Physics 6010, Fall 2016

Constraints and Lagrange Multipliers.

Relevant Sections in Text: §1.3–1.6

Constraints

Often times we consider dynamical systems which are defined using some kind of restrictions on the motion. For example, the spherical pendulum can be defined as a particle moving in 3-d such that its distance from a given point is fixed. Thus the *true* configuration space is defined by giving a simpler (usually bigger) configuration space along with some *constraints* which restrict the motion to some subspace. Constraints provide a phenomenological way to account for a variety of interactions between systems. We now give a systematic treatment of this idea and show how to handle it using the Lagrangian formalism.

For simplicity we will only consider *holonomic* constraints, which are restrictions which can be expressed in the form of the vanishing of some set of functions – the constraints – on the configuration space and time:

$$C_\alpha(q,t) = 0, \quad \alpha = 1, 2, \dots m.$$

We assume these functions are smooth and independent so that if there are n coordinates q^i, then at each time t the constraints restrict the motion to a nice $n-m$ dimensional space. For example, the spherical pendulum has a single constraint on the three Cartesian configruation variables (x, y, z):

$$C(x, y, z) = x^2 + y^2 + z^2 - l^2 = 0.$$

This constraint restricts the configuration to a two dimensional sphere of radius l centered at the origin. To see another example of such constraints, see our previous discussion of the double pendulum and pendulum with moving point of support.

We note that the constraints will restrict the velocities:

$$\frac{d}{dt}C_\alpha = \frac{\partial C_\alpha}{\partial q^i}\dot{q}^i + \frac{\partial C_\alpha}{\partial t} = 0.$$

For example in the spherical pendulum we have

$$x\dot{x} + y\dot{y} + z\dot{z} = 0.$$

There are two ways to deal with such constraints. Firstly, one can simply solve the constraints, *i.e.,* find an independent set of generalized coordinates. We have been doing this all along in our examples with constraints. For the spherical pendulum, we solve the constraint by

$$x = l \sin\theta\cos\phi, \quad y = l\sin\theta\sin\phi, \quad z = l\cos\theta,$$

and express everything in terms of θ and ϕ, in particular the Lagrangian and EL equations. In principle this can always be done, but in practice this might be difficult. There is another method in which one can find the equations of motion without having to explicitly solve the constraints. This is known as the method of *Lagrange multipliers*. This method is not just popular in mechanics, but also features in "constrained optimization" problems, *e.g.,* in economics. As we shall see, the Lagrange multiplier method is more than just an alternative approach to constraints – it provides additional physical information about the forces which maintain the constraints.

Lagrange Multipliers

The method of Lagrange multipliers in the calculus of variations has an analog in ordinary calculus. Suppose we are trying to find the critical points of a function $f(x, y)$ subject to a constraint $C(x, y) = 0$. That is to say, we want to find where *on the curve defined by the constraint* the function has a maximum, minimum, saddle point. Again, we could try to solve the constraint, getting a solution of the form $y = g(x)$. Then we could substitute this into the function f to get a (new) function $h(x) = f(x, g(x))$. Then we find the critical points by solving $h'(x) = 0$ for $x = x_0$ whence the critical point is $(x_0, g(x_0))$ This is analogous to our treatment of constraints in the variational calculus thus far (where we solved the constraints via generalized coordinates before constructing the Lagrangian and EL equations). There is another method, due to Lagrange, which does not require explicit solution of the constraints and which gives useful physical information about the constraints.

To begin with, when finding a critical point (x_0, y_0) subject to the constraint $C(x, y) = 0$ we are looking for a point on the curve $C(x, y) = 0$ such that a displacement tangent to the curve does not change the value of f to first order. Let the tangent vector to $C(x, y) = 0$ at the point (x_0, y_0) on the curve be denoted by $\vec{t}$. We want

$$\vec{t} \cdot \nabla f(x_0, y_0) = 0 \quad \text{where } C(x_0, y_0) = 0.$$

Evidently, at the critical point the gradient of f is orthogonal to the curve $C(x, y) = 0$. Now, any vector orthogonal to the curve – orthogonal to $\vec{t}$ at (x_0, y_0) – will be proportional to the gradient of C at that point.* Thus the condition for a critical point (x_0, y_0) of f

* This follows from the basic calculus result that the gradient of a function is orthogonal to the locus of points where the function takes a constant value.

