Chapter 16

Variational methods

The calculus of variations originated with attempts to solve Queen Dido's problem[1], known to mathematicians as the *isoperimetric problem*: to determine the shape of that closed plane curve of fixed length that encloses the maximum possible area.

The next major development was the brachistochrone problem: to find the curve of a thin wire joining two points at different elevations, down which a bead slides without friction, in minimum time. According to lore, this problem was first posed by Galileo; the Swiss mathematician Jakob Bernoulli subsequently (in the late 17th Century) offered a reward for its solution. The first to solve it was apparently Isaac Newton, who submitted his solution anonymously. Bernoulli was not fooled, however, remarking "One knows the lion by his claw."

Problems of extrema remain a rich and difficult area of modern mathematics. Plateau's problem—to find the surface of minimum area that spans a general closed curve in space—is unsolved to this day[2].

The possibility of expressing physical principles in a mathematical form, in

[1] According to legend, Dido, in return for a service to a North African chieftain, was given as much land as she could enclose within a bull's hide. She cleverly first cut the hide into thongs, tied them together to make a long string, and enclosed with them a circular plot of ground, upon which she founded the city of Cathage.

[2] An æthestic aspect of Plateau's problem arises from the fact that soap films on bent wire frames tend to minimize the energy of surface tension, thereby minimizing their area.

which the desired result makes some quantity that depends on it an extremum, has turned out to be of enormous importance. It is the basis of Hamilton's Principle in mechanics, has led subsequently to the path integral formulation of quantum mechanics, and is also the basis of quantum field theory, not to mention string theory. The practical applications of variational methods are legion, in fields ranging from structural engineering to economics and finance. This chapter will touch on the basic ideas and some representative applications.

16.1 Classical variational calculus

In differential calculus we often make use of the property that the derivative of a function (of one or several variables) vanishes at a (local) minimum or maximum to locate such extremal points. As we have seen in Chapter 9, we can think of functions—especially continuous ones—as vectors in an abstract space. If this space is a separable Hilbert space such as $\mathcal{L}_2$ we can think of the function as a denumerable set of independent variables[3]. In that case a *functional* $\mathcal{F}(\{f\})$ that depends on the function $f(x)$ is the generalization, from an ordinary function of several variables, to a function of *infinitely* many variables.

16.1.1 Functionals

Here are some specific examples of functionals:

1. Suppose a curve in the plane is represented by the vector $\vec{r} = (x, y = f(x))$; then the length of the curve is

$$\ell = \int_a^b dx \sqrt{1 + (f'(x))^2}$$

and the area enclosed between it and the x-axis is

$$A = \int_a^b f(x)\, dx\,.$$

[3]...for example the coefficients in an expansion of $f(x)$ in a complete set of orthonormal functions $\phi_k(x)$.

2. If an arbitrary curve $f(x)$ represents a stationary uniform chain of mass μ per unit length, then its (potential) energy is

$$V = \mu g \int_a^b dx\, f(x) \sqrt{1 + (f'(x))^2}\,.$$

3. If the curve $f(x)$ represents a stiff wire in the vertical plane (x horizontal, y vertical), along which a bead can slide without friction under the influence of gravity, then the time the bead takes to slide down the wire is

$$T = \int_a^b \frac{ds(x)}{v(x)} \equiv \int_a^b dx \sqrt{\frac{1 + (f'(x))^2}{2gf(x)}}$$

4. If

$$\vec{r}(\tau) = \begin{pmatrix} x(\tau) \\ y(\tau) \\ z(\tau) \end{pmatrix}$$

represents the parametric curve of a light ray in a medium of variable refractive index $n(\vec{r})$, then

$$T = \frac{1}{c} \int_{\tau_1}^{\tau_2} n(\vec{r}(\tau)) \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}\, d\tau$$

is the time for the light to go from $\vec{r}(\tau_1)$ to $\vec{r}(\tau_2)$, where $\dot{x} = dx/d\tau$, *etc.*

5. The energy of a thin beam of uniform density μ per unit length, with a central load W at the midway point is

$$E(\{\xi\}) = \int_0^L \left[\frac{YI}{2} \left(\frac{d^2\xi}{dz^2} \right)^2 + \mu g \xi(z) + W \delta(z - L/2)\, \xi(z) \right] dz\,,$$

where Y is Young's modulus and I the "areal moment of inertia".

16.1.2 The Euler-Lagrange equation

The central problem of the calculus of variations is to find extrema of functionals. The basic idea is this: replace the unknown function f in the functional $\mathcal{F}\{f\}$ by $f+\eta$, where $\eta(x)$ is a "small" function. The variation of $\mathcal{F}$ under this change is

$$\delta\mathcal{F} = \mathcal{F}\left(\{f+\eta\}\right) - \mathcal{F}\left(\{f\}\right) ;$$

if f is the function that makes the functional extremal, the variation must vanish, to first order in η. To make this more explicit, consider a functional that depends only on $f(x)$, $f'(x)$ and explicitly on x:

$$\mathcal{F}\left(\{f\}\right) = \int_a^b \Phi\left(f(x), f'(x), x\right)\, dx\,. \tag{16.1}$$

Clearly,

$$\begin{aligned}
&\delta\mathcal{F}\left(\{f\}\right) = \\
&\quad = \int_a^b \left[\Phi\left(f(x)+\eta(x), f'(x)+\eta'(x), x\right) - \Phi\left(f(x), f'(x), x\right)\right] dx \\
&\quad = \int_a^b \left[\frac{\partial\Phi}{\partial f}\eta + \frac{\partial\Phi}{\partial f'}\eta' + \ldots\right] dx
\end{aligned}$$

