Calculus of Variations and its Applications

Mason Kamb

May 2017

Contents

1 Introduction

Many problems in mathematics and the physical sciences are naturally formulated in terms of identifying a function that minimizes some quantity of interest. A natural example from geometry is the seemingly simple question: which is the shortest length path between two points in $\mathbb{R}^n$? While everyone knows that such a path is the straight line segment connecting the two points, proving that such is the case is more subtle than such a simple question would suggest. A more complex example of a question in the same vein is to ask: given some open set Ω, and some boundary conditions, can we identify a surface defined on this set, satisfying the boundary conditions, that has the minimum possible area?

Similar questions arise in physics. In optics, Fermat's principle of least time tells us that a light ray always takes a path that leads to a minimum travel time between its starting and endpoints. While in a single, homogenous medium, light will follow a straight line path, it is not immediately clear which path light takes when traversing media, or moving through a medium with a continuously varying speed of light. In mechanics and field theory, the path of a physical state through state space is characterized by an ‘action principle,’ which states that the path that the state takes minimizes (or stationarizes) a functional known as the action of the system.

While these sorts of infinite-dimensional optimization problems are not amenable to the machinery of finite-dimensional real analysis, there exists a generalization of these tools for attacking these questions by analytical methods, known as the ’calculus of variations.’ While the setting of the calculus of variations is over functionals on general normed vector spaces, specifically vector spaces of functions, the methods of results of the calculus of variations are remarkably simple and powerful, and bear a great deal of resemblance to the machinery of finite-dimensional real analysis. In this paper, we provide an overview of these mehtods, and discuss their application to a few specific physical and mathematical problems.

2 Basic Definitions and Lemmas

2.1 Normed Vector Spaces and Functionals

Definition: A normed linear space $\mathcal{R}$ is a set that, in addition to satisfying the usual properties of linear spaces, is equipped with a norm, i.e. a function $f : \mathcal{R} \to \mathbb{R}$ satisfying the following axioms:

1. If $x \in \mathcal{R}$, then $f(x) \geq 0$.
2. $f(x) = 0$ exactly when $x = 0$.
3. For any scalar α, $f(\alpha x) = |\alpha| f(x)$.
4. For any elements $x, y \in \mathcal{R}$, $f(x + y) \leq f(x) + f(y)$.

In general, the action of f on an element $x \in \mathcal{R}$ is denoted by $\|x\|$. From now on, unless otherwise specified, we take $\mathcal{R}$ to be an arbitrary normed vector space.

Some examples of normed linear spaces include:

- $\mathbb{R}^n$, equipped with any norm of the form

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$

  for $p \geq 1$. These norms are known as ‘p-norms’ for real vectors. Taking the limit as $p \to \infty$, we obtain the norm

$$\|x\|_\infty = \max \{|x_j|\}$$

- $L^p(\Omega)$, the space of real-valued functions on a compact set $\Omega \subset \mathbb{R}^n$ whose first p powers are integrable, equipped with the norm

$$\|f\|_p = \left(\int_\Omega |f|^p \right)^{1/p}$$

These norms are known as 'p-norms' for functions. On $L^\infty(\Omega)$, the space of functions whose every power is integrable, we can associate the norm

$$\|f\|_{\sup} = \sup \{y : y = |f(x)|,\, x \in \Omega\}$$

This norm is known as the 'sup norm' for functions. The sup norm is typically included as a p-norm.

- $D^k(\Omega)$, the space of real-valued, n times differentiable functions on $\Omega \subset \mathbb{R}^n$, equipped with the norm

$$\|f\|_{k,p} = \|f\|_p + \|\partial_1 f\|_p + \|\partial_2 f\|_p + \cdots + \|\partial_{11} f\|_p + \cdots$$

i.e., the sum of the p norms of f and its partials up to the kth order. We will denote $\|f\|_{k,\infty}$, i.e. the sum of the sup norms of f and its partials simply as $\|f\|_k$.

