1 Constraint forces: D'Alembert's principle

Many forces that one deals with in mechanics are specified straightforwardly in terms of direction and magnitude. Constraint forces are different. The strength of a constraint force is whatever it takes to maintain the associated restriction on the motion of the system. A very familiar constraint from freshman physics is the requirement that a sliding block stay on an incline. The constraint force is the normal force the incline provides to counter the component of the weight in that direction. For a single particle, a holonomic constraint is a restriction that the particle live on a smooth surface or curve, which may change with time. Corners and edges are difficult things to deal with, and technically disqualify a constraint from being holonomic. If the motion never comes near the edge, however, you can always pretend it doesn't exist. For systems of many particles, the unconstrained configuration space is $\mathcal{Q}_0 = E_{3N}$, a $3N$-dimensional Euclidean space with three coordinates for each particle. Holonomic constraints restrict the configuration to a smooth lower-dimensional "surface" within this space. Sometimes the reduction in dimension is very dramatic. A rigid body for instance has a gazillion constraints keeping all the particles in their proper relative positions.

Although the magnitude of the forces which maintain constraints cannot always be determined easily, there is a general principle about their direction, which goes by the name

D'Alembert's principle:

> Freeze the configuration of a mechanical system and the constraint forces $\mathbf{F}^c_\alpha$ acting on it at some instant. Let $\delta\mathbf{x}_\alpha$ be infinitesimal variations of the positions staying within $\mathcal{Q}(t)$. Then
>
> $$\delta W \stackrel{def}{=} \sum_\alpha \mathbf{F}^c_\alpha \cdot \delta\mathbf{x}_\alpha = 0.$$

This covers both time-independent and time-dependent constraints. For the moment, we'll think about only time-independent holonomic constraints, so that the motion must always lie on some fixed smooth configuration space $\mathcal{Q} \subset \mathcal{Q}_0$. Infinitesimal variations of position conforming to the constraint $\mathcal{Q}(t)$ are sometimes called 'virtual displacements'. Then D'Alembert's principle says that the constraint forces do no work under virtual displacements. To understand it, consider first a single particle constrained to some (2-d) surface. The

principle is equivalent to the statement that the constraint force is perpendicular to surface. If it were not, the constraint force would be pushing the particle in some direction *within* the constraint surface. That clearly makes no sense. By definition, the constraint acts only to keep the particle on the surface and doesn't care precisely where on the surface it is. The work done by the constraint force under some tiny displacement measures the degree to which the force wants the particle to move in that direction. Once this simple case is understood, the general one is not far behind. The tiny motions which satisfy the constraint may be more complicated, but the idea is precisely the same. δW indicates the degree to which the constraint forces are encouraging the system to move in the direction specified by $\{\delta \mathbf{x}_\alpha\}$. The very definition of a constraint says that it is indifferent to which direction things move as long as they conform to the constraint. So, δW must be zero.

We can also extend the form which says that the constraint force is perpendicular to the constraint surface by making appropriate definitions (this will be more generally useful,too). Arrange the forces on the various particles into one long vector

$$\mathbf{F}^c = \begin{pmatrix} \mathbf{F}_1^c \\ \mathbf{F}_2^c \\ \vdots \\ \mathbf{F}_N^c, \end{pmatrix}$$

and the coordinates of all the particles into another one

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N. \end{pmatrix}$$

Now we agree that dot products of such giant vectors are to be taken by multiplying each of the corresponding $3N$ components together, i.e.,

$$\mathbf{F}^c \cdot \delta\mathbf{X} \stackrel{def}{=} \sum_\alpha \mathbf{F}_\alpha^c \cdot \delta\mathbf{x}_\alpha.$$

Then, D'Alembert's priciple says that the constraint force $\mathbf{F}_c$ is orthogonal to the constraint surface, since it is orthogonal to any direction $\delta\mathbf{X}$ tangent to it.

So that's it for time-independent constraints. Time-dependent constraints are a little trickier, but not much. Now we are dealing with a constraint surface $\mathcal{Q}(t) \subset \mathcal{Q}_0$ which moves as time passes. The motion of the surface is everywhere and at all times perpendicular to itself. This is a matter of definition again. Shift a plane parallel to itself. Have you really changed anything? No. The only difference from the time-independent case is that the constraint forces are now required not to maintain zero velocity orthogonal to the surface, but a possibly non-zero component in those directions. (there can be many directions orthogonal to the surface $\mathcal{Q}(t)$ – consider a bead sliding on a wire for a very simple example.) Thus, the same argument as before carries over. The constraint forces are, in this case also, totally indifferent to motion tangent to $\mathcal{Q}(t)$.

