Notes on Functionals

B. Svetitsky

March 11, 2005

1 Functionals and functional derivatives

Consider a function $F(y)$. You can differentiate it, $\frac{dF}{dy} = F'(y)$. Then if you start at a point y_0 and you move a distance[1] dy, the function F changes by an amount $dF = F(y_0 + dy) - F(y_0) = F'(y)|_{y_0}\, dy$.

Now consider a function of several variables, y_1, y_2, ... Writing the function as $F(y_1, y_2, \ldots)$, it has partial derivatives $\partial F/\partial y_1$, $\partial F/\partial y_2$, ... If I start at a point y_1^0, y_2^0, ... and move to a new point via a displacement dy_1, dy_2, ..., the function F will change according to

$$dF = \left.\frac{\partial F}{\partial y_1}\right|_{y^0} dy_1 + \left.\frac{\partial F}{\partial y_2}\right|_{y^0} dy_2 + \cdots \ . \tag{1}$$

Example. Take a function in 3-dimensional space, $\phi(\mathbf{r}) = \phi(x, y, z)$. Defining $d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}$, where dx, dy, and dz are independent step sizes, we have

$$d\phi = \nabla\phi \cdot d\mathbf{r} = \frac{\partial\phi}{\partial x}\,dx + \frac{\partial\phi}{\partial y}\,dy + \frac{\partial\phi}{\partial z}\,dz \ .$$

We can write the independent variables y_1, y_2, ... collectively as y_n ($n = 1, 2, \ldots$). y then looks like a function of the integer variable n. F is thus a function of the *function* y.

But a function $y(x)$ in physics usually depends on a variable x that takes on all real values in some interval $[a, b]$. To relate this to what we have

[1] Important point: dy is entirely independent of y! Think of it as a step of size ϵ, which can be chosen arbitrarily, with no relation to the place y_0 from which you step.

discussed so far, let's choose N points on the interval $[a, b]$ with the points a distance ϵ apart, where $N\epsilon = b - a$. The n^{th} point is at $x = x_n = a + n\epsilon$. We can represent the function $y(x)$ by its values on the N points, so that we consider the function $y_n = y(x_n) = y(a + n\epsilon)$, which would give more and more information about the original $y(x)$ as $N \to \infty$, $\epsilon \to 0$.

We can define a function of all the $\{y_n\}$, namely $F(\{y_n\})$. In the limit $N \to \infty$, the function F becomes a function of the function $y(x)$. We then call F a *functional* of $y(x)$, written $F[y]$. It is a function of *all* the values of $y(x)$ in the interval $[a, b]$: an *infinite* number of independent variables!

A functional takes as input a function $y(x)$ on a domain—not the value of the function at a specific point x, but *all* the values of y at *all* the x's in the domain. Its output is a *number*.

Example. Define $F[y] = \int_0^1 y(x)^2\,dx$. Then $F[y = x] = \frac{1}{3}$ and $F[y = \sin \pi x] = \frac{1}{2}$. F depends on the entire functional form of $y(x)$ in the interval $[0, 1]$.

Example. The simplest functional simply evaluates the input function at a single, specific point. For instance, if $F[y(x)] = y(3)$, then $F[y = x^2] = 9$ and $F[y = \cos \pi x] = -1$.

If we change the values of $\{y_n\}$, the function $F(\{y_n\})$ will change according to (1). Let's rewrite this as

$$dF = \sum_{n=1}^{N} \left.\frac{\partial F}{\partial y_n}\right|_{y^0} dy_n \ . \tag{2}$$

How does this look in the $N \to \infty$ limit? Recall the definition of an integral:

$$\int_a^b dx\, f(x) = \lim_{\epsilon \to 0} \sum_{n=1}^{N} \epsilon f(x_n) \ . \tag{3}$$

Rewrite (2) as

$$dF = \sum_{n=1}^{N} \epsilon \left(\frac{1}{\epsilon} \left.\frac{\partial F}{\partial y_n}\right|_{y^0} \right) dy_n \ . \tag{4}$$