An important notion is the relative 'coarseness' and 'fineness' of norms. A norm f is 'finer' than another norm g if there exists some M such that $f(x) < \epsilon \implies g(x) < M\epsilon$ for all $x \in \mathcal{R}$; f is 'coarser' if the converse is true. An important result in finite-dimensional real analysis is that for any two norms f and g on $\mathbb{R}^n$ are equally fine/coarse; i.e., there exists constant α and β such that for all $x \in \mathbb{R}^n$, $\alpha f(x) \leq g(x) \leq \beta f(x)$. Thus in finite-dimensional analysis, the choice of norm is less relevant, and is primarily useful only for ease of calculation.

On the other hand, the choice of norm on general normed linear spaces can be more subtle. For example, $\mathrm{D}^n(\Omega)$ may be equipped with $\|f\|_{\sup}$ or $\|f\|_n$. The choice of norm is in a certain sense arbitrary, but depends on the nature of the application considered. The $\sup$ norm may be easier to work with; however, it is 'coarser' than $\|f\|_n$. Consider, by way of example, the function $f_\epsilon(x) = \epsilon \sin\left(\frac{2\pi x}{\epsilon^2}\right)$, defined on the interval $[-\pi, \pi]$. Note that as $\epsilon \to 0$, $\|f_\epsilon\|_{\sup} \to 0$, but $\|f_\epsilon\|_1$ diverges.

This extra precision can be important in applications, especially in physically-motivated problems, wherein higher derivatives may represent physically meaningful quantities, and it is necessary to have some analytic control over these variables in order to obtain realistic solutions.

We now define basic analytic concepts on $\mathcal{R}$. Most of the machinery discussed in this section bears a great similarity to that of finite-dimensional real analysis.

Definition: A functional J is a function mapping $\mathcal{R}$ to $\mathbb{R}$.

From now on, unless otherwise specified, we take J to be an arbitrary functional on $\mathcal{R}$.

Definition: We say that

$$\lim_{x \to y} J[x] = L$$

if, for all $\epsilon > 0$, for some $\delta > 0$, for all elements $x \in \mathcal{R}$ such that $\|x - y\| < \delta$, $|J[x] - L| < \epsilon$. We say that J is continuous at y if $\lim_{x \to y} J[x] = J[y]$.

Definition: J is said to be bounded if there is some $M \in \mathbb{R}$ such that $|J[x]| \leq M\|x\|$ for all $x \in \mathcal{R}$.

2.2 Extremizing Functionals

Definition: We say that J is Fréchet differentiable at an element $\rho \in \mathcal{R}$ if, for some continuous linear functional δJ,

$$J[\rho + \phi] - J[\rho] = \delta J[\phi] + e[\phi]\|\phi\| \tag{1}$$

where $\phi \in \mathcal{R}$ is arbitrary, and $\lim_{\phi \to \mathbf{0}} e[\phi] = 0$.

δJ may be referred to as the Fréchet derivative of J at ρ, or the 'variation' of J about ρ. Note that if δJ exists, then it is given by

$$\delta J[\phi] = \left[\frac{d}{d\epsilon} J[f + \epsilon\phi]\right]\Bigg|_{\epsilon=0} \tag{2}$$

Theorem: Suppose that a functional J has a local extremum at x, i.e. in some neighborhood $\|x - y\| < \epsilon$, J does not change sign. If J is Fréchet differentiable at x, $DJ[\phi] = 0$.

Proof. Suppose x is a local minimum of J. Let $\phi \in \mathcal{R}$ be arbitrary. Note that for sufficiently small $\epsilon > 0$, the relations

$$J[x + \epsilon\phi] - J[x] \geq 0$$
$$J[x - \epsilon\phi] - J[x] \geq 0$$

both hold. Dividing by ϵ and passing to the limit as $\epsilon \to 0$, we have

$$\delta J_x[\phi] \geq 0$$
$$-\delta J_x[\phi] \geq 0$$

Thus $\delta J_x[\phi] = 0$. The case where J has a local maximum at x follows from a similar argument. □

In many cases, we will not be interested in extremizing a functional on a large vector space $\mathcal{R}$, but rather on a more restricted affine subspace.