You can make up arguments that look fancier, but they have this idea hidden in their starting points.

2 brief review of variational calculus

2.1 Functionals depending on a single function

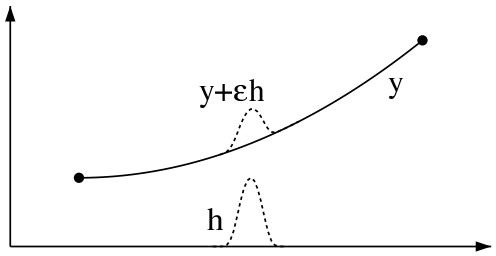
Let's recall some of the basic results from the calculus of variations. Going back to the beginning, consider a functional

$$I[y] = \int_a^b f(y(x), y'(x), x)\, dx. \tag{1}$$

The domain of I is the smooth functions having specified values $y(a) = y_a$ and $y(b) = y_b$ at the endpoints. You can take 'smooth' to mean sufficiently differentiable to make the manipulations meaningful. (In the current context, twice continuously differentiable will do – but that's a technical matter which you shouldn't worry about.)

Now alter $y(x)$ slightly by adding $\epsilon h(x)$ to it, with $h(a) = h(b) = 0$ (so as not to disturb the values of the function at the endpoints). The shape of this deformation is specified by $h(x)$ and the size is controlled by ϵ. Then the value of the functional I changes by

$$\begin{aligned} I[y + \epsilon h] - I[y] &= \int \{f(y(x) + \epsilon h(x), y'(x) + \epsilon h'(x), x) - f(y(x), y'(x), x)\}\, dx \\ &= \epsilon \int \sum_i \frac{\delta I}{\delta y(x)} h(x)\, dx + \mathcal{O}(\epsilon^2). \end{aligned} \tag{2}$$



$\mathcal{O}(\epsilon^2)$ indicates a 'correction term' which vanishes at least as fast as a constant times ϵ^2 as $\epsilon \to 0$. The first factor inside the integral is called the functional derivative of S (with respect to $y(x)$) and is given by

$$\frac{\delta I}{\delta y(x)} = \left\{ \frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} \right\} \Bigg|_{y(x),y'(x)}. \tag{3}$$

This formula is central to the whole business, but you already have derivations available from several sources, so I won't repeat it here.

The name 'functional derivative' requires a little explanation. First, differentiating equation (2),

$$\frac{d}{d\epsilon} I[y + \epsilon h]\Big|_{\epsilon=0} = \int \sum_i \frac{\delta I}{\delta y(x)} h(x)\, dx. \tag{4}$$

Now consider a special choice of h, the 'bump function' depicted in figure 2.1, with an area one under it and centered at x_0. $\delta I/\delta y(x)$ is a continuous function, so if the bump is made narrow enough,

$$\frac{d}{d\epsilon} I[y + \epsilon h]\Big|_{\epsilon=0} \approx \frac{\delta I}{\delta y(x_0)}.$$

The functional derivative $\delta I/\delta y(x)$ measures the response of I to a change of the function y near x.

In a similar vein, an analogy can be drawn to a function $f(x_1, \ldots, x_n)$ of n variables, which we view as cartesian coordinates in a Euclidean space. The change in f under a small displacement of the argument is

$$f(\mathbf{x} + \epsilon\mathbf{y}) - f(\mathbf{x}) = \epsilon \sum_i [\nabla f]_i y_i + \mathcal{O}(\epsilon^2). \tag{5}$$

The structures of Eqns. (5) and (2) are exactly parallel. It is almost as though $\delta S/\delta q^i(t)$ and $h^i(t)$ were the components of infinite dimensional vectors, each different i and t corresponding to a different component, except that instead of a sum over the t label we have an integral. This is a good rough way to think about it.

A slightly different notation one often encounters is

$$\delta I = \int \frac{\delta I}{\delta y(x)} \delta y(x)\, dt. \tag{6}$$

This says the same thing as Eq. (2). The variation ϵh of the trajectory has become δy. δI is the 'first-order' variation of I, the meaning of which is obtained by putting the ϵ back in.

