Additional material

§A A review of some results in calculus

I include here a brief review of some basic results in many variable calculus that will appear often during the course.

Derivatives and partial derivatives

Let me introduce the notation

$$\delta f(x) = f(x + \delta x) - f(x)$$

for the variation of a function f as we change its argument. We typically want to make δx small, and understand how δf depends on δx. The answer is that

$$\delta f = \frac{df}{dx}\delta x + \mathcal{O}((\delta x)^2)$$

where the last term is a "correction term", satisfying

$$\lim_{\delta x \to 0} \frac{\mathcal{O}((\delta x)^2)}{\delta x} = 0\,.$$

Notice that this is just a restatement of the usual definition of the derivative

$$\frac{df}{dx} = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{(x + \delta x) - x}$$

in a form which is more convenient for our applications.

This definition extends straightforwardly to functions of several variables. We define the partial derivatives of $f(x_1, \ldots, x_n)$ by

$$\frac{\partial f}{\partial x_i} = \lim_{\delta x_i \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + \delta x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{\delta x}\,.$$

Note that in the case of functions of a single variable the definitions of the partial and ordinary derivatives coincide.

We can now express the change in $f(\vec{x})$ under small changes $\delta\vec{x}$ of $\vec{x}$ as

$$\delta f(\vec{x}) = f(\vec{x} + \delta\vec{x}) - f(\vec{x}) \approx \sum_{i=1}^{n} \delta x_i \, \frac{\partial f}{\partial x_i} = \delta\vec{x} \cdot \vec{\nabla} f + \mathcal{O}(\delta\vec{x}^2)$$

where we have defined the vector of derivatives

$$(\vec{\nabla} f)_i \coloneqq \frac{\partial f(\vec{x})}{\partial x_i}\,.$$

When the variation is infinitesimal we write $\delta x_i \to dx_i$, and we have

$$df(\vec{x}) = \sum_{i=1}^{n} dx_i \, \frac{\partial f}{\partial x_i} = d\vec{x} \cdot \vec{\nabla} f\,.$$

The chain rule and commuting derivatives

Assume now that the vector $\vec{x}$ is a function of time, which we denote as $\vec{x}(t)$, and that we have a function $f(\vec{x}(t), t)$ as above (where we have included a possible explicit dependence on the time coordinate). Note that f is now implicitly a function of t via its dependence on $\vec{x}(t)$, in addition to any possible explicit dependence on t it might have. The variation of this function as t varies is given by the chain rule

$$\frac{df}{dt} = \left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}\right) + \frac{\partial f}{\partial t} = \vec{\nabla} f \cdot \frac{d\vec{x}}{dt} + \frac{\partial f}{\partial t}\,.$$

There is a version of this rule for the case of multiple variables. Say that you have a set of variables $(x_1, \ldots, x_n)$ that depend on other variables $(y_1, \ldots, y_m)$. Then:

$$\frac{\partial f}{\partial y_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial y_j}\,.$$

Another theorem that we will use later is that if we partially differentiate a function first with respect to x_i and then with respect to x_j we obtain the same as if we differentiated in the opposite order provided that the result is continuous:

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)\,.$$

This result is known as *Schwarz's theorem.* During this course all second derivatives will be continuous, so we will apply this result freely.

As it is a relatively common mistake, let me note that in general partial derivatives associates to variables belonging to different coordinates do *not* commute. Denoting by x_i the first set of coordinates, and u_i a second set, we have

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial u_j}\right) \neq \frac{\partial}{\partial u_j}\left(\frac{\partial f}{\partial x_i}\right)\,.$$

As a simple example, consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ that tells us how far a point is from the vertical axis. We choose as x_i the Cartesian coordinates (x, y), and as u_j the polar coordinates (r, θ). In Cartesian coordinates we have simply $f(x, y) = x$, while in polar coordinates we have $f(x, \theta) = r\cos\theta$. We have

$$\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial r}(1) = 0$$

while

$$\begin{aligned}\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial r}\right) &= \frac{\partial}{\partial x}(\cos\theta) = \frac{\partial}{\partial x}\left(\frac{x}{\sqrt{x^2+y^2}}\right)\\ &= \frac{1}{\sqrt{x^2+y^2}} - \frac{x^2}{(x^2+y^2)^{3/2}}\\ &\neq 0\,.\end{aligned}$$