Taking the limit $\epsilon \to 0$, with $x = a + n\epsilon$, and introducing the notation $dy_n = \delta y(x)$, (4) becomes

$$dF = \int_a^b dx \left.\frac{\delta F}{\delta y(x)}\right|_{y^0(x)} \delta y(x) \ . \tag{5}$$

Here $y^0(x)$ is the particular function $y(x)$ that is the starting point for the arbitrary infinitesimal change $\delta y(x)$. The $1/\epsilon$ has been absorbed into $\delta F/\delta y(x)$; this can be taken to be the definition of the *functional derivative* $\delta F/\delta y(x)$.

The meaning of (5) is the same as the meaning of (1). The change in F is a sum of terms proportional to the infinitesimal changes $\delta y(x)$, with constants of proportionality that are just the functional derivative (i.e., the partial derivatives) $\delta F/\delta y(x)$. You can think of this derivative as giving the response of the functional F to a small change in y, with the change localized at x.

The preceding discussion gives a definition of the functional derivative, but it does not give a useful method for calculating it since for each problem we would have to define carefully a mesh of points x_n and a function F of the discrete set $y(x_n)$. More usually, we have a functional $F[y]$, defined for functions y of a continuum variable x, and we need its functional derivative. We can start with (5) as a *definition* of the functional derivative, and use it to calculate.

Example. Let

$$F[y] = \int_0^1 y(x)^2\,dx\ . \tag{6}$$

To calculate the functional derivative, we calculate the change dF that is due to an infinitesimal change $\delta y(x)$ in the independent variables:

$$\begin{aligned} F[y+\delta y] &= \int_0^1 [y(x)+\delta y(x)]^2\,dx \\ &= \int_0^1 [y(x)^2 + 2y(x)\,\delta y(x) + \delta y(x)^2]\,dx \end{aligned} \tag{7}$$

Now we throw away $(\delta y)^2$, since δy is an infinitesimal and we have in mind the $\delta y \to 0$ limit. Thus to first order in δy,

$$\begin{aligned} F[y+\delta y] &= \int_0^1 [y(x)^2 + 2y(x)\,\delta y(x)]\,dx \\ &= F[y] + \int_0^1 2y(x)\,\delta y(x)\,dx\ . \end{aligned} \tag{8}$$

The infinitesimal change in F due to δy is then

$$dF = F[y+\delta y] - F[y] = \int_0^1 2y(x)\,\delta y(x)\,dx\ . \tag{9}$$

The crucial step is to compare (9) to (5). We thus identify

$$\frac{\delta F}{\delta y(x)} = 2y(x) \ . \tag{10}$$

This is the prototype for all calculations of the functional derivative.

2 Generalizations

1. A functional can depend on more than one function. Then the functional derivative with respect to each argument is defined in the obvious way, each calculated while keeping the other function constant.

   Example. If

   $$F[y(x), z(x)] = \int_0^1 y(x)^2 z(x)^3 \, dx \ . \tag{11}$$

   then

   $$\begin{aligned} \frac{\delta F}{\delta y(x)} &= 2y(x)z(x)^3 \\ \frac{\delta F}{\delta z(x)} &= 3y(x)^2 z(x)^2 \ . \end{aligned} \tag{12}$$

2. The independent function can be a function of more than one variable.

   Example. Taking $\mathcal{V}$ to be a specific domain of integration in space, let

   $$F[f(x, y, z)] = \int_{\mathcal{V}} dV \ f^3 \ . \tag{13}$$

   Then

   $$\frac{\delta F}{\delta f(x, y, z)} = 3f^2 \ , \tag{14}$$

   just as in the one-dimensional case. F can also be written as $F[f(\mathbf{r})]$.

3. A functional doesn't have to be a simple integral.

   Example. Take

   $$F[f(x, y)] = \int_0^1 dx \, x^2 \left[\int_0^1 dy \, \sin f(x, y)^4 \right]^2 \ . \tag{15}$$

   Here, as always, the definition (5) gives a straightforward method for calculating the functional derivative $\delta F / \delta f$.