For example, we may have a functional J defined on $D^k(\Omega)$, but we may only be interested in functions $f \in S$ on $D^k(\Omega)$ that satisfy $F[f] = r$ for some other suitable smooth functional F and some scalar r. To this end, we can construct a vector space $\hat{D}^k(\Omega)$ of functions $\hat{f}$ that satisfy $F[\hat{f}] = 0$ for $x \in \partial\Omega$, and define the reduced functional $\hat{J}[\hat{f}] = J[\hat{f} + g]$, where $g \in S$. Thus by extremizing $\hat{J}$, we may obtain the extrema of J on S. Furthermore, note that for $\phi \in \hat{D}^k(\Omega)$,

$$\delta\hat{J}_{\hat{f}}[\phi] = \frac{d}{d\epsilon}\left[\hat{J}[\hat{f} + \epsilon\phi]\right] = \frac{d}{d\epsilon}\left[J[g + \hat{f} + \epsilon\phi]\right] = \delta J_{g+\hat{f}}[\phi]$$

Due to this, we may compute the extrema of J on S by identifying those functions $f \in S$ for which $DJ_f[\phi] = 0$, for ϕ satisfying $F[\phi] = 0$. A particularly useful special case of this principle is for extremizing functionals on a space of functions f that satisfy certain boundary conditions, e.g. $f(x) = g(x)$ for $x \in \partial\Omega$. In this case, we can construct a functional $F[f] = \int_{\partial\Omega}(f - g)^2\, d\mu$, which is zero exactly when f satisfies the appropriate boundary conditions. Thus, to extremize a functional on a space of functions satisfying certian boundary conditions, it suffices to identify those functions f such that $DJ_f[\phi] = 0$ for $\phi(\partial\Omega) = 0$.

Now that we have developed a means to identify extrema of functionals on normed vector spaces, we consider the problem of classifying these extrema as minima, maxima, or stationary points. In order to do so, we present the following definitions:

Definition: A functional $G[x, y]$ defined on $\mathcal{R} \times \mathcal{R}$ is said to be bilinear if for each fixed x, $G[x, y]$ is a linear functional in y, and for each fixed y, $G[x, y]$ is a linear functional in x.

Definition: A functional J is said to be quadratic if there is a bilinear functional G such that $J[x] = G[x, x]$.

Definition: A functional J is said to be twice differentiable at $\rho \in \mathcal{R}$ if there exists a continuous quadratic functional $\delta^2 J$ such that

$$J[\rho + \phi] - J[\rho] = \delta J[\phi] + \delta^2 J[\phi] + e[\phi]\|\phi\|^2 \tag{3}$$

where $\phi \in \mathcal{R}$ is arbitrary, and $\lim_{\phi \to \mathbf{0}} e[\phi] = 0$. If such a functional exists, then it is known as the 'second variation' of J about ρ, and can be given by

$$\delta^2 J[\phi] = \left[\frac{d^2}{d\epsilon} J[\rho + \phi\epsilon]\right]\bigg|_{\epsilon=0} \tag{4}$$

Definition: A quadratic functional φ is strongly positive if there exists a $k > 0$ such that

$$\varphi[\phi] \geq k\|\phi\|^2$$

for all $\phi \in \mathcal{R}$. φ is strongly negative if there exists a $k < 0$ such that

$$\varphi[\phi] \leq k\|\phi\|^2$$

Theorem: If J is twice differentiable at ρ, and δJ_ρ^2 is strongly positive, then J has a local minimum at ρ.

Proof. Recall that for all $\phi \in \mathcal{R}$

$$J[\rho + \phi] - J[\rho] = \delta J[\phi] + \delta^2 J[\phi] + e[\phi]\|\phi\|^2$$

Since J is minimal at ρ, the derivative vanishes, and we are left with

$$\begin{aligned} J[\rho + \phi] - J[\rho] &= \delta^2 J[\phi] + e[\phi]\|\phi\|^2 \\ &\geq (k + e[\phi])\|\phi\|^2 \end{aligned}$$

For $\|\phi\|$ sufficiently small, $k + e[\phi]$ is positive, and thus we obtain

$$J[\rho + \phi] - J[\rho] \geq 0$$

in that neighborhood. $\square$

This theorem can be generalized to cases where we are not interested in the entire function space but on a more restricted subspace, in the same way as we generalized our analysis involving single differentiability.