Starting at any point, you can find a local minimum of a function f by just pushing that point backwards against the gradient ∇f until you reach a point where the gradient vanishes. There's no guarantee you'll ever find a minimum, but going downhill is the best way to find the bottom of the valley if it has one. To be completely honest, it may not be a minimum in a strict sense. Consider $z(x, y) = x^2$. The described procedure will run you down to $x = 0$ without changing y, because the function is horizontal in that direction.

Now, we can do exactly the same thing with functionals. Take the specific example,

$$I[y] = \int_a^b (y')^2\, dx.$$

Then, you quicky compute

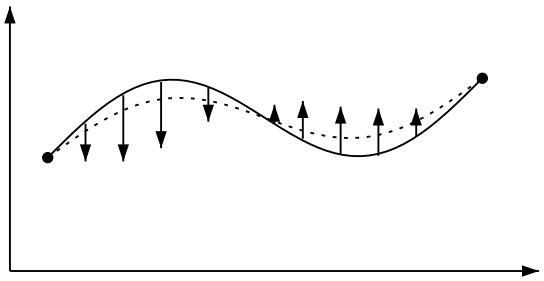
$$\frac{\delta I}{\delta y(x)} = -y''(x).$$

Figure 2.1 shows a function (solid line) which runs between the specified endpoints. At each point along the graph, attach an arrow of length $|\delta I/\delta y(x)|$ and pointing up if the functional derivative is negative, down if it's positive, i.e. opposite $\delta I/\delta y(x)$. pointing up if By the foregoing discussion, it should be clear that shifting y in the direction of the arrows at every point will reduce the value of the functional I. To be very specific, if you take

$$h(x) = -\frac{\delta I}{\delta y(x)},$$

and plug this into equation (2), you will see that every point x gives a negative (or maybe zero) contribution to the integral and I is bound to decrease. So

shift y a bit, which gives the dotted curve, recompute the functional derivative, draw new arrows, shift a bit more,... and eventually you will reach a point where $\delta I/\delta y \equiv 0$ (close anyway). No further shift of y will reduce I. (At least to first order in the shift. There's always the possibility of saddle points, of course, just as with ordinary functions, e.g. $z = x^2 - y^2$ at $x = y = 0$). What



you will wind up with in this example is a straight line. The I you're trying to minize is the integral of $(y')^2$. $y' = 0$ would surely do that, but it's not allowed because it's got to go from A to B. The best you can do is a straight line between them.

So now you have obtained a function y_0 for which

$$\frac{d}{d\epsilon} I[y_0 + \epsilon h]\Big|_{\epsilon=0} = 0$$

for every allowed h in equation (2). Equivalently,

$$\frac{\delta I}{\delta y(x)}\Big|_{y_0} = 0 \quad \text{at all } x,$$

or $\delta S = 0$ (equation 6), which in turn is the same thing as

$$\left\{ \frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} \right\}\Bigg|_{y_0, y_0'} = 0.$$

This is the functional analog of a critical point, and there's nothing wrong with calling it that. A different language is customary, however. We say that I is stationary at y_0.

2.2 Functionals depending on many functions

In our use of the functional calculus, the single function y will be replaced by the trajectory of a mechanical system, i.e. positions as a function of time. So we have to think how all this gets modified. If we have a functional $I[y_1, y_2, \ldots, y_n]$ depending on y_1, y_2,..., y_n, we can, for the most part, think of it as a functional $I[\mathbf{y}]$ of a vector-valued function $\mathbf{y}(x)$. Then y_1, y_2, ..., y_n are the components of this vector. Notice that it may perfectly well have more than three components. This assumes that the y_i are independent of each other. Drawing arrows as was done in figure 2.1, you'd get something

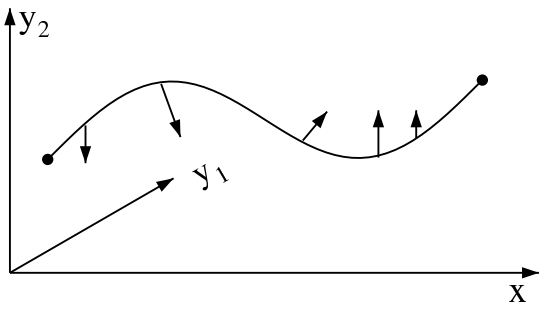


Figure 1: A function $\mathbf{y}(x)$ taking values in a two-dimensional space

like figure 1. Note that the arrows here are all parallel to the y_1-y_2 plane, they don't have any x-component. This picture gets a little unwieldy and it's easier to supress x entirely and just draw the projection into the y_1-y_2 plane. If you do that, you'll get pictures like figure 2. You must remember that the curve here is parametrized by x. To minimize I, drag each point along the direction of the arrow, being careful to keep its x-label attached.