3 The Euler equation

When the functional is a simple integral, Euler's equation gives a powerful formula for quick calculation of the functional derivative. Start with the case

$$F[y] = \int L(x, y(x))\, dx \ , \tag{16}$$

i.e., a simple integral where the integrand is some function of x and $y(x)$. Varying $y(x)$,

$$F[y + \delta y] = \int L(x, y + \delta y)\, dx = \int \left(L(x, y) + \frac{\partial L(x, y)}{\partial y}\, \delta y \right) dx \ , \tag{17}$$

where in the second equality we have again thrown away terms of order $(\delta y)^2$ and higher. Referring to the definition (5), we find

$$\frac{\delta F}{\delta y(x)} = \frac{\partial L(x, y)}{\partial y} \ . \tag{18}$$

Eq. (18) is a handy formula that is applicable whenever the functional is of the form (16).

Example. For $F[y] = \int x^3 e^{-y(x)}\, dx$, eq. (18) gives the result $\delta F/\delta y = -x^3 e^{-y(x)}$.

Euler's formula is not (18), but rather applies to a case somewhat more complicated. Consider the case where the integrand L contains both y and $y' = dy/dx$,

$$F[y] = \int L(x, y, y')\, dx \ . \tag{19}$$

A variation of $y(x)$ by some specific $\delta y(x)$ gives

$$F[y + \delta y] = \int L(x, y + \delta y, y' + \delta y')\, dx \tag{20}$$

where $\delta y' = d(\delta y)/dx$ is the derivative of the small variation $\delta y(x)$. Expanding to first order in δy and its derivative,

$$F[y + \delta y] \simeq \int \left(L(x, y, y') + \frac{\partial L(x, y, y')}{\partial y}\, \delta y + \frac{\partial L(x, y, y')}{\partial y'}\, \delta y' \right) dx \ . \tag{21}$$

Note that we have differentiated $L(x, y, y')$ as if y and y' are unconnected, independent variables. This is not as confusing as it sounds.

Example. If $L = x^2y^3y'^4$, then $\partial L/\partial x = 2xy^3y'^4$, $\partial L/\partial y = x^2 3y^2y'^4$, and $\partial L/\partial y' = x^2y^3 4y'^3$.

Eq. (21) is not yet in the form of (5) so we cannot yet extract the functional derivative. The problem is to turn $\delta y'$ into δy in the last term; we do this with an integration by parts:

$$\int_a^b \frac{\partial L(x,y,y')}{\partial y'}\,\delta y'\,dx = \left[\frac{\partial L(x,y,y')}{\partial y'}\,\delta y(x)\right]_a^b - \int_a^b \frac{d}{dx}\left(\frac{\partial L(x,y,y')}{\partial y'}\right)\delta y(x)\,dx\ . \quad (22)$$

The first term on the right hand side of (22) is a *boundary term*: It is proportional to the values of δy at the boundaries of the interval $[a,b]$. Using (22) in (21), we obtain

$$dF = \int \left(\frac{\partial L(x,y,y')}{\partial y} - \frac{d}{dx}\frac{\partial L(x,y,y')}{\partial y'}\right)\delta y(x)\,dx + [\textit{boundary terms}]\ . \quad (23)$$

Now we can compare to (5) and conclude that

$$\frac{\delta F}{\delta y(x)} = \frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'} \quad (24)$$

as long as x is not at a boundary of the interval; if x is at a boundary, there is an added term that can be read off (23). This is Euler's formula, an extremely useful short cut to the functional derivative when the functional is of the form (19).

Simple generalizations of Euler's formula:

1. The integrand L can contain higher derivatives of $y(x)$. The extension of Euler's equation is derived in the same way as (24).