3 Extremizing Integral Functionals

Many of the canonical applications of the calculus of variations consist of extremizing functionals of the form:

$$J[f] = \int_\Omega F(f(x),\, \partial_1 f(x),\, \partial_2 f(x), \cdots,\, \partial_{11} f(x), \cdots,\, x)\, d\mu \tag{5}$$

where $\Omega \subset \mathbb{R}^n$ is a bounded open set, and $f \in D^k(\Omega)$. We may at times want to restrict f to satisfy certain boundary conditions, or other data.

We first prove a general theorem which guarantees that the analytic techniques we have developed are applicable to functions J of this form. Surprisingly, these results are not proved in Gelfand and Fomin's book [1].

Theorem: If F is continuous, J is continuous with respect to the norm $\|f\|_k$. If F is differentiable, then J is differentiable.

Proof. Let $\epsilon > 0$ be arbitrary, and let $\delta_1 > 0$ be appropriately small so that $\|\mathbf{x} - \mathbf{y}\| < \delta_1 \implies |F(\mathbf{x}) - F(\mathbf{y})| < \epsilon$. Suppose $f, g \in D^n(\Omega)$ satisfy $\|f - g\|_n < \delta$. It follows that

$$\begin{aligned}
|J[f] - J[g]| &= \left| \int_\Omega F(f, \partial_1 f, \partial_2 f, \cdots, \partial_{11} f, \cdots, x)\, d\mu - \int_\Omega F(g, \partial_1 g, \partial_2 g, \cdots, \partial_{11} g, \cdots, x)\, d\mu \right| \\
&\leq \int_\Omega |F(f, \partial_1 f, \partial_2 f, \cdots, \partial_{11} f, \cdots, x) - F(g, \partial_1 g, \partial_2 g, \cdots, \partial_{11} g, \cdots, x)|\, d\mu \\
&\leq \int_\Omega \epsilon\, d\mu = \mu(\Omega)\epsilon
\end{aligned}$$

Thus J is continuous. Furthermore, note that for $\phi \in \mathcal{D}^k(\Omega)$,

$$\begin{aligned}
J[f + \phi] - J[f] = \int_\Omega & F(f + \phi, \partial_1 f + \partial_1 \phi, \partial_2 f + \partial_2 \phi, \cdots, \partial_{11} f + \partial_{11} \phi, \cdots, x) \\
&- F(f, \partial_1 f, \partial_2 f, \cdots, \partial_{11} f, \cdots, x)\, d\mu
\end{aligned}$$

By the definition of differentiability for functions on $\mathbb{R}^n$, we can rewrite this as

$$\int_\Omega \nabla F(f, \partial_1 f, \cdots) \cdot (\phi, \partial_1 \phi, \cdots)\, d\mu + \int_\Omega e(\phi, \partial_1 \phi, \cdots) \|(\phi, \partial_1 \phi, \cdots)\|\, d\mu$$

where e is a function such $e(\mathbf{x}) \to 0$ as $\|\mathbf{x}\| \to 0$. Note that the first term is a composition of linear functionals in ϕ, and thus is linear. Furthermore, since integrals and derivatives of order up to k are both continuous with respect to the norm $\|f\|_k$, it follows that this term is a continuous linear functional in ϕ. Furthermore, note that

$$\left| \int_\Omega e(\phi, \partial_1 \phi, \cdots) \|(\phi, \partial_1 \phi, \cdots)\|\, d\mu \right| \leq \mu(\Omega)\, \|e(\phi, \partial_1 \phi, \cdots)\|_0 \|\phi\|_k$$

As $\|\phi\|_k \to 0$, $\|e(\phi, \partial_1 \phi, \cdots)\|_0 \to 0$ by assumption. It follows that J is Fréchet differentiable.

□

A similar argument shows that if F is twice differentiable, then J is twice differentiable. This result can be easily generalized to cases where f is vector-valued rather than scalar-valued.

Thus, we may seek the extrema of functionals of this form by identifying their critical points. Before we may proceed to analyze specific cases of this equation in detail, we prove a helpful lemma.

3.1 Fundamental Lemma of the Calculus of Variations

Lemma: Let $\Omega \subset \mathbb{R}^n$, and suppose $\alpha \in C(\Omega)$. Suppose also that

$$\int_\Omega \alpha \cdot h\, d\mu = 0$$

for every $h \in D^\infty(\Omega)$ satisfying $h(\partial\Omega) = 0$. Then α is identically zero on Ω.