We then set the terms linear in η to zero. The difficulty we now face is that while we imagine we know the "small" deviation η, we know nothing about its derivative. So the idea is to re-express the term containing $(\eta)'$ so we get rid of the derivative. This can be accomplished using integration by parts:

$$\begin{aligned}
\int_a^b \frac{\partial\Phi}{\partial f'}\eta'(x)\, dx &= \int_a^b \frac{d}{dx}\left(\frac{\partial\Phi}{\partial f'}\eta\right) dx - \int_a^b \eta\frac{d}{dx}\left(\frac{\partial\Phi}{\partial f'}\right) dx \\
&= \left(\frac{\partial\Phi}{\partial f'}\eta\right)\Big|_a^b - \int_a^b \eta\frac{d}{dx}\left(\frac{\partial\Phi}{\partial f'}\right) dx\,.
\end{aligned}$$

We now note that in most cases we want to specify the values $f(a)$ and $f(b)$. For example, in the brachistochrone problem the end points (x, y) are fixed. The function $\eta(x)$ must therefore vanish at the ends of the interval, so the perfect derivative term vanishes:

$$\left(\frac{\partial\Phi}{\partial f'}\eta\right)\Bigg|_a^b = 0$$

and we can write, to first order in η,

$$\delta\mathcal{F} = \int_a^b \left[\frac{\partial\Phi}{\partial f} - \frac{d}{dx}\left(\frac{\partial\Phi}{\partial f'}\right)\right]\eta\,(x)\;dx = 0\,. \tag{16.2}$$

Now the variation $\eta(x)$ can be *any* function that vanishes at $x = a$ and $x = b$. Hence we may take it to be proportional to successive members of a complete orthonormal system. The extremal condition then amounts to the statement that the contents of the square bracket in Eq. 16.2 is orthogonal to every member of a complete set. This means that the bracketed function itself must vanish for all $x \in [a, b]$, except possibly on a set of Lebesgue measure zero. That is, we have found

$$\frac{\partial\Phi}{\partial f} - \frac{d}{dx}\left(\frac{\partial\Phi}{\partial f'}\right) = 0\,. \tag{16.3}$$

Equation 16.3 is known as the Euler-Lagrange equation.

.....

Exercise 16.1
If we approximate the function $f(x)$ by the first N terms of its expansion in a complete orthonormal set $\{\phi_k\}$,

$$f\,(x) = \sum_{k=1}^{N} f_{_k}\phi_k\,(x)\,,$$

we may think of the functional $\mathcal{F}$ as a function of a finite number of variables, f_k. With $\mathcal{F}$ defined as in Eq. 16.1, show that the condition for $\mathcal{F}$ to have

an extremum:

$$\frac{\partial F}{\partial f_k} = 0, \quad k = 1, \ldots, N$$

can be put in the form of the Euler-Lagrange equation, as long as

$$\eta_N(x) = \sum_{k=1}^{N} \delta f_k \phi_k(x)$$

vanishes at the endpoints of the interval.

.....

16.1.3 Applications

Let us begin with the brachistochrone problem. A bead slides frictionlessly on a wire that is bent into a plane curve, in the $x-z$ plane, and falls from an initial point (a, z_a) to a lower point (b, z_b). The time of the fall is therefore

$$T = \int_a^b \frac{ds(x)}{v(x)} \equiv \int_a^b dx \sqrt{\frac{1 + (z'(x))^2}{2g\, z(x)}}\,.$$

The function Φ is

$$\Phi(z, z') = \sqrt{\frac{1 + (z'(x))^2}{z(x)}}$$

(the factor $2g$ is immaterial) and the Euler-Lagrange equation becomes

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

This leads to an impressively ugly second-order nonlinear differential equation to solve, so it is best to employ a trick. We turn the problem on its side, so to speak, and take z for the variable of integration. Then letting $\dot{x} = dx/dz$ we have

$$\delta \int_{z_a}^{z_b} dz \sqrt{\frac{1+\dot{x}^2}{z}} = 0$$

or (since there is no explicit dependence on x)

$$\frac{d}{dz}\left(\frac{\dot{x}}{1+\dot{x}^2} \sqrt{\frac{1+\dot{x}^2}{z}} \right) = 0\,.$$

This can be integrated immediately, and squared, to give the first-order equation

$$\frac{1}{z} \cdot \frac{\dot{x}^2}{1+\dot{x}^2} = A^2$$

which leads to the separable form

$$dx = \pm dz \sqrt{\frac{A^2 z}{1 - A^2 z}}\,.$$

We want the $-$ sign because on physical grounds the altitude $z(x)$ must decrease monotonically with x. The solution is the equation of a cycloid,

$$A^2 x + B = \sqrt{A^2 z\,(1 - A^2 z)} - \sin^{-1}\sqrt{A^2 z}\,,$$

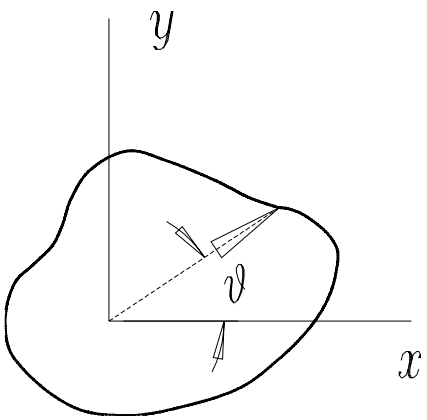
where the constants A^2 and B must be adjusted to match the initial and final conditions.