   Example. If
   $$F[y] = \int L(x,y,y',y'')\,dx \quad (25)$$
   then
   $$\frac{\delta F}{\delta y(x)} = \frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'} + \frac{d^2}{dx^2}\frac{\partial L}{\partial y''}\ . \quad (26)$$
   The second derivative comes from the need to integrate by parts twice to deal with $\delta y''$, and two minus signs make a plus sign.

2. If the argument of the functional is a function of more than one variable, then partial derivatives can appear in L.

   Example. Let

$$F[f(x,y,z)] = \int L(x,y,z,f,f_x,f_y,f_z)\,dx\,dy\,dz\ . \tag{27}$$

   where we use the notation $f_x \equiv \frac{\partial f}{\partial x}$, etc. Then

$$\frac{\delta F}{\delta f(x,y,z)} = \frac{\partial L}{\partial f} - \frac{\partial}{\partial x}\frac{\partial L}{\partial f_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial f_y} - \frac{\partial}{\partial z}\frac{\partial L}{\partial f_z}\ . \tag{28}$$

Examples of the use of Euler's formula can be found in the next section.

4 Minimization and classical mechanics

Classical mechanics rests upon Newton's Laws. A central theme in more advanced treatments (called analytical mechanics) is the connection of Newton's Laws to an *action principle*. Mathematically, this is a simple application of what we have developed above.

Take a single particle that is free to move along a line, under the influence of a force derivable from a potential. If $x(t)$ is the coordinate of the particle, then its kinetic energy is

$$T = \frac{1}{2}m\dot{x}^2 \tag{29}$$

where we use the notation $\dot{x} \equiv \frac{dx}{dt}$ for the velocity. The potential energy is $V(x)$, so that the force acting on the particle is $F(x) = -\frac{dV}{dx}$.

Given the position x and velocity $\dot{x}$ of the particle at any given time, we can define a new function called the *Lagrangian*,

$$L = T - V = \frac{1}{2}m\dot{x}^2 - V(x)\ . \tag{30}$$

Now consider the following *boundary value problem:* Given that the particle is known to be at position x_1 at time t_1, and at some other position x_2 at a later time t_2, where was the particle at the times in between, that is, what is the function $x(t)$—the *trajectory*—for $t_1 < t < t_2$? According to

Hamilton, we begin by defining the *action* of a trajectory as the following functional[2] of $x(t)$:

$$S[x(t)] = \int_{t_1}^{t_2} L(t, x, \dot{x})\, dt \ . \tag{31}$$

Hamilton's *Principle of Least Action* states that the true trajectory $x(t)$ is that function that minimizes the action (31).

How does one minimize a function $y(x)$? One looks for points where its derivative vanishes. How does one minimize a *functional?* The same! Hamilton's Principle can be written as

$$\frac{\delta S}{\delta x(t)} = 0 \tag{32}$$

for all $t_1 < t < t_2$. The action S is precisely of the form (19), where Euler's formula applies, so (32) is equivalent to

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0 \ . \tag{33}$$

This is called the Euler-Lagrange equation of the action (31), or simply the Lagrange equation.[3]

Let's show that this really works. Inserting the Lagrangian (30) into the Lagrange equation (33), we obtain

$$\frac{dV}{dx} - \frac{d}{dt} m\dot{x} = 0 \ , \tag{34}$$

which is just $F = ma$. This is a second-order ordinary differential equation for $x(t)$. To complete the minimization of the action, one must solve (34) for $x(t)$ (this is the hard part!) to get a solution containing two arbitrary constants; then these constants are fixed by the boundary values $x(t_1)$ and $x(t_2)$.

This was a rather trivial example, but it does constitute a proof (in reverse) of Hamilton's Principle for this simple case. The real power of the Lagrangian method as derived from Least Action can only be felt when the mechanical system is complex, and we leave this to a course in Analytical Mechanics.

[2] L can contain t explicitly if the potential V is time-dependent. We ignore this in the following.