Proof. Suppose α is positive at some point $\hat{x} \in \Omega$. By continuity, there exists some neighborhood $S = \|x - \hat{x}\|_1 < \delta$ where α is positive. Consider the function

$$h(x) = \begin{cases} 0 & x \notin S \\ \prod_{j=1}^n e^{-1/(x_j - \hat{x}_j)^2} & x \in S \end{cases}$$

h is continuous and infinitely differentiable on $\mathbb{R}^n$, and satisfies $h(x) = 0$ for $x \in \partial\Omega$, so it satisfies the hypotheses required above. Furthermore,

$$\int_\Omega \alpha \cdot h\, d\mu = \int_S \alpha \cdot h\, d\mu > 0$$

since αh is entirely positive on S. This is a contradiction, which proves the lemma. □

3.2 Euler-Lagrange Equations

We will now be able to attack the motivating question: given an integral functional J of the form given above, for which f does $\delta J_f = 0$?

For simplicity, instead of the more general form of the functional given above, we consider a functional J of the form:

$$J[f] = \int_a^b F(x, f, f', \cdots, f^{(k)})\, dx$$

Furthermore, we require that f and its $k-1$ first derivatives be specified on the boundary of $[a, b]$. Consider an arbitrary smooth perturbation ϕ that vanishes, along with its $k-1$ first derivatives, on the boundary of $[a, b]$. Note that

$$\begin{aligned}\delta J_f[\phi] &= \frac{d}{d\epsilon} \int_a^b F(x, f + \epsilon\phi, f' + \epsilon\phi', \cdots, f^{(k)} + \epsilon\phi^{(k)})\, dx \bigg|_{\epsilon=0} \\ &= \int_a^b \frac{\partial F}{\partial f}\phi + \frac{\partial F}{\partial f'}\phi' + \cdots + \frac{\partial F}{\partial f^{(k)}}\phi^{(k)}\, dx\end{aligned}$$

For simplicity, we consider the integrals of each of the $k+1$ terms separately. Integrating the jth term by parts, we obtain

$$\int_a^b \frac{\partial F}{\partial f^{(j)}}\phi^{(j-1)}\, dx = \frac{\partial F}{\partial f^{(j)}}\phi^{(j)}\bigg|_a^b - \int_a^b \frac{d}{dx}\frac{\partial F}{\partial f^{(j)}}\phi^{(j-1)}\, dx$$

Since $\phi^{(j-1)} = 0$ on $\{a, b\}$ for $j \leq k$, the first term vanishes. Integrating by parts $j-1$ more times, we obtain

$$\int_a^b \frac{\partial F}{\partial f^{(j)}}\phi^{(j)}\, dx = \int_a^b (-1)^j \frac{d^j}{dx^j}\frac{\partial F}{\partial f^{(j)}}\phi\, dx$$

Recombining these terms, we have

$$\delta J_f[\phi] = \int_a^b \phi\left[\sum_{j=0}^{k}\left((-1)^j \frac{d^j}{dx^j}\frac{\partial F}{\partial f^{(j)}}\right)\right] dx$$

If we then require that this quantity be zero for all ϕ satisfying the appropriate smoothness and boundary conditions, we may apply the fundamental lemma of the calculus of variations to obtain

$$\sum_{j=0}^{k}(-1)^j \frac{d^j}{dx^j}\frac{\partial F}{\partial f^{(j)}} = 0 \tag{6}$$

The f that satisfy this equation are the extrema of J. These equations are known as the Euler-Lagrange equations.

The derivation in the case where f is defined on a subset of $\mathbb{R}^n$ is similar, and the resulting Euler-Lagrange equation is nearly identical in form, except that the terms $\frac{\partial F}{\partial f^{(j)}}$ are replaced with the sum of the partial derivatives of F with respect to all of the jth-order partial derivatives of f. In the case where f is a multivariate function, or where J depends on two functions rather than one, we simply write down the Euler-Lagrange equations for each component.