(where $C(x_0, y_0) = 0$) is that the gradient of f and the gradient of C are proportional at (x_0, y_0). We write

$$\nabla f + \lambda \nabla C = 0, \quad \text{where } C(x_0, y_0) = 0$$

for some λ. This requirement is meant to hold only on the curve $C = 0$, so without loss of generality we can take as the critical point condition

$$\nabla(f + \lambda C) = 0, \quad C = 0.$$

This constitutes *three* conditions on 3 unknowns; the unknowns being (x, y) and λ. The function λ is known as a *Lagrange multiplier*. In fact, if we artificially enlarge our x-y plane to a 3-d space parametrized by (x, y, λ) we can replace the above critical point condition with

$$\tilde{\nabla}(f + \lambda C) = 0,$$

where $\tilde{\nabla}$ is the gradient in (x, y, λ) space. You should prove this as an exercise.

To summarize: the critical points (x_0, y_0) of a function $f(x, y)$ constrained to a curve $C(x, y) = 0$ can be obtained by finding unconstrained critical points (x_0, y_0, λ_0) of a function in the space of variables (x, y, λ):

$$\tilde{f}(x, y, \lambda) = f(x, y) + \lambda C(x, y).$$

We can do the same thing with our variational principle. Suppose we have an action for n degrees of freedom q^i, $i = 1, 2, \ldots, n$:

$$S[q] = \int_{t_1}^{t_2} dt\, L(q(t), \dot{q}(t), t)$$

where the configuation space is subject to m constraints

$$C_\alpha(q, t) = 0, \quad \alpha = 1, 2, \ldots, m.$$

Let the solutions to the constraints be given in terms of generalized coordinates s^A, $A = 1, 2, \ldots, n - m$,

$$q^i = F^i(s, t),$$

i.e.,

$$C_\alpha(F^i(s, t), t) = 0.$$

The functions F^i determine the graph of the solution set of $C_\alpha = 0$ in the configuration space. The correct equations of motion can be obtained by substituting the solutions $q^i = F^i(s, t)$ into the Lagrangian for q^i, thus defining a Lagrangian for s^A, and computing the resulting EL equations for s^A. Using the same technology you used in your homework to study the effect of a point transformation on the EL equations (principally the chain rule), it is not hard to see that the correct equations of motion are then

$$\frac{\partial F^i}{\partial s^A}\left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)_{q=F(s,t)} = 0.$$

We note that the functions $\frac{\partial F^i}{\partial s^A}$ have the geometric meaning of (a basis of) $(n-m)$ tangent vectors to the $(n-m)$-dimensional surface $C_\alpha = 0$. Thus the equations of motion are the statement that the projections of the EL equations along the surface must vanish.

Now we introduce the Lagrange multiplier method. We consider a modified action,

$$\tilde{S}[q, \lambda] = \int_{t_1}^{t_2} dt\, \tilde{L} = S[q] + \int_{t_1}^{t_2} dt\, \lambda^\alpha(t) C_\alpha(q(t), t),$$

in which we have added m new configuration variables λ^α, $\alpha = 1, 2, \dots, m$; these are the Lagrange multipliers. The variation of the new action is

$$\delta\tilde{S} = \int_{t_1}^{t_2} dt\, \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)\delta q^i + \int_{t_1}^{t_2} dt\, \left(\delta\lambda^\alpha C_\alpha + \lambda^\alpha \frac{\partial C_\alpha}{\partial q^i}\delta q^i\right).$$

The EL equations of motion coming from $\tilde{L}$ are

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} + \lambda^\alpha \frac{\partial C_\alpha}{\partial q^i} = 0,$$

which come from the variations in q^i and also

$$C_\alpha = 0,$$

which come from variations of λ^α. We have $(n+m)$ equations for $(n+m)$ unknowns. In principle they can be solved to get the q^i and the λ^α as functions of t.

What is the meaning of these equations? Well, the constraints are there, of course. But what about the modified EL expressions? The EL equations you *would* have gotten from L now have a "force term", $\lambda^\alpha \frac{\partial C_\alpha}{\partial q^i}$. The force term is geometrically orthogonal to the surface $C_\alpha = 0$ in configuration space. This you can see from the identity (exercise)

$$0 = \frac{\partial}{\partial s^A} C_\alpha(F(s), t) = \frac{\partial C_\alpha}{\partial q^i}\frac{\partial F^i}{\partial s^A}.$$

(Recall that $\frac{\partial F^i}{\partial s^A}$ represend $n-m$ vectors tangent to the surface defined by $C_\alpha = 0$.) Thus the meaning of the EL equations coming from $\tilde{L}$ is that the EL expressions coming from L no longer have to vanish, they simply have to be orthogonal to the constraint surface since the equations of motion say that

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = -\lambda^\alpha \frac{\partial C_\alpha}{\partial q^i}.$$

One physically interprets this "force term" as the force required to keep the motion on this surface.