As a second application, let us solve Dido's problem. We express the curve parametrically (in Cartesian coordinates) in terms of an angle θ:

$$A \;=\; \frac{1}{2}\int_0^{2\pi} d\theta \left[x^2(\theta) + y^2(\theta) \right]$$

$$\ell = \int_0^{2\pi} d\theta \sqrt{\dot{x}^2(\theta) + \dot{y}^2(\theta)}$$

as shown below



where

$$\dot{x} = \tfrac{dx}{d\theta}, \quad \dot{y} = \tfrac{dy}{d\theta}\,.$$

We want to maximize the area A while keeping the perimeter ℓ fixed[4]; this is an example of a *constrained* variational problem. We employ the same device used when finding the extremum of a function of a finite number of variables, subject to constraints—namely a Lagrange multiplier for each constraint. We therefore want to set

$$\delta(\lambda A + \ell) = 0$$

which leads to the Euler-Lagrange equations (there are two equations, one for each function $x(\theta)$ and $y(\theta)$)

$$\lambda x + \frac{\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}(\dot{x}\ddot{y} - \dot{y}\ddot{x}) = 0$$

[4] ...or minimize ℓ while keeping A constant.

$$\lambda y - \frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}} (\dot{x}\ddot{y} - \dot{y}\ddot{x}) = 0$$

These can be combined to get

$$\lambda (x\dot{x} + y\dot{y}) = 0 \,,$$

whose solution is $x^2 + y^2 = \text{constant}$.

Alternatively, we could have expressed the isoperimetric problem in polar coordinates,

$$A = \tfrac{1}{2} \int_0^{2\pi} d\theta \, r^2 (\theta)$$

$$\ell = \int_0^{2\pi} d\theta \, \sqrt{r^2 (\theta) + \dot{r}^2 (\theta)}$$

where $\dot{r} = dr/d\theta$, thus obtaining the Euler-Lagrange equation

$$r - \lambda \frac{r}{\sqrt{r^2 + \dot{r}^2}} + \lambda \frac{d}{d\theta} \left(\frac{\dot{r}}{\sqrt{r^2 + \dot{r}^2}} \right) = 0 \,.$$

We see by inspection[5] that

$$r (\theta) = \lambda = \text{constant}$$

—that is, a circle—is a solution.

16.1.4 Variable end points

The curve of a loaded beam depends on the conditions imposed at the ends. If the beam is a cantilever (that is, unsupported at one end) we cannot specify *ab initio* the displacement of the (unsupported) free end at $z = L$.

[5] Solving directly is messy and takes us somewhat afield.

All we can say is that $\xi(0) = \xi'(0) = 0$ (because obviously, the supported end must be clamped or it cannot remain in static equilibrium). If we vary the energy (here we consider the case with a weight hung on the free end of the beam) to find its minimum, we obtain

$$\delta \int_0^L \left[\frac{YI}{2} \left(\xi''\right)^2 + \mu g \xi + W \delta \left(z - L + \varepsilon\right) \xi \right] = 0$$

or, with $f(z) = \mu g \xi + W \delta \left(z - L + \varepsilon\right)$,

$$\int_0^L \left[YI \frac{d^4\xi}{dz^4} + f(z) \right] \delta\xi + \left(YI \frac{d^2\xi}{dz^2} \delta\xi' \right) \Bigg|_0^L - \left(YI \frac{d^3\xi}{dz^3} \delta\xi \right) \Bigg|_0^L = 0 \,.$$

(Note that we get around any problems with the delta function by hanging the weight a distance ε in from the end. We can let $\varepsilon \to 0$ after we are done.) Because the function $\delta\xi(z)$ is arbitrary, the Euler-Lagrange equation

$$YI \frac{d^4\xi}{dz^4} + f(z) = 0$$

is still satisfied. However, we now also have the end-point contributions

$$\left(YI \frac{d^2\xi}{dz^2} \delta\xi' \right) \Bigg|_0^L - \left(YI \frac{d^3\xi}{dz^3} \delta\xi \right) \Bigg|_0^L$$

that must vanish as well. The conditions at the clamped end require

$$\delta\xi(0) = \delta\xi'(0) = 0$$

so we deduce the subsidiary conditions

$$\left. \frac{d^2\xi}{dz^2} \right|_{z=L} = \left. \frac{d^3\xi}{dz^3} \right|_{z=L} = 0 \,.$$

To solve this problem we simply integrate:

$$\left(\frac{d^3\xi}{dz^3}\right)\Bigg|_0^z = -\frac{1}{YI}\int_0^z d\zeta\, f(\zeta) = -\frac{\mu g}{YI}z - \frac{W}{YI}\theta(z-L+\varepsilon)$$

or

$$\frac{d^3\xi}{dz^3} = A - \frac{\mu g}{YI}z - \frac{W}{YI}\theta(z-L+\varepsilon)\ .$$