Since each y_i can be varied independently of the others, a stationary function $\mathbf{y}$ for I must be stationary under variations of each y_i alone, leaving all the other y's fixed. What that means is that the stationarity condition is given by the collection of Euler equations

$$\frac{\delta I}{\delta y_i(x)}\Big|_{\mathbf{y}_0} = \left\{\frac{\partial f}{\partial y_i} - \frac{d}{dx}\frac{\partial f}{\partial y_i'}\right\}\Bigg|_{\mathbf{y}_0,\mathbf{y}_0'} = 0, \quad \text{for all } i. \tag{7}$$

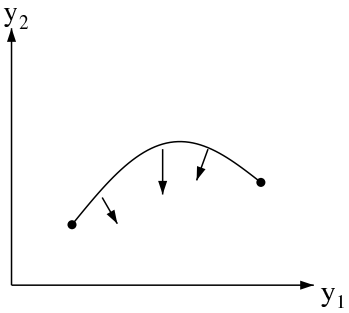


Figure 2: Projection of $\mathbf{y}(x)$ into the y_1-y_2 plane.

An important thing to notice: any complete set of coordinates will work here. The condition

$$\frac{d}{d\epsilon} I[\mathbf{y} + \epsilon \mathbf{h}] = 0$$

does not depend upon what coordinates we use to describe the functions $\mathbf{y}(x)$ and $\mathbf{h}(x)$. So stationarity of I is completely independent of such a choice. Whatever coordinates we choose, we'll still get a set of n Euler equations. The explicit form in terms of the y's will depend upon that choice because when f is rewritten in terms of new ones, it will look different. That's a good thing, because a smart choice of coordinates may make the equations much simpler.

3 Hamilton's Principle and Newton's Equation of Motion

For each configuration space trajectory $q(t)$ of a mechanical system between an initial time t_0 and final time t_1, we define the action

$$S[q] = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)\, dt$$

where the Lagrangian is

$$L = T - U$$

for a purely mechanical system. This notion is the main character in Hamilton's Principle, which can be taken as the fundamental principle of mechanics, replacing Newton's Laws. **Hamilton's Principle** asserts that

> The actual trajectory of a system between initial configuration q at time t_0 and final configuration $\overline{q}$ at time $\overline{t}$, is one which makes the action stationary.

If the system is subject to constraints, it is implicitly understood that the action is to be restricted to trajectories which actually obey them. In other words, if you vary the trajectory to one which violates the constraints, Hamilton's Principle doesn't tell you what will happen.

We need to see that this is equivalent to Newton's equation of motion (2nd Law) in order to accept it as an alternative fundamental principle. To make things more concrete, we'll focus on the specific example of a single particle constrained to lie in a surface $\mathcal{Q}$. Our particle is then acted upon by a constraint force and another force which derives from a potential $U(\mathbf{r})$. When you have understood this, you should have no difficulty whatever in dealing with the general case. (If you want to be even more specific, you could think about a sphere. Then the constraint is $|\mathbf{r}|^2 = R^2$.)

Now, pick any trajectory $\mathbf{r}(t)$ with the designated starting and ending points. From

$$L = T - U = \frac{1}{2}m|\dot{\mathbf{r}}|^2 - U(\mathbf{r})$$

evaluate

$$\frac{\delta S}{\delta \mathbf{r}(t)} = \frac{\partial L}{\partial \mathbf{r}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}} = -\frac{\partial U}{\partial \mathbf{r}} - \frac{d\mathbf{p}}{dt}. \tag{8}$$

This is a vector at each point of your chosen trajectory. Since $-\partial U/\partial \mathbf{r}$ is the force provided by the potential, Newton's 2nd law asserts that, *in order for the particle to move along your chosen trajectory*,

$$\mathbf{F}_{\text{extra}} = \frac{d\mathbf{p}}{dt} - \frac{\partial U}{\partial \mathbf{r}}$$

is the force which must supplied from somewhere besides the potential U. But there is only one other source of forces on the particle, the costraint. The constraint force, by its very nature acts perpendicular to the constraint surface