3.3 Some Examples

The Euler-Lagrange equations are a very useful result in variational analysis, since many naturally occurring problems in mathematics, physics and other domains of application can be formulated in terms of minimizing or maximizing an integral on a given domain. A few examples of these applications follow:

1. **Shortest-Path Problem:** Consider the problem of determining the shortest-distance path between two points **x** and **y** in $\mathbb{R}^n$, i.e. the path $\mathbf{f}(t)$ that minimizes the functional

$$J[\mathbf{f}] = \int_0^1 \|\mathbf{f}'\|\, dt$$

over the space of paths with $\mathbf{f}(0) = \mathbf{x}$ and $\mathbf{f}(1) = \mathbf{y}$. The Euler-Lagrange equations reduce to

$$\frac{\partial\|\mathbf{f}'\|}{\partial f_j} - \frac{d}{dt}\frac{\partial\|\mathbf{f}'\|}{\partial f_j'} = 0$$

for each component f_j of $\mathbf{f}$. Note that the first term in these equations is zero, so we are left with

$$0 = \frac{d}{dt}\frac{\partial\|\mathbf{f}'\|}{\partial f_j'} \implies \frac{\partial\|\mathbf{f}'\|}{\partial f_j'} = C$$

for some constant C. This holds for all f_j' only when each f_j' is constant. It follows that the shortest-distance continuous path between two points has zero second derivative. The only solution given this condition is the straight line path

$$\mathbf{f}(t) = \mathbf{x}(1-t) + \mathbf{y}t \tag{7}$$

2. **Brachistochrone Problem:** Consider the problem of determining the shape of the slope $f(x)$ that will allow a frictionless bead to descend from a point $(a, 0)$ to a point (b, B) below. The functional to be minimized is the travel time. To obtain this functional in terms of a time-independent path function f, we use the conservation of energy to derive the following identity:

$$v(y) = \sqrt{2gy}$$

We are thus led to the following expression for travel time:

$$T[f] = \int_a^b \sqrt{\frac{1+f'^2}{2gf}}\,dx \tag{8}$$

Instead of directly applying the Euler-Lagrange equations, we make use of a related identity, known as the Beltrami identity, which states that

$$L - \frac{\partial L}{\partial f'} f' = C \tag{9}$$

where L is the integrand of the functional being extremized, and C is some constant. This equation is equivalent to the Euler-Lagrange equation in the case where L depends only on f and its first derivative, and not the spatial coordinate x. A proof of this can be found in [1].

From this we obtain

$$\sqrt{\frac{1+f'^2}{2gf}} - f'\frac{f'}{\sqrt{(1+f'^2)(2gf)}} = \frac{1}{\sqrt{2gf(1+f'^2)}} = C \implies f(1+f'^2) = \frac{1}{2C^2}$$

It can be shown that the solution to this equation is the parametric curve

$$x = \frac{4}{C^2}(\theta - \sin(\theta))y = \frac{4}{C^2}(1-\cos(\theta)) \tag{10}$$

These euqations are the equations of a cycloid curve.

3. **Plateau's Problem:** Consider the problem of determining the surface φ over the domain $\Omega \subset \mathbb{R}^n$, satisfying some boundary condition $\varphi(x) = g(x)$ on $\partial\Omega$. The area functional is given by

$$J[\varphi] = \int_\Omega S\,dA \tag{11}$$

where $S = \sqrt{1 + \nabla\varphi \cdot \nabla\varphi}$. The Euler-Lagrange equation is given by

$$\frac{\partial S}{\partial \varphi} - \sum_{j=1}^n \frac{\partial}{\partial x_j}\frac{\partial S}{\partial \varphi_j} = 0$$

Note that S does not depend explicitly on φ, so we are left with

$$\sum_{j=1}^n \frac{\partial}{\partial x_j}\frac{\partial S}{\partial \varphi_j} = 0$$

Expanding this equation, we obtain

$$\sum_{j=1}^n \left[\frac{\varphi_{jj}}{S} - \varphi_j \sum_{k=1}^n \frac{\varphi_k}{S}\varphi_{kj}\right] = 0$$

After some algebraic manipulation, this can be reduced to a simpler form:

$$\nabla \cdot \left(\frac{\nabla \varphi}{S(\varphi)} \right) = 0 \tag{12}$$

While the above derivation is quite simple, we have omitted several complexities relating the existence, uniqueness and smoothness of a minimal area surface to the geometry of $\partial\Omega$ and the nature of the boundary conditions. Some aspects of this problem are currently being researched [2].