Integrating again,

$$\frac{d^2\xi}{dz^2} = B + Az - \frac{\mu g}{2YI}z^2 - \frac{W}{YI}(z-L+\varepsilon)\ \theta(z-L+\varepsilon)$$

and again,

$$\frac{d\xi}{dz} = \left(\left.\frac{d\xi}{dz}\right|_{z=0} \equiv 0\right) + Bz + \frac{A}{2}z^2 - \frac{\mu g}{6YI}z^3 - \frac{W}{2YI}(z-L+\varepsilon)^2\ \theta(z-L+\varepsilon)$$

and a final time yields

$$\xi = (\xi(0)\equiv 0) + \frac{B}{2}z^2 + \frac{A}{6}z^3 - \frac{\mu g}{24YI}z^4 - \frac{W}{6YI}(z-L+\varepsilon)^3\ \theta(z-L+\varepsilon)\ .$$

The two non-vanishing constants of integration, A, B, are determined from the end-point conditions at $z = L$:

$$\xi''(L) = 0 \quad \Rightarrow\ B + AL - \tfrac{\mu g}{2YI}L^2 = 0$$

$$\xi'''(L) = 0 \quad \Rightarrow\ A - \tfrac{\mu g}{YI}L - \tfrac{W}{YI} = 0$$

leading to

$$\xi(z) = -\frac{1}{2}\left(\frac{\mu g L^2}{2YI} + \frac{WL}{YI}\right)z^2 + \frac{1}{6}\left(\frac{\mu g L}{YI} + \frac{W}{YI}\right)z^3 - \frac{\mu g}{YI}z^4\,.$$

.....

Exercise 16.2
Find the curve of a uniform chain of mass μ per unit length, hanging from two level supports a distance L apart, by minimizing its potential energy

$$V = \mu g \int_a^b dx\, y(x)\sqrt{1 + (y'(x))^2}$$

subject to a suitable constraint.

.....

Exercise 16.3
Find the curve of a uniform thin beam of mass μ per unit length, loaded with a point weight W halfway between the level end-supports, for the two cases:

a) the ends are clamped to be horizontal (that is, both the displacement ξ, and its first derivative, $d\xi/dz$ vanish at both ends);

b) the ends simply rest on the supports (that is, the displacement ξ vanishes at the ends, but its derivative is unconstrained there).

.....

Exercise 16.4
In §16.1.1, the fourth example is Fermat's Principle of Least Time. Consider two cases:

a) Light travels from one medium of uniform refractive index n_1 across a plane interface to another medium of (uniform) index n_2. Show that the ray obeys Snell's Law.

b) The index has the form $n(x, y, z) = 1+\alpha z$. What is the equation of the curve traversed by a light ray, taking x as the parameter representing the curve—that is, what are $y(x)$ and $z(x)$? (This is the explanation of the mirage.)

.....

16.2 Hamilton's principle

D'Alembert's principle of virtual work [**?**] allows us to derive the equations of motion for mechanical systems, subject to *holonomic* constraints, directly from Newton's Second Law of motion,

$$\frac{d\vec{p}}{dt} = \vec{F}\,.$$

For each generalized coordinate $q(t)$ we find an equation of the form

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0\,. \tag{16.4}$$

We immediately recognize this as the Euler-Lagrange equation, where L is the Lagrangian of the system, expressed in terms of kinetic energy T and potential energy V

$$L = T - V\,.$$

Sir William Rowan Hamilton noticed that Eq. 16.4 could be expressed as a variational problem,

$$\delta \int_{t_1}^{t_2} dt\, L\,(q, \dot{q}; t) = 0\,, \tag{16.5}$$

now called variously Hamilton's Principle, or the Principle of Least Action[6], since Eq. 16.4 is just the corresponding Euler-Lagrange equation for the function $q(t)$ that makes the functional in Eq. 16.5 an extremum.

. . . . . .

Exercise 16.5
The Lagrangian of a pendulum, consisting of a bob of mass m and negligible physical extension on the end of a massless rigid rod of length ℓ, suspended

[6]The latter name derives from the fact that the time-integral of the Lagrangian in Eq. 16.5 is called the *action*.

below a fixed pivot is

$$L = \frac{m}{2}\left(\dot{x}^2 + \dot{y}^2\right) - mgy$$

where x and y are the position coordinates of the bob, and $\dot{x}$ and $\dot{y}$ respectively the components of its velocity.

a) Express the Lagrangian in terms of a single generalized coordinate, the angle θ the rod makes with the vertical.

b) Derive the equation of motion for $\theta(t)$.

c) Reduce this equation to quadrature (that is, express the answer as an integral).

.....

16.3 The role of symmetry

Principle of Least Action is important in theoretical physics because it provides a way to incorporate the fundamental symmetries of a system *ab initio*, thereby restricting the possible forms that the Lagrangian can take. Consider, for example, a non-interacting particle in three dimensions. Translation invariance in space and time implies the Lagrangian cannot depend either on position $\vec{x}$ or on the time, t. Hence it can depend only on the velocity,

$$\vec{v} = \frac{d\vec{x}}{dt}\,.$$

Rotation invariance implies the Lagrangian must be a scalar, *i.e.* a function of $\vec{v}\cdot\vec{v}$. The simplest such function is

$$L = \frac{1}{2}m\vec{v}\cdot\vec{v}\,,$$

which we see is the usual non-relativistic kinetic energy.