Leibniz's rule

Assume that you have a function of x expressed in integral form:

$$f(x) = \int_{a(x)}^{b(x)} dt\, g(x,t)\,.$$

Then

$$\frac{df}{dx} = g(x,b(x))\frac{db}{dx} - g(x,a(x))\frac{da}{dx} + \int_{a(x)}^{b(x)} dt\, \frac{\partial g(x,t)}{\partial x}\,. \tag{A.0.1}$$

Notation for time derivatives

Finally, an additional piece of notation: the time coordinate t will play a special role during this course, so for convenience we introduce special notation for derivatives with respect to t. Given a function $x(t)$ we will write

$$\dot{x} \equiv \frac{dx}{dt}$$

and similarly for higher order time derivatives. For instance,

$$\ddot{x} \equiv \frac{d^2x}{dt^2}\,.$$

§A.1 Two useful lemmas for coordinate changes

Consider two sets of generalised coordinates $\{u_i\}$ and $\{q_i\}$ related by $u_i = u_i(q_1, q_2, ...q_n, t)$. Note that we allow for the change of coordinates to depend on time.[20] Such transformation is known as a *point transformation.* We also note that the $\{u_i\}$ coordinates depend on $\{q_i\}$ (and possibly t), but not on $\{\dot{q}_i\}$. This is no longer true if we take a time derivative: generically $\dot{u}_i$ will depend on $\{q_i, \dot{q}_i, t\}$. We start by proving the following two simple lemmas:

Lemma (A). *If $u_i = u_i(q_1, q_2, ...q_n, t)$ then*

$$\frac{\partial u_i}{\partial q_j} = \frac{\partial \dot{u}_i}{\partial \dot{q}_j}.$$

[20]For instance, we could have $u_i = e^t q_i$, giving a sort of "expanding" set of coordinates. Such things appear fairly naturally when one is studying cosmology, for example.

Proof. By the Chain Rule

$$\dot{u}_i = \sum_{k=1}^{n} \frac{\partial u_i}{\partial q_k} \dot{q}_k + \frac{\partial u_i}{\partial t}$$

Further differentiating with respect to $\dot{q}_j$ just picks out the coefficient of $\dot{q}_j$ (since u_i does not depend on $\dot{q}_i$ its derivative $\partial u_i / \partial q_k$ does not either) giving the advertised result

$$\frac{\partial}{\partial \dot{q}_j}(\dot{u}_i) = \frac{\partial}{\partial \dot{q}_j}\left(\sum_{k=1}^{n} \frac{\partial u_i}{\partial q_k}\dot{q}_k + \frac{\partial u_i}{\partial t}\right) = \frac{\partial u_i}{\partial q_j}. \qquad \square$$

Lemma (B)**.** *If* $u_i = u_i(q_1, q_2, ...q_n, t)$ *then*

$$\frac{\partial \dot{u}_i}{\partial q_j} = \frac{d}{dt}\left(\frac{\partial u_i}{\partial q_j}\right).$$

Proof. We again use the Chain Rule, and the fact that partial derivatives on the same set of coordinates commute (if the result is continuous):

$$\frac{d}{dt}\left(\frac{\partial u_i}{\partial q_j}\right) = \sum_{k=1}^{n} \frac{\partial^2 u_i}{\partial q_k \partial q_j}\dot{q}_k + \frac{\partial^2 u_i}{\partial t \partial q_j} = \frac{\partial}{\partial q_j}\left(\sum_{k=1}^{n} \frac{\partial u_i}{\partial q_k}\dot{q}_k + \frac{\partial u_i}{\partial t}\right) = \frac{\partial}{\partial q_j}(\dot{u}_i). \qquad \square$$

Example: invariance of the Euler-Lagrange equations under coordinate changes

As an example of how the theorems above are useful, let us prove explicitly that the choice of generalized coordinates does not affect the form of the Euler-Lagrange equations.