4. **Newtonian Mechanics:** In mechanics, we may want to describe the motions of a system of N particles. Newton's laws of motion take the form

$$\frac{d}{dt} m_i x'_{ij} = -\frac{\partial V}{\partial x_{ij}} \tag{13}$$

where x_{ij} is the jth spatial coordinate of the position vector $\mathbf{x}_i$ of the ith particle, and V is the potential energy of the system. We define the Lagrangian of the system as

$$L(\mathbf{x}_1, \cdots, \mathbf{x}'_1, \cdots, t) = \frac{1}{2} \sum_{j=1}^{n} m_j \|\mathbf{x}'_j\|^2 - V(\mathbf{x}_1, \cdots, \mathbf{x}'_1, \cdots, t) \tag{14}$$

which represents the total kinetic energy of the system, minus the potential energy. We define the Action functional of a trajectory $\mathbf{x}(t)$, consisting of the positions of each particle over time as

$$S[\mathbf{x}] = \int_a^b L(\mathbf{x}, \mathbf{x}', t)\, dt \tag{15}$$

Consider the problem of minimizing S over the space of trajectories $\mathbf{x}$ satisfying certain boundary conditions $\mathbf{x}(a) = \mathbf{x}_a$ and $\mathbf{x}(b) = \mathbf{x}_b$. The Euler-Lagrange equations read

$$\frac{\partial L}{\partial x_{ij}} - \frac{d}{dt} \frac{\partial L}{\partial x'_{ij}} = 0$$

for each component x_{ij} Note that by substituting the expression for the Lagrangian given above, we obtain

$$\frac{\partial V}{\partial x_{ij}} - \frac{d}{dt} m_j x'_{ij} = 0$$

which is equivalent to our original formulation of Newton's laws. This correspondence is known as the *principle of least action*.

3.4 The Second Variation

In the above set of examples, we ommitted any discussion of whether the critical functions of each functional were indeed extrema. To this end, we return to the notion of the second variation, developed in section 1.

The general formula for the second variation of an integral functional with a suitably differentiable integrand F, about a function f, is given by

$$\delta^2 F[\phi] = \int_\Omega \frac{d}{d\epsilon} F(f + \epsilon\phi, \partial_1 + \epsilon\partial_1\phi, \cdots, \partial_{11} f + \epsilon\partial_{11}\phi, \cdots, x)\, d\mu \bigg|_{\epsilon=0} \tag{16}$$

This can be tedious to analyze directly for fully general integral functionals, so in this section we restrict our attention to functionals of the form

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

where F is at least twice differentiable. The second variation of J at f is

$$\begin{aligned} \delta^2 J[\phi] &= \int_a^b \frac{d^2}{d\epsilon^2} F(f + \epsilon\phi, f' + \epsilon\phi', x)\, dx \bigg|_{\epsilon=0} \\ &= \int_a^b \partial_{11} F \phi^2 + 2\partial_{12} F \phi\phi' + \partial_{22} F \phi'^2\, dx \end{aligned}$$

If we restrict our attention to those ϕ that satisfy the boundary conditions $\phi(a) = \phi(b) = 0$, we can rewrite this using the following result,

$$\int_a^b 2\partial_{12} F \phi\phi' = -\int_a^b \left(\frac{d}{dx}\partial_{12} F\right) f^2\, dx$$

which follows by integration by parts. We are then left with

$$\partial^2 J[\phi] = \int_a^b P f^2 + Q f'^2\, dx \tag{17}$$

where

$$P(x) = \frac{1}{2}\partial_{11} F(f, f', x) \tag{18}$$

and

$$Q(x) = \frac{1}{2}\left(\partial_{12} F(f, f', x) - \frac{d}{dx}\partial_{12} F(f, f', x)\right) \tag{19}$$

We are then led to ask which conditions we need for $\delta^2 J$ to be strongly positive or negative at an extremal function f. To this end, we consider the extrema of $\delta^2 J$. The associated Euler-Lagrange equation is

$$-\frac{d}{dx}(P\phi') + Q\phi = 0 \tag{20}$$

While $\phi = 0$ is clearly a solution to this equation, other, nontrivial solutions can exist. We define the following term to characterize an important class of these solutions:

Definition: $\hat{a}$ is said to be a conjugate point of a if there exists a solution of (20) that is zero at a and $\hat{a}$, but is not identically zero.