We could have determined the form of the Lagrangian, up to the factor $m/2$, by imposing the symmetry of Galilean invariance. According to Galileo, if the velocity of a particle in frame S is $\vec{v}$, then in frame S' moving with velocity $\vec{V}$ relative to the first, the velocity is

$$\vec{v}' = \vec{v} - \vec{V}\,,$$

and the time differential is

$$dt' = dt\,.$$

Since the Lagrangian is a general function of $\vec{v}^2$, we have in system S',

$$A' = \int_{t_1}^{t_2} dt\, L\left(\left(\vec{v} - \vec{V}\right)^2\right)$$

We want A' to be reducible to A plus, at most, a constant. Suppose $\vec{V}$ is small—then we can expand in Taylor's series to find

$$A' \approx \int_{t_1}^{t_2} dt\, \left[L\left(\vec{v}^2\right) - 2\vec{v}\cdot\vec{V}\frac{\partial L}{\partial\left(\vec{v}^2\right)} + \ldots\right].$$

Now, in order for the term

$$\Delta A' = -2\vec{V}\cdot\int_{t_1}^{t_2} dt\, \left(\vec{v}\frac{\partial L}{\partial\left(\vec{v}^2\right)}\right)$$

to be a constant—that is, independent of the path taken by the particle[7]—the integrand must be a perfect differential with respect to time. Since

[7] This condition must be satisfied in order that varying the term $\Delta A'$ not affect the dynamics.

$\vec{v}$ is already a perfect differential, namely $d\vec{x}/dt$, the only way the entire integrand can be a perfect differential is if

$$\frac{\partial L}{\partial\left(\vec{v}^2\right)} = \text{constant}\,,$$

that is, if $L \propto \vec{v}^2$. In plain terms, the only possible Galilean-invariant Lagrangian that describes a non-interacting particle is

$$L = \frac{1}{2} m \vec{v}^2\,.$$

Suppose we wanted to impose Lorentz invariance, rather than Galilean invariance, as the fundamental symmetry. As Einstein showed, the correct formula for addition of velocities is, for small $\vec{V}$ (but large $\vec{v}$),

$$\vec{v}' \approx \frac{\vec{v} - \vec{V}}{1 - \vec{v}\cdot\vec{V}\Big/c^2}\,.$$

We must also correct the time: in the moving frame we have, under the same conditions,

$$dt' \approx dt\left(1 - \vec{v}\cdot\vec{V}\Big/c^2\right)\,.$$

Hence

$$A' \approx A - \vec{V}\cdot\int_{t_1}^{t_2} dt\,\vec{v}\left(\frac{L}{c^2} + 2\left(1 - \vec{v}^2\Big/c^2\right)\frac{\partial L}{\partial\left(\vec{v}^2\right)}\right);$$

the condition that the term linear in $\vec{V}$ be independent of path yields the relation

$$\frac{L}{c^2} + 2\left(1 - \vec{v}^2\Big/c^2\right)\frac{\partial L}{\partial\left(\vec{v}^2\right)} = \alpha = \text{constant}\,,$$

whose solution is

$$L = \alpha c^2 + \beta\sqrt{1 - \vec{v}^2\big/c^2}\,.$$

To make contact with the nonrelativistic result for small $\vec{v}$ we choose $\alpha = m$, $\beta = -mc^2$; thus the unique (up to an additive constant that contributes nothing to the dynamics) Lagrangian is

$$L = mc^2 - mc^2\sqrt{1 - \vec{v}^2\big/c^2}\,.$$

It is perhaps worth pointing out that

$$dt\sqrt{1 - \vec{v}^2\big/c^2} \equiv d\tau\,,$$

where the differential $d\tau$ of the "proper time" is a manifestly Lorentz-invariant quantity, so the integral $\int d\tau$ is also manifestly Lorentz-invariant.

16.4 Variational problems in continuous systems

The equation of motion of a vibrating stretched string is

$$T\frac{\partial^2\psi}{\partial x^2} - \mu\frac{\partial^2\psi}{\partial t^2} = 0 \tag{16.6}$$

where x is the coordinate along the string, $\psi(x,t)$ is the instantaneous displacement of the piece of string between x and $x + dx$ in the direction perpendicular to the string, T is the string tension, and $\mu(x)$ the mass per unit length. If we look at the string as composed of masses $\mu(x)\Delta x$ connected by massless segments of tension T, and if we add up the kinetic and potential energies of the masses, we get a Lagrangian

$$L \quad = \quad \frac{1}{2}\sum_{x_k = k\Delta x}\left[\mu\left(x_k\right)\Delta x\left(\dot{\psi}\left(x_k\right)\right)^2 - \frac{T}{\Delta x}\left(\psi\left(x_{k+1}\right) - \psi\left(x_k\right)\right)^2\right]$$

$$\underset{\Delta x \to 0}{\to} \quad \frac{1}{2}\int_0^\ell dx \left[\mu\left(x\right)\left(\frac{\partial\psi}{\partial t}\right)^2 - T\left(\frac{\partial\psi}{\partial x}\right)^2\right].$$

The action is therefore

$$A = \frac{1}{2}\int dt \int dx \left[\mu\left(x\right)\left(\frac{\partial\psi}{\partial t}\right)^2 - T\left(\frac{\partial\psi}{\partial x}\right)^2\right]$$

and we can apply Hamilton's principle as before. Imagine the string is clamped at the ends; then $\psi(0,t) = \psi(\ell,t) = 0$, hence the small variation $\eta(x,t)$ also vanishes at the supports. The result is that (assuming η also vanishes at the beginning and end of the time interval in the action)

$$\delta A = \int dt \int dx \left[T\frac{\partial^2\psi}{\partial x^2} - \mu\frac{\partial^2\psi}{\partial t^2}\right]\eta\left(x,t\right) = 0$$

which, because of the generality of $\eta(x,t)$, gives us the wave equation 16.6.