[3] Since Hamilton instructs us to solve a boundary value problem, with $x(t_1)$ and $x(t_2)$ fixed, the variation $\delta x(t)$ is fixed to be zero at t_1 and t_2. Thus x in (33) is *not* a boundary point, and there are no boundary terms to worry about.

It should be noted that setting $\delta S/\delta x(t) = 0$ does not necessarily lead to a minimum of the action. It could just as well be a maximum or a saddle point, depending on the signs of the second derivatives at the solution. Fortunately, Hamilton's Principle, while it is often called the Principle of Least Action, does not really require a minimum of S but only a *stationary point*, which means that any solution of $\delta S/\delta x(t) = 0$ will do. As we have seen, this is just what gets us back to Newton's Second Law, which is all that is needed.

This example, Hamilton's Principle in mechanics, shows the most common application of functional differentiation, namely, finding the minimum (or at least a stationary point) of a functional. Hamilton's Principle is but one example of a *variational principle*, a theme which recurs in many fields of physics. In optics one has Fermat's Law of Least Time; in quantum mechanics one has the Rayleigh-Ritz variational method and Schwinger's action principle; and in classical mechanics there are other forms of Least Action in other contexts. Most mathematics textbooks discuss functional methods with these applications in mind, so they treat only the minimization problem and not the wider range of application of functional differentiation. The most common name for this subject in the textbooks is *calculus of variations*.

A slightly more complex case is that of a particle free to move in three dimensions, so that it has three coordinates $\mathbf{r}(t) = [x(t), y(t), z(t)]$. Then the action is a functional of the three functions x, y, z,

$$S[\mathbf{r}(t)] = \int_{t_1}^{t_2} L(t, \mathbf{r}, \dot{\mathbf{r}})\, dt \ . \tag{35}$$

The Lagrange equations come from varying separately with respect to x, y, and z,

$$0 = \frac{\delta S}{\delta x(t)} = \frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} \ , \tag{36}$$

with another equation for $y(t)$ and another for $z(t)$. The three equations can be summarized as

$$0 = \nabla_{\mathbf{r}} L - \frac{d}{dt}\nabla_{\dot{\mathbf{r}}} L. \tag{37}$$

The first ∇ is just the ordinary gradient, containing derivatives with respect to the components of $\mathbf{r}$; the second ∇ contains derivatives with respect to the components of $\dot{\mathbf{r}}$:

$$\nabla_{\dot{\mathbf{r}}} L = \mathbf{i}\frac{\partial L}{\partial \dot{x}} + \mathbf{j}\frac{\partial L}{\partial \dot{y}} + \mathbf{k}\frac{\partial L}{\partial \dot{z}} \ . \tag{38}$$

Example. Take a particle in a central potential, for which

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(r) \ . \tag{39}$$

Then

$$\nabla_{\mathbf{r}} L = -V'(r)\nabla r = -V'(r)\hat{\mathbf{r}} \tag{40}$$

$$\nabla_{\dot{\mathbf{r}}} L = m\dot{\mathbf{r}} \tag{41}$$

and so the Lagrange equations are

$$m\ddot{\mathbf{r}} = -\hat{\mathbf{r}}V'(r) \tag{42}$$

Of course, the problem can also be done component by component, using (36) and its siblings.

5 Field theory

Let us continue our visit to classical mechanics. We have discussed the action principle as it applies to a single particle moving in one and three dimensions. Now consider the mechanics of a violin string, which is allowed to vibrate up and down (but not sideways) while tied at its ends. An element of the string is denoted by its coordinate x along the string, and in its motion it is displaced in a direction transverse to the string by a distance $\psi(x,t)$. Now the function ψ constitutes an infinite set of coordinates for which x is the index. (Recall the correspondence $y_n \to y(x)$ in Section 1.) A function $\psi(x,t)$ of space and time is a trajectory of the entire string. It is not hard to show that the kinetic and potential energies of the string are given by

$$T = \int_0^\ell dx\, \frac{\mu}{2}\left(\frac{\partial\psi}{\partial t}\right)^2 \tag{43}$$