Theorem: Let $f(x)$ be an extremal function of J, i.e. let y satisfy the appropriate Euler-Lagrange equation. Then f is a minimum (resp. maximum) if:

1. $P(x) \geq 0$ (resp. $P(x) \leq 0$) for all $x \in [a, b]$.
2. (a, b) contains no conjugate points of a.

The first condition by itself is a necessary, but not sufficient condition for this result. The full proof of this is somewhat detailed, and we do not include it in this paper. We refer the interested reader to Gelfand and Fomin's text [1] for more details.

We use this approach to verify that the extrema of the functional (15) are indeed minima. Instead of examining this functional in the full multivariate case, we apply it to the case of a single particle moving in one dimension, where the action is given by

$$S[y] = \int_a^b \mathcal{L}(y(t), y'(t), t)\, dt$$

In this case, we have

$$P(t) = \frac{1}{2}\partial_{y'y'}(\frac{1}{2}my'^2 - U(y)) = \frac{m}{2}$$

Assuming the particle has positive mass, then this is positive. Furthermore,

$$Q(t) = \frac{1}{2}\left(\partial_{yy'}(\frac{1}{2}my'^2 - U(y)) - \frac{d}{dx}\partial_{yy'}(\frac{1}{2}my'^2 - U(y))\right)$$

vanishes, since $\frac{1}{2}my'^2$ does not depend explicitly on y, and $U(y)$ does not depend explicitly on y'. The associated Euler-Lagrange equation for $\delta^2 S$ becomes

$$\frac{m}{2}\phi'' = 0$$

This constrains ϕ to be a linear. The only linear function that satisfies $\phi(a) = \phi(b) = 0$ is $\phi = 0$, so there exist no conjugate points of a in (a, b).

It follows that the extrema of S are indeed minima, and thus that we have a principle of *least* action and not just a principle of *critical* action.

Conclusions and Outlook

This paper has given a very cursory overview of the calculus of variations. While the machinery developed in this paper suffices for elementary applications of the sort that we have discussed in the preceding sections, the Calculus of Variations can be extended in many interesting directions to encompass other classes problems, as well as admit other solution techniques. Some of these directions include:

1. *Approximation Methods*. While we have seen that minimizing certain (Integral) functionals leads to a system of PDEs, which we may then proceed to solve via other methods, we can instead attempt to minimize a functional J by considering a sequence of approximate functions f_k such that $J[f_k] \to \inf J$. This can be fruitful both from a theoretical perspective, as well as for physical applications, where we may take an approximate solution f_k with large k as our 'solution' to a variational problem and operate on it. There are a number of analytic questions that arise when considering such sequences, some of which are elementary, but beyond the scope of this paper. More details can be found in [1].

2. *Inverting PDEs*. We have seen the extrema of certain functionals are often characterized by partial differential equations. This can sometimes be exploited in the reverse– functionals can be associated to classes of PDEs by their minima. Applying analytic techniques to function spaces can then be a fruitful tactic for gleaning information about solutions to these PDEs. Understanding when the process of 'inverting' a class of PDEs is possible, and what information can be gleaned via this process is an active area of research [3].

3. *Calculus of Variations on Manifolds*. In this paper, we have been considering functionals defined on functions that map subsets of $\mathbb{R}^n$ to subsets of $\mathbb{R}^m$. However, many interesting problems instead consider functions f that map subsets of a manifold $\mathcal{M}$ to another manifold $\mathcal{N}$. This complicates our analysis significantly, since the set of all such functions (typically) does not form a vector space. Nevertheless, variational techniques can be extended to encompass these cases as well, to great fruition. One example of such a problem arises in a general relativity, where the trajectories of particles are given by geodesics (shortest line paths) on curved 4-manifolds.

These are just a few of the many applications and generalizations of the Calculus of Variations that appear in mathematics, which merit more involved discussion.

References

1. I.M. Gelfand and S.V. Fomin, Calculus of Variations, (1963).

2. A. Fomenko, The Plateau problem. (1990).

3. D. Zenkov, The inverse problem of the calculus of variations : Local and global theory. (2015).