The energy (Hamiltonian) of the system becomes

$$H = \frac{1}{2}\int_0^\ell dx \left[\mu\left(x\right)\left(\frac{\partial\psi}{\partial t}\right)^2 + T\left(\frac{\partial\psi}{\partial x}\right)^2\right] \tag{16.7}$$

where we have made the contact transformation[8] that is the continuum equivalent of

$$H = \dot{q}\frac{\partial L}{\partial\dot{q}} - L\,.$$

Note that the Hamiltonian, Eq. 16.7, is manifestly real and positive.

Suppose we consider sinusoidal vibrations,

$$\psi\left(x,t\right) = \cos\left(\omega t\right)\varphi\left(x\right)\,.$$

[8] see, *e.g.* , Goldstein, *Classical Mechanics.*

The action then becomes

$$A = \frac{1}{2}\int_{t_1}^{t_2} dt \int_0^\ell dx \left[\omega^2 \mu\left(x\right)\varphi^2 \sin^2(\omega t) - T\left(\frac{d\varphi}{dx}\right)^2 \cos^2(\omega t)\right],$$

which, for time intervals $\gg 2\pi\omega^{-1}$, has the limit

$$\underset{t_2 \gg t_1}{\rightarrow} \frac{1}{4}\int_0^\ell dx \left[\omega^2 \mu\left(x\right)\varphi^2 - T\left(\frac{d\varphi}{dx}\right)^2\right].$$

The Principle of Least Action is then equivalent to the constrained extremum problem

$$\delta \int_0^\ell dx \left[T\left(\frac{d\varphi}{dx}\right)^2 - \lambda\mu\left(x\right)\left(\varphi\left(x\right)\right)^2\right] = 0$$

—that is, extremizing the string tension energy while keeping the (weighted) normalization integral

$$N = \int_0^\ell dx\, \mu\left(x\right)\varphi^2(x)$$

fixed. The Lagrange multiplier λ is just ω^2 in the resulting eigenvalue equation

$$-\frac{d^2\varphi}{dx^2} = \omega^2 \frac{\mu\left(x\right)}{T}\varphi\left(x\right). \tag{16.8}$$

If the linear mass-density is uniform , $\mu =$ constant, Eq. 16.8 is the equivalent of the time-independent Schrödinger equation for a particle confined in a one-dimensional box. This suggests that we might represent all the linear partial differential equations of mathematical physics as appropriate action principles, and that seems to be so.

16.5 Bilinear functionals

We saw in Chapter 11 that the matrix elements $\langle\varphi|\, H\, |\psi\rangle$ of linear operators in Hilbert space could be regarded as bilinear functionals of the functions φ and ψ. The eigenvalue equation

$$H\psi = \lambda\psi$$

may therefore be obtained by varying with respect to $\varphi^\dagger$ subject to the constraint $\langle\varphi \mid \psi\rangle = \text{constant}$. The relation between the variational calculus and linear eigenvalue problems leads to some useful approximation techniques.

16.5.1 Rayleigh-Ritz upper bound

For example, suppose the operator H is hermitian[9] on some Hilbert space $\mathcal{H}$: then its eigenfunctions form a complete orthogonal system $\{\varphi_k\}_0^\infty$ with corresponding (real) eigenvalues h_k. We may then express any diagonal matrix element of H as

$$\langle\psi|\, H\, |\psi\rangle = \sum_{k=0}^{\infty} |c_k|^2\, h_k\,,$$

where

$$\psi\,(x) = \sum_{k=0}^{\infty} \langle\varphi_k \mid \psi\rangle\, \varphi_k\,(x) = \sum_{k=0}^{\infty} c_k \varphi_k\,(x)\,.$$

Now consider the ratio

$$\frac{\langle\psi|\, H\, |\psi\rangle}{\langle\psi \mid \psi\rangle} \equiv \frac{\sum\limits_{k=0}^{\infty} |c_k|^2\, h_k}{\sum\limits_{k=0}^{\infty} |c_k|^2}\ ; \tag{16.9}$$

[9] ... or better yet, essentially self-adjoint.

since $|c_k|^2 > 0$, the right hand side of Eq. 16.9 can be minimized by replacing all the h_k by the smallest eigenvalue. That is, for any ψ

$$h_{\min} \leq \frac{\sum_{k=0}^{\infty} |c_k|^2 h_k}{\sum_{k=0}^{\infty} |c_k|^2} \equiv \frac{\langle\psi| H |\psi\rangle}{\langle\psi \mid \psi\rangle} \stackrel{df}{=} \bar{h} .$$

This fact leads to the Rayleigh-Ritz variational method for estimating the lowest eigenvalue of the operator H: we choose a function ψ that contains undetermined parameters, compute the left side of Eq. 16.9, obtaining a function of those parameters, then minimize that function, obtaining a finite set of algebraic equations to solve. By this means the ground state energy (and hence the electron removal energy) of the helium atom was calculated to great precision, historically providing enormous confidence in the new quantum mechanics.

. . . . . .