$$V = \int_0^\ell dx\, \frac{\tau}{2}\left(\frac{\partial\psi}{\partial x}\right)^2 \tag{44}$$

where ℓ is the length of the string; μ is its density (mass per unit length); and τ is the tension of the string. As before $L = T - V$, and thus the action

is

$$\begin{aligned} S[\psi(x,t)] &= \int dt\,(T-V) \\ &= \int dt\,dx\,\left[\frac{\mu}{2}\left(\frac{\partial\psi}{\partial t}\right)^2 - \frac{\tau}{2}\left(\frac{\partial\psi}{\partial x}\right)^2\right] . \end{aligned} \tag{45}$$

This is of the form

$$S[\psi(x,t)] = \int dt\,dx\,\mathcal{L}(\psi,\psi_x,\psi_t) \tag{46}$$

and so the Euler-Lagrange equation is [cf. (28)]

$$\begin{aligned} 0 &= \frac{\delta S}{\delta\psi(x,t)} \\ &= \frac{\partial\mathcal{L}}{\partial\psi} - \frac{\partial}{\partial x}\frac{\partial\mathcal{L}}{\partial\psi_x} - \frac{\partial}{\partial t}\frac{\partial\mathcal{L}}{\partial\psi_t} \\ &= \tau\frac{\partial^2\psi}{\partial x^2} - \mu\frac{\partial^2\psi}{\partial t^2} \end{aligned} \tag{47}$$

This partial differential equation is the wave equation,

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} , \tag{48}$$

where we identify the wave velocity as $v = \sqrt{\tau/\mu}$.

The string is the simplest example of a problem in *continuum dynamics,* also called *classical field theory.* $\psi(x,t)$ is the *field,* a variable associated with each point in space that evolves dynamically in time. (Think of the electric field, governed by Maxwell's Equations.) S is the action for the field theory, and the wave equation is the equation of motion, which is a partial differential equation because of the added element of x-dependence.

For any trajectory $\psi(x,t)$, the energy is the *sum* of the kinetic and potential energies,

$$\begin{aligned} E[\psi(x,t)] &= T+V \\ &= \int dx\,\left[\frac{\mu}{2}\left(\frac{\partial\psi}{\partial t}\right)^2 + \frac{\tau}{2}\left(\frac{\partial\psi}{\partial x}\right)^2\right] . \end{aligned} \tag{49}$$

Note that there is no integral over time in defining the energy. One might be interested in finding the time-independent shape of the string that minimizes the energy. Here the kinetic energy is zero, and one minimizes the functional

$$E[\psi(x)] = \int dx \, \frac{\tau}{2} \left(\frac{\partial \psi}{\partial x} \right)^2 \quad . \tag{50}$$

The Euler equation here is

$$\frac{d^2\psi}{dx^2} = 0 \tag{51}$$

which gives the trivial solution (satisfying the boundary conditions) $\psi = 0$. This is also obvious from the form of (50). It is of course easy to invent an energy-minimization problem with a non-trivial solution.[4] The point of this exercise is to show that the same problem might invite both time-dependent and time-independent variational calculations.

Of course, not all field theory problems are one-dimensional like the vibrating string. Examples of two-dimensional field theory are the vibrations of a drumhead or surface waves in water, where the physics is described by a function $\psi(x, y, t)$ and an action functional $S = \int dt \, dx \, dy \, \mathcal{L}(\psi, \psi_t, \psi_x, \psi_y)$. Electromagnetism is a *three*-dimensional field theory, with six separate fields that are the components of $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$.

6 Constraints

Frequently a minimization problem includes a constraint. In ordinary multivariable calculus, for instance, one might be asked to find the minimum of $f(x, y, z)$ under the constraint that $g(x, y, z) = 0$. This is usually done by the method of Lagrange multipliers.