Theorem. *Assume that we have two sets of generalized coordinates* $\{u_1, \ldots, u_n\}$ *and* $\{q_1, \ldots, q_n\}$ *related by an invertible change of coordinates* $u_i = u_i(q_1, \ldots, q_n, t)$. *Then the Euler-Lagrange equations*

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = 0 \quad \forall i \in \{1, \ldots, n\}$$

are equivalent to

$$\frac{\partial L}{\partial u_k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}_k}\right) = 0 \quad \forall k \in \{1, \ldots, n\}$$

Proof. We will prove the result by repeated application of the Chain Rule. For the first term in the Euler-Lagrange equations we get

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{d}{dt}\left(\sum_{k=1}^{n} \frac{\partial L}{\partial u_k}\underbrace{\frac{\partial u_k}{\partial \dot{q}_i}}_{=0} + \sum_{k=1}^{n} \frac{\partial L}{\partial \dot{u}_k}\frac{\partial \dot{u}_k}{\partial \dot{q}_i} + \frac{\partial L}{\partial t}\underbrace{\frac{\partial t}{\partial q_i}}_{=0}\right)$$

which using Lemma (A) becomes

$$= \frac{d}{dt}\left(\sum_{k=1}^{n} \frac{\partial L}{\partial \dot{u}_k}\frac{\partial u_k}{\partial q_i}\right)$$

$$= \sum_{k=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}_k}\right)\right]\frac{\partial u_k}{\partial q_i} + \sum_{k=1}^{n}\left(\frac{\partial L}{\partial \dot{u}_k}\right)\frac{d}{dt}\left(\frac{\partial u_k}{\partial q_i}\right)$$

which is now, using Lemma (B)

$$= \sum_{k=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}_k}\right)\right]\frac{\partial u_k}{\partial q_i} + \sum_{k=1}^{n}\frac{\partial L}{\partial \dot{u}_k}\frac{\partial \dot{u}_k}{\partial q_i}\,.$$

The second term in the Euler-Lagrange equations is easier. Again using the Chain Rule:

$$\frac{\partial L}{\partial q_i} = \sum_{k=1}^{n}\frac{\partial L}{\partial u_k}\frac{\partial u_k}{\partial q_i} + \sum_{k=1}^{n}\frac{\partial L}{\partial \dot{u}_k}\frac{\partial \dot{u}_k}{\partial q_i} + \frac{\partial L}{\partial t}\underbrace{\frac{\partial t}{\partial q_i}}_{=0}\,.$$

Taking the difference of both equations we get

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = \sum_{k=1}^{n}\left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}_k}\right) - \frac{\partial L}{\partial u_k}\right)\frac{\partial u_k}{\partial q_i}\,.$$

We are almost there. In order to exhibit the rest of the argument most clearly, we will switch to matrix notation. Denote the matrix associated to the change of variables by

$$\mathsf{J}_{ik} \coloneqq \frac{\partial u_k}{\partial q_i}\,.$$

This matrix (known as the "Jacobian matrix") is invertible, since by assumption the change of coordinates is invertible. Denote the vector of Euler-Lagrange equations on the q coordinates by

$$\mathsf{E}_i^{(q)} = \frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right)$$

and similarly for the u coordinates

$$\mathsf{E}_k^{(u)} = \frac{\partial L}{\partial u_k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{u}_k}\right)\,.$$

Using these definitions we can rewrite the Euler-Lagrange equations as the vector equations $\vec{\mathsf{E}}^{(q)} = 0$ and $\vec{\mathsf{E}}^{(u)} = 0$, and we have just shown that

$$\vec{\mathsf{E}}^{(q)} = \mathsf{J}\,\vec{\mathsf{E}}^{(u)}$$

with J invertible, so $\vec{\mathsf{E}}^{(q)} = 0$ iff $\vec{\mathsf{E}}^{(u)} = 0$, which is what we wanted to show. □