Exercise 16.6
Find an upper bound to the ground state energy of a one-dimensional quantum-mechanical harmonic oscillator whose Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\Omega^2 x^2 ,$$

using a trial function that is the ground state of a particle in a 1-dimensional box of width L. (That is, take L for your variational parameter.) Make sure the wave function is symmetric about $x = 0$.

. . . . . .

Exercise 16.7
Compute an upper bound to the ground state energy of the hydrogenic atom in three dimensions, with Hamiltonian

$$H = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{r} ,$$

using a trial wave function of the form

$$\psi(\vec{r}) = e^{-\lambda r^2} .$$

.....

16.5.2 Lower bounds

We may evaluate the ratio[10]

$$K(\lambda, \{\psi\}) = \frac{\langle (H - \bar{h})\,\psi \mid (H - \lambda)\,\psi \rangle}{\langle \psi \mid \psi \rangle} \equiv \frac{\langle H\psi \mid H\psi \rangle}{\langle \psi \mid \psi \rangle} - \bar{h}^2 \stackrel{df}{=} \Delta - \bar{h}^2$$

for any λ and any ψ. We then may expand ψ in eigenfunctions of H to obtain

$$K(\lambda, \{\psi\}) = \Delta - \bar{h}^2 = \frac{\sum_{k=0}^{\infty} |c_k|^2 (h_n - \bar{h})(h_n - \lambda)}{\sum_{k=0}^{\infty} |c_k|^2} .$$

There are two ways we might use this relation:

1. let $\lambda = \bar{h}$ and suppose we know that

   $$(h_{\min} - \bar{h})^2 \leq (h_n - \bar{h})^2 \ \forall\, n\,;$$

   clearly, then,

   $$\Delta - \bar{h}^2 = \frac{\sum_{k=0}^{\infty} |c_k|^2 (h_n - \bar{h})^2}{\sum_{k=0}^{\infty} |c_k|^2} \geq (h_{\min} - \bar{h})^2 \geq 0\,,$$

[10] The methods in this subsection come from Goertzel and Tralli, *Some Mathematical Methods of Physics.*

or

$$\bar{h} - \sqrt{\Delta - \bar{h}^2} \leq h_{\min} \leq \bar{h} + \sqrt{\Delta - \bar{h}^2}\,.$$

(For this bound to have the correct ordering and signs, either H must be positive, or else if $h_{\min} < 0$, the Rayleigh-Ritz upper bound $\bar{h}$ must also be negative.)

Example:

Let us consider the vibrating string of unit length. Then we can obtain an upper bound on the lowest eigenvalue, $h_{\min}$, (which we know is π^2) of $-D^2 \equiv -d^2/dx^2$ using a differentiable function $\psi(x)$ which vanishes at $x = 0$ and at $x = 1$. Such a function is $x(1-x)$. Then

$$\bar{h} = \frac{\int_0^1 dx \left[\frac{d}{dx}\left(x\left(1-x\right)\right)\right]^2}{\int_0^1 dx \left[x\left(1-x\right)\right]^2} = 10 > \pi^2\,.$$

We evaluate the lower bound next:

$$\Delta = \frac{\int_0^1 dx \left[\frac{d^2}{dx^2}\left(x\left(1-x\right)\right)\right]^2}{\int_0^1 dx \left[x\left(1-x\right)\right]^2} = 120$$

or, combining the two bounds,

$$10 - \sqrt{120 - 100} \leq h_{\min} \leq \bar{h} = 10\,.$$

Admittedly $10 - \sqrt{20} = 5.527\ldots$ is not a very close lower bound to π^2, but it is definitely a lower bound. With more information we can do better, as we see below.

2. Suppose we have approximate knowledge of the second eigenvalue, h_1 (we suppose the eigenvalues are ordered, $h_{\min} \equiv h_0 \leq h_1 \leq h_2 \ldots$). Then if we know that $h_0 \leq \bar{h} \leq h_1$, choose

$$\lambda = h_0 + h_1 - \bar{h}\,,$$

giving

$$\Delta - \bar{h}^2 = \frac{\sum_{k=0}^{\infty} |c_k|^2 \left(h_n - \bar{h}\right)\left(h_n + \bar{h} - h_0 - h_1\right)}{\sum_{k=0}^{\infty} |c_k|^2}\,.$$

However,

$$(h_n - \bar{h})\,(h_n + \bar{h} - h_0 - h_1) \geq (\bar{h} - h_0)\,(h_1 - \bar{h})\ , \quad \forall\, n\,,$$

hence

$$\Delta - \bar{h}^2 \geq (\bar{h} - h_0)\,(h_1 - \bar{h})\ .$$

Example:

Again consider the vibrating string of unit length, with $\psi(x) = x(1-x)$, $\bar{h} = 10$, $\Delta = 120$. Now we know that in 1-dimensional problems $h_1 \approx 4h_0$; for 2-dimensional problems $h_1 \approx \frac{5}{2}h_0$; and in general, $h_1 \approx \frac{n+3}{n}h_0$, where n is the dimensionality. Thus, if we let $h_1 \approx 4h_0$, we get

$$120 - 100 \geq (10 - h_0)\,(4h_0 - 10)\ ,$$

or $h_0 \geq 9.260\ldots$. Had we guessed $h_1 \approx 3h_0$, the lower bound would have been slightly worse, $h_0 \geq 8.774\ldots$.

. . . . . .

Exercise 16.8
Find a lower bound for the ground state energy of the harmonic oscillator, making a plausible guess for the energy of the first excited state.

. . . . . .