Example. To minimize $f(x, y) = (x^2+y^2-1)^2$ with the constraint $x-y^2 = 2$, we introduce a Lagrange multiplier λ and define $h(x, y) = f(x, y) + \lambda(x - y^2)$. Now we minimize h in the usual way:

$$0 = \frac{\partial h}{\partial x} = 4x(x^2 + y^2 - 1) + \lambda \tag{52}$$

$$0 = \frac{\partial h}{\partial y} = 4y(x^2 + y^2 - 1) - 2\lambda y \tag{53}$$

[4] Just change the boundary conditions on ψ, for example. A more interesting alternative is to add $\int dx \, \left(a\psi^4 - b\psi^2 \right)$ to the energy (50).

These two equations, together with the constraint $x-y^2=2$, give three equations for the three unknowns x, y, and λ.

An explanation of why the method works can be found in most first-year textbooks.

If there is more than one constraint, we use more than one Lagrange multiplier, one multiplier for each constraint.

Example. To minimize $f(x,y,z)=(x^2+y^2+z^2-1)^2$ with the constraints $x-y^2=2$, $x^2+y^3+z^4=4$, we introduce Lagrange multipliers λ_1 and λ_2 and define $h(x,y,z)=f(x,y,z)+\lambda_1(x-y^2)+\lambda_2(x^2+y^3+z^4)$. Setting three partial derivatives ∇h to zero, along with the two constraints, gives five algebraic equations for the five unknowns x, y, z, λ_1, λ_2.

Constrained minimization of a *functional* is no different. To minimize $S[y(x)]$ under the constraint $T[y]=0$, define the combined functional $R[y]=S[y]+\lambda T[y]$. Then write the minimization equation

$$\frac{\delta R}{\delta y(x)}=0\ . \tag{54}$$

This will yield an equation for $y(x)$ as always, with λ appearing as a parameter. The constraint equation $T[y]=0$ gives another equation, so that there is enough to fix λ, too.

Example. Let's minimize $S[y(x)]=\int_0^1 y'^2\,dx$, subject to the boundary conditions $y(0)=y(1)=0$ and to the constraint $T[y]\equiv\int_0^1 y\,dx=1$. Then $R[y]=\int_0^1 (y'^2+\lambda y)\,dx$, and the minimization (54) gives the Euler equation

$$\lambda-2y''=0\ . \tag{55}$$

This has the general solution $y=\frac{1}{4}\lambda x^2+c_1x+c_2$. Plugging this into the constraint gives $T[y]=\frac{1}{12}\lambda+\frac{1}{2}c_1+c_2=1$. Solving this equation together with the boundary conditions gives $c_1=6$, $c_2=0$, and $\lambda=-24$, so the solution is $y=6x(1-x)$.

If there is more than one constraint, say $T_1[y]=0$ and $T_2[y]=0$, then we use two Lagrange multipliers and we minimize $R[y]=S[y]+\lambda_1T_1[y]+\lambda_2T_2[y]$.

A more complex problem is that of a *locally constrained* minimization. Here there is typically a functional of (at least) *two* functions, $S[x(t),y(t)]$,

along with an equation that constrains $x(t)$ and $y(t)$ *at every time* t. Since there are an infinite number of constraints—one at every time t—there must be an infinite number of Lagrange multipliers—one at every time t. So we have a Lagrange multiplier *function* $\lambda(t)$. If the constraint is $f(x(t), y(t)) = 0$ then we define the new functional

$$R[x(t), y(t)] = S[x, y] + \int \lambda(t) f(x(t), y(t))\, dt \ . \tag{56}$$

When we minimize R with respect to $x(t)$ and $y(t)$, we get differential equations in which the unknown *function* $\lambda(t)$ appears. In addition, the algebraic equation $f(x(t), y(t)) = 0$ is available to help fix $\lambda(t)$.

