* if the media through which the light is traveling are homogeneous and a *variational calculus* problem if the medium is nonhomogeneous.

To apply Fermat's principle, we seek to minimize the travel time $T[u(x)]$ of light along all the possible paths $u(x)$ that the light could take through a medium between the points x_0 and x_1 at times t_0 and t_1, respectively. This requires us to minimize

$$T[u(x)] = \int_{t_0}^{t_1} dt = \int_{x_0}^{x_1} \frac{ds}{v(x,u)} = \int_{x_0}^{x_1} \frac{n(x,u)}{c} ds, \tag{1.1}$$

[4] In fact, the travel time is not always a minimum, which is an important subtlety that we will return to in due course.

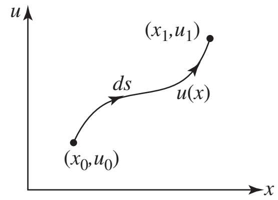

Figure 1.1. Path of light through a medium with variable index of refraction.

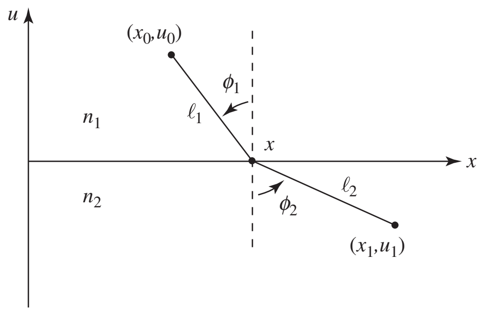

Figure 1.2. Path of light through two homogeneous media.

where $v(x,u)$ is the speed of light in the medium, c is the speed of light in a vacuum, $n(x,u)$ is the index of refraction in the medium, and ds is a differential element along the path $u(x)$ as illustrated in Figure 1.1. Equation (1.1) sums (through integration) the time it takes light to travel along each infinitesimally short element ds of the path $u(x)$ to obtain the total travel time $T[u(x)]$.

First, let us consider light traveling through two homogeneous media that each have constant index of refraction with their interface coinciding with the x-axis as shown in Figure 1.2. Because the indices of refraction n_1 and n_2 are constant in their respective media, the light travels along a straight path through each medium. That is, the path having the shortest travel time is the same as that having the shortest length in a homogeneous medium. For fixed end points (x_0,u_0) and (x_1,u_1), therefore, we seek to find the value of x where the actual path of the light intersects with the interface between the two media. With constant indices of refraction, the travel time (1.1) is

$$\begin{aligned}
T(x) &= \frac{1}{c}\left[n_1\int_{x_0}^{x} ds + n_2\int_{x}^{x_1} ds\right] \\
&= \frac{1}{c}\left[n_1\ell_1 + n_2\ell_2\right] \\
T(x) &= \frac{1}{c}\left[n_1\sqrt{(x-x_0)^2+u_0^2} + n_2\sqrt{(x_1-x)^2+u_1^2}\right].
\end{aligned}$$

Accordingly, the travel time $T(x)$ is an algebraic function of x, and we use *differential calculus* to minimize it. We seek the value of x for which $T(x)$ is a minimum corresponding to the value of x where $dT/dx = 0$. Evaluating $dT/dx = 0$ leads to

$$\frac{1}{2}n_1\left[(x-x_0)^2+u_0^2\right]^{-1/2}\left[2(x-x_0)\right] + \frac{1}{2}n_2\left[(x_1-x)^2+u_1^2\right]^{-1/2}\left[-2(x_1-x)\right] = 0.$$

Note that the square root factors are the lengths ℓ_1 and ℓ_2, and given that

$$x - x_0 = \ell_1 \sin\phi_1, \quad x_1 - x = \ell_2 \sin\phi_2$$

from the geometry, we have

$$\frac{n_1}{\ell_1}(\ell_1 \sin\phi_1) = \frac{n_2}{\ell_2}(\ell_2 \sin\phi_2).$$

This leads to *Snell's law* for two homogenous media

$$n_1 \sin\phi_1 = n_2 \sin\phi_2,$$

which indicates that the ratio of the sines of the angles of incidence and refraction is the same as the ratio of the indices of refraction in the two homogeneous media.

Now consider the general case with a variable index of refraction $n(x,u)$ throughout the medium, in which case Fermat's principle (1.1) becomes

$$T[u(x)] = \frac{1}{c}\int_{x_0}^{x_1} n(x,u)ds = \frac{1}{c}\int_{x_0}^{x_1} n(x,u)\sqrt{1+\left(\frac{du}{dx}\right)^2}dx.$$

The last expression follows from the fact that $ds = \sqrt{dx^2 + du^2}$ from the geometry, and because $u = u(x)$, the total differential of $u(x)$ is $du = \frac{du}{dx}dx$. This leads to the result that $ds = \sqrt{1+[u'(x)]^2}dx$. Consequently, in the case of nonhomogeneous media, we seek the path represented by the function $u(x)$ that results in the minimum travel time $T[u(x)]$. Note that evaluation of the definite integral for each possible path $u(x)$ produces a unique scalar value for the travel time $T[u(x)]$. We seek the path that produces the least travel time according to Fermat's principle. In other words, the *functional* $T[u(x)]$, which is a definite integral involving an unknown function $u(x)$, is to be minimized. This requires *calculus of variations*.

Observe that the variational form of the governing equation arises naturally from formulating the problem from first principles. That is, applying the physical law (Fermat's principle) directly leads to the variational form. In summary, we see that finding extrema of an *algebraic function* is a problem in differential calculus, while finding extrema of a *functional* (definite integral) is a problem in variational calculus.

1.3.2 Shape of a Liquid Drop

Second, let us consider the shape that a liquid drop takes when placed on a smooth horizontal surface as shown in Figure 1.3. The shape of the liquid drop will be axisymmetric about the vertical u-axis and be such that the total energy of the drop is a minimum.

The two forms of energy of the liquid drop are that due to its potential energy under a gravitational field and the surface energy due to surface tension at the liquid–gas interface. First, let us consider the potential energy of the drop. The potential energy per unit volume of a horizontal portion of the drop with differential thickness du is $\rho g u(r)$, and the volume of this infinitesimally thin portion is $\pi r^2 du$. Here, ρ is the density of the liquid, g is acceleration due to gravity, and $u(r)$ is the vertical