Exercise 16.9
Find upper and lower bounds for the ground state energy of the quartic anharmonic oscillator, whose Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\Omega^2 x^2 + \beta^2 x^4\,.$$

. . . . . .

16.5.3 Bounds for positive operators

A positive operator $A \geq 0$ is a Hermitian operator for which $\langle\Phi|\, A\,|\Phi\rangle \geq 0$ for any vector $|\Phi\rangle$ in the Hilbert space. We say an operator $A \geq C$ if both are Hermitian and

$$\langle\Phi|\,(A - C)\,|\Phi\rangle \geq 0\,.$$

Using the trial vector

$$|\Phi\rangle = |\Psi\rangle - \sum_{n=1}^{N} c_n\,|\varphi_n\rangle$$

where the $\{\varphi_n\}_1^N$ are any set of N vectors in $\mathcal{H}$ and Ψ is an arbitrary vector, we see that

$$\begin{aligned}\langle\Phi|\, A\,|\Phi\rangle = \langle\psi|\, A\,|\psi\rangle \quad &- \quad \sum_{n=1}^{N}\left(c_n\,\langle\Psi|\, A\,|\varphi_n\rangle + c_n^*\,\langle\varphi_n|\, A\,|\Psi\rangle\right)\\ &+ \quad \sum_{m,n} c_m^* c_n\,\langle\varphi_m|\, A\,|\varphi_n\rangle\\ &\geq \quad 0\,.\end{aligned}$$

Treating the c_n^* as variational parameters we may minimize and evaluate the expression for the minimizing values, to find

$$\langle\Phi|\, A\,|\Phi\rangle_{\min} = \langle\psi|\, A\,|\psi\rangle - \sum_{m,n}\langle\psi|\, A\,|\varphi_m\rangle\, B_{mn}\,\langle\varphi_n|\, A\,|\Psi\rangle \geq 0\,,$$

where B_{mn} is the inverse of the $N \times N$ matrix $\langle\varphi_m|\, A\,|\varphi_n\rangle$. Since Ψ is arbitrary, in operator terms we have found an operator that is a lower bound for A:

$$A \geq \sum_{m,n} A\,|\varphi_m\rangle\, B_{mn}\,\langle\varphi_n|\, A\,. \tag{16.10}$$

The virtue of the bound Eq. 16.10 is that by increasing the number of functions in the set $\{\varphi_n\}_1^N$ we can increase the lower-bounding operator and thereby improve—say—the lower bound on an eigenvalue.

.....

Exercise 16.10
The electron-electron interaction in the helium atom is manifestly positive, since

$$\begin{aligned}\langle \Psi | V | \Psi \rangle &= e^2 \int d^3 r_1 \int d^3 r_2 \frac{|\Psi(\vec{r}_1, \vec{r}_2)|^2}{|\vec{r}_1 - \vec{r}_2|} \\ &\equiv 4\pi e^2 \int d^3 R \int d^3 q \frac{\left|\tilde{\Psi}\left(\vec{R}, \vec{q}\right)\right|^2}{q^2}\end{aligned}$$

where

$$\tilde{\Psi}\left(\vec{R}, \vec{q}\right) = \int d^3 r \Psi\left(\vec{R} + \frac{1}{2}\vec{r}, \vec{R} - \frac{1}{2}\vec{r}\right) e^{i\vec{q}\cdot\vec{r}}.$$

Therefore it is possible to bound this term below by an operator of finite rank N, as in Eq. 16.10, and thereby obtain a series of (increasing) lower bounds to the ground state energy of helium. Choose two suitable trial functions and thereby obtain a rank-2 lower bound for the ground state energy of He.

.....

16.6 Feynman-Hellman theorem

Suppose a Hermitian operator $H(\lambda)$ depends on a parameter λ. Then, naturally, all the eigenvalues, $h_n(\lambda)$, and eigenvectors, $|\varphi_n(\lambda)\rangle$, will be functions of λ as well. We may express $h_n(\lambda)$ as

$$h_n(\lambda) = \frac{\langle \varphi_n(\lambda) | H(\lambda) | \varphi_n(\lambda) \rangle}{\langle \varphi_n(\lambda) \mid \varphi_n(\lambda) \rangle},$$

which gives, upon differentiation with respect to λ,

$$\frac{dh_n(\lambda)}{d\lambda} \equiv \frac{\left\langle \varphi_n(\lambda) \left| \left(\frac{\partial H(\lambda)}{\partial \lambda} \right) \right| \varphi_n(\lambda) \right\rangle}{\langle \varphi_n(\lambda) \mid \varphi_n(\lambda) \rangle}. \tag{16.11}$$

. . . . . .

Exercise 16.11
Fill in the steps necessary to derive Eq. 16.11.

. . . . . .

Equation 16.11 has many uses. For example, suppose we are looking at the bound ($E < 0$) energy eigenvalues of a particular quantum-mechanical Hamiltonian, $H_0 + \lambda V$ where we think of V as a perturbation. Then if, for example, $V < 0$ the discrete energy eigenvalues of the system are monitonically *decreasing* functions of λ. Moreover, Eq. 16.11 establishes the first-order perturbation formula,

$$h_n(\lambda) \approx h_n(0) + \lambda \frac{\langle \varphi_n(0) | V | \varphi_n(0) \rangle}{\langle \varphi_n(0) \mid \varphi_n(0) \rangle},$$

via Taylor's theorem.