Example. Consider a particle moving in a plane and constrained to move along the curve $y = x^2$. The action is

$$S[x(t), y(t)] = \int \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right)\, dt \tag{57}$$

and the constraint is $y - x^2 = 0$. We define

$$R[x(t), y(t)] = \int \left[\frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right) + \lambda(t) \left(y - x^2\right)\right]\, dt \ , \tag{58}$$

the minimization of which gives the Euler-Lagrange equations

$$m\ddot{x} + 2\lambda(t) x = 0 \tag{59}$$

$$m\ddot{y} + \lambda(t) = 0 \tag{60}$$

Since in these differential equations $\lambda(t)$ is an unknown function, their solution is quite difficult.

The example shows that the Lagrange multiplier method for local constraints is not really a calculational tool. It is, however, very important as a conceptual device in the analytical mechanics of constrained systems, and reappears in classical and quantum field theory.

7 The delta function

Finally, let us return to the simplest functional of all,

$$F[y(x)] = y(x_0) \ , \tag{61}$$

which just evaluates its argument $y(x)$ at the specific point x_0. What is its functional derivative $\delta F/\delta y(x)$? F only depends on the value of y at x_0, so if we vary y at some *other* place, F will be unchanged. Thus

$$\frac{\delta F}{\delta y(x)} = 0, \qquad x \neq x_0 \ . \tag{62}$$

If we change y exactly at x_0, however, F obviously changes.

An easy way to proceed is to rewrite (61) as an integral functional. Let's say there is some function $\delta(x - x_0)$ that allows us to write

$$F[y(x)] = \int \delta(x - x_0)\, y(x)\, dx \ . \tag{63}$$

What are the properties of $\delta(x)$? Since $F[y]$ has no dependence on $y(x)$ for $x \neq x_0$, we clearly have

$$\delta(x - x_0) = 0 \text{ whenever } x \neq x_0 \ . \tag{64}$$

What is $\delta(x - x_0)$ at $x = x_0$? If it were finite, the integral (63) would always be zero, because of the infinitesimal measure dx. Thus $\delta(x - x_0)$ must be *infinite* at $x = x_0$!

The actual value of $\delta(0)$ will never really concern us. All we have to know is how to integrate with $\delta(x)$. In fact, we can take the following [which is just (61) and (63)] to be the *definition* of $\delta(x)$:

$$y(x_0) = \int_a^b \delta(x - x_0)\, y(x)\, dx \ . \tag{65}$$

In view of (63), we can state the following result, which must be used with care:

$$\frac{\delta F}{\delta y(x)} = \delta(x - x_0) \ , \tag{66}$$

or, more compactly,

$$\frac{\delta y(x_0)}{\delta y(x)} = \delta(x - x_0) \ . \tag{67}$$

Properly speaking, $\delta(x)$ is not a function at all, since its infinite value takes us out of the usual domain of definition of functions. Mathematicians call it a *distribution*, a limit of a sequence of functions that really only has

meaning in integral expressions such as (65). Let us evaluate (65) for the special case $y(x) = 1$, choosing as well $x_0 = 0$. We get

$$1 = \int_a^b \delta(x)\, dx \ . \tag{68}$$

So the area under the δ function is 1 (even though its width is zero!). One possible realization of $\delta(x)$ as a sequence of functions is the set of gaussians

$$\delta_N(x) = \sqrt{\frac{N}{\pi}} e^{-Nx^2} \ . \tag{69}$$

Each δ_N has unit area, and δ_N becomes higher and narrower as $N \to \infty$. Mathematicians will always be careful to insert $\delta_N(x)$ into integrals like (65) and to evaluate the integral *before* taking the $N \to \infty$ limit.

Expressions such as (67) have meaning only when they are multiplied by some function of x_0 and integrated over x_0; then we are returned to the usual kind of functional derivative. At a physicist's level of rigor, however, (67) can be used to deal with chain rules and with higher functional derivatives, taking care to remember that the limit $N \to \infty$ may contain pitfalls.

References

[1] F. B. Hildebrand, *Advanced Calculus for Applications.*

[2] H. Goldstein, *Classical Mechanics.*

[3] V. I. Arnold, *Mathematical Methods of Classical Mechanics.*