It is easy to verify that these modified equations, $(n+m)$ in number, are equivalent the correct $(n-m)$ equations obtained for s^A earlier. Indeed, we have the m equations of constraint. And, given this constraint, to say the EL expression coming from L is orthogonal to $C_\alpha = 0$ is the same as saying its projection tangent to the surface vanishes, *i.e.*,

$$\frac{\partial F^i}{\partial s^A}\left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)_{q=F(s,t)} = \frac{\partial F^i}{\partial s^A}\left(-\lambda^\alpha \frac{\partial C_\alpha}{\partial q^i}\right) = 0,$$

which is precisely the content of the equations for the s^A we obtained above.

Example: Plane pendulum revisited

Let us study the plane pendulum using Lagrange multipliers. We model the system as moving in a plane with coordinates (x,y) subject the constraint

$$C = x^2 + y^2 - l^2 = 0.$$

Without the constraint the Lagrangian would be simply

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy.$$

According to our general prescription for incorporating the constraint, we construct the modified Lagrangian

$$\tilde{L} = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy + \lambda(x^2 + y^2 - l^2).$$

The critical points for the action built from $\tilde{L}$, with the configuration space parametrized by (x,y,λ), should give us the critical points along the surface $C = 0$. To find the critical points we construct the EL equations as usual. We get

$$x^2 + y^2 - l^2 = 0,$$

coming from the variation of λ, and

$$2\lambda x - m\ddot{x} = 0, \quad 2\lambda y - mg - m\ddot{y} = 0,$$

coming from the variations of x and y, respectively.

Here we can see more explicitly how the Lagrange multiplier defines a force term beyond the gravitational force. This "force of constraint" represents the force of the rigid pendulum arm upon the particle and is given by

$$\vec{F}_{constraint} = 2\lambda x\hat{x} + 2\lambda y\hat{y}.$$

The typical analysis of EL equations involving Lagrange multipliers can now be nicely demonstrated. First, the three EL equations can be solved for λ (exercise)

$$\lambda = \frac{m}{2l^2}\left(x\ddot{x} + y\ddot{y} + gy\right).$$

Next, differentiation of the constraint twice reveals:

$$\ddot{C} = 0 \quad \Longrightarrow \quad x\ddot{x} + y\ddot{y} = -(\dot{x}^2 + \dot{y}^2),$$

so that the multiplier λ can be solved for in terms of the original velocity phase space variables:

$$\lambda = -\frac{m}{2l^2}(\dot{x}^2 + \dot{y}^2 - gy).$$

Substituting this result back into the EL equations for x and y we get the equations of motion for x and y with the effect of the constraint — physically, the tension in the rod — taken into account:

$$m\ddot{x} = -\frac{m}{l^2}(\dot{x}^2 + \dot{y}^2 - gy)x, \quad m\ddot{y} = -\frac{m}{l^2}(\dot{x}^2 + \dot{y}^2 - gy)y - mg.$$

Note we never had to solve the constraint! Still, as a nice exercise you can check that, after solving the constraint with $x = l\cos\phi$, $y = -l\sin\phi$, these remaining 2 equations are equivalent the familiar equation of motion for a plane pendulum, namely,

$$\ddot{\phi} = -\frac{g}{l}\sin\phi,$$

where ϕ is the angular displacement from equilibrium.

Using Lagrangian multipliers, the equations of motion for x and y tell us that the pendulum moves according to a superposition of forces consisting of (i) gravity, (ii) the force of constraint $\vec{F}_{constraint}$ needed to keep the mass moving in a circle of radius l. This latter force is supplied by the Lagrange multiplier terms in the equation of motion. Indeed, thanks to these Lagrange multiplier terms, the radial component of the net force is (exercise)

$$\frac{\vec{r}}{l}\cdot\vec{F} = -\frac{m}{l}(\dot{x}^2 + \dot{y}^2),$$

which is the centripetal force, as it should be.

To summarize: Given a dynamical system with coordinates q^i and Lagrangian L, we can impose constraints $C_\alpha(q,t) = 0$ by the following recipe.

(i) Add variables λ^α – the Lagrange multipliers – to the configuration space,

(ii) Define a Lagrangian on the augmented velocity phase space $\tilde{L} = L + \lambda^\alpha C_\alpha$,

(iii) Compute the usual EL equations from $\tilde{L}$ for the q^i and λ^α degrees of freedom.

The resulting equations will include the constraints themselves as equations of motion coming from variations of λ^α. The equations coming from the variations of the q^i will have extra terms involving the multipliers. For Newtonian systems these terms represent the forces in the system which are necessary to enforce the constraints. If desired, one can use the equations of motion, the constraints, and the time derivatives of the constraints to solve for the multipliers in terms of the velocity phase space. One can then reduce the original equations to only be built from the original degrees of freedom.

Thus the Lagrange multiplier method has distinct advantages over our previous approach in which we just solve the constraints at the beginning.. Namely, you do not have to explcitly solve the constraints in order to compute the equations of motion, and the equations of motion have additional physical information: the forces of constraint.