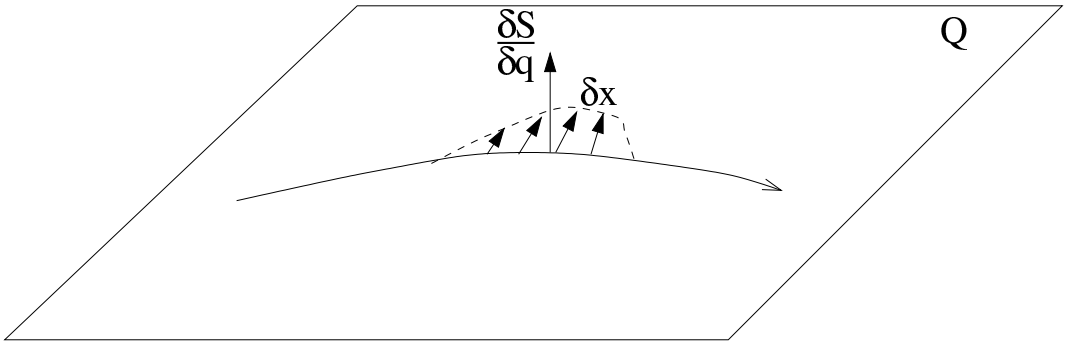


Figure 3: When $\delta S/\delta \mathbf{r}$ is perpendicular to $\mathcal{Q}$, the action is stationary to variations of the trajectory which stay within $\mathcal{Q}$.

$\mathcal{Q}$. Putting these last two equations together

$$\frac{\delta S}{\delta \mathbf{r}(t)} = -\mathbf{F}_{\text{extra}}. \tag{9}$$

Now, on the one hand, Newton assures us that, unless this vector is perpendicular to $\mathcal{Q}$ at all points of the trajectory, we've made a bad guess – the trajectory is wrong. On the other hand, the variational calculus says that we can reduce the action $S[q]$ of the trajectory by shifting it along the direction F_{extra}. Since we're not going to allow the trajectory to be pulled off the constraint surface $\mathcal{Q}$, we can only shift along the projection of $\delta S/\delta \mathbf{r}$ into the plane tangent to $\mathcal{Q}$. Thus, restricted to trajectories lying in $\mathcal{Q}$, the action is stationary precisely when $\delta S/\delta \mathbf{r}$ is perpendicular to $\mathcal{Q}$. (See figure 3) In other words, when $\delta S/\delta \mathbf{r}$ is perpendicular to $\mathcal{Q}$ at all points of the trajectory, there is no way

$$\delta S = \int \frac{\delta S}{\delta \mathbf{r}(t)} \delta \mathbf{r}(t),$$

can be anything but zero if $\delta \mathbf{r}$ is tangent to the surface.

Conclusion: Newton's Law is satisfied for precisely the same trajectories that make the action stationary among trajectories which satisfy the constraint.

3.1 parting shots

One very nice things about the Lagrangian approach is that we can avoid dealing with the constraint forces explicitly. We simply set up the variational principle for motions which are conformant to the constraint and go from there. We also are free to use any coordinates to describe the motion in $\mathcal{Q}$. This is indeed possible using Newtonian methods, too, but it's generally a lot more cumbersome.

But, if we like, we are also free to use coordinates on the bigger configuration space $\mathcal{Q}_0$, and incorporate the constraint by a Lagrange multiplier function. Namely, we can usually express a constraint as $g(\mathbf{X}(t)) = t$. This is actually very easy. Just find the constraint surface $\mathcal{Q}(t)$ and define g to be equal to t there! (Problems with multi-valuedness are illusory, since this only need work for short intervals of time and small pieces of configuration space) For example, a bead on a wire rotating around the z-axis is constrained to $\phi = \omega t$ in spherical coordinates, so take $g(\mathbf{r}) = \phi/\omega$. Then, since the constraint force acts perpendicular to the surface of constant g, it can be faked by use of a possibly time-dependent potential $\lambda(t)g$. We just have to tune $\lambda(t)$ to ensure that the resulting motion really does follow the constraint.