AN INTRODUCTION TO GLOBAL ANALYSIS
AND THE CALCULUS OF VARIATIONS

M. A. GUEST
Preface

These notes form the outline of a course given at the University of Rochester in the Fall Semester of 1988. The intention of the course was to give a brief and gentle introduction to smooth manifolds, in the (natural, but non-trivial) context of the calculus of variations. On the one hand, manifolds are thus seen to arise naturally from concrete problems. On the other hand, the ancient subject of the calculus of variations is seen in a new and refreshing light. Of course, the calculus of variations did have a great influence on the development of global analysis, but textbooks on smooth manifolds and global analysis have tended to suppress this somewhat, and textbooks on the calculus of variations have rarely sympathized with the global point of view. With the recent renewal of interest in these areas (stemming from developments in mathematical field theories, in physics), it is perhaps appropriate again to acknowledge openly the close connections between the two subjects.

Lectures 1 to 5 present some examples in the classical calculus of variations, in as simple-minded a way as possible. The only prerequisites for reading these is some knowledge of calculus (especially the chain rule!). Lectures 6 to 10 give a brief treatment of smooth manifolds. (Very few details are given here as many modern texts are available.) Finally, some attempt is made to draw things together in lectures 11 to 14. While these notes barely scratch the surface of the "modern" calculus of variations, they should prepare the reader for more specialized graduate courses in this direction.

December 1988
Lecture 1: Functionals and the Euler-Lagrange equation

A fundamental problem in calculus is to find the maxima and minima of a real-valued function

\[ f : U \to \mathbb{R} \]

on a subset \( U \) of Euclidean space \( \mathbb{R}^n \). The calculus of variations is based on the more general problem of finding the maxima and minima of a real-valued function

\[ f : \mathcal{U} \to \mathbb{R} \]

on a set of functions \( \mathcal{U} \).

**Definition:** A function \( y \) is said to be a critical point of \( f \) if and only if \( (\frac{d}{dt}) f(y_t) \big|_{t=0} = 0 \) for all smooth families \( \{ y_t \} \) of functions in \( \mathcal{U} \) with \( y_0 = y \). (The precise meaning of "smooth" will be explained in lecture 3.)

It is clear that critical points should be useful in searching for maxima and minima of \( f \). The precise relation between critical points and maxima or minima obviously depends on the nature of the set \( \mathcal{U} \), so we shall postpone such matters for the moment.

In classical terminology, \( f \) is called a *functional*, and a critical point is called an *extremal* of \( f \). A *variational problem* is a problem in which the extremals (or maxima and minima) of a functional are sought.

**Problem A.** What curve connecting the points \((0,1)\) and \((1,0)\) gives the least "time of descent" for a particle released at \((0,1)\) and allowed to slide down to \((1,0)\) ?

![Diagram](image)

This is the famous "brachistochrone problem" proposed by Johannes Bernoulli in 1696,
and solved by him, his brother Jacob, and Newton (amongst others) shortly afterwards. (Newton had by this time relinquished his academic post in order to become Master of the Royal Mint.) The answer is that the curve must be an arc of a cycloid.

The cycloid is also the "tautochrone"; it has the property that the time of descent to (1,0) is actually independent of the point on the curve where the particle is released!

Problem B. Which curve connecting the points (-1,a) and (1,a) gives the least surface area when revolved about the x-axis? This is a very simple problem in the theory of "minimal surfaces". The answer is the catenary if a is greater than approximately 1.51; otherwise there is no solution (which is $C^1$). The catenary arises as the solution to another variational problem, namely to find the position of a hanging (heavy) chain such that the potential energy of the chain is minimized.

Problem C. Which curve (of the type considered above in problem B) gives the least volume of revolution? Obviously there is no answer to this; the volume can be made arbitrarily close to zero.

To explain the solutions of a variety of variational problems of this nature, we shall consider $\mathcal{U}_*$ and $f$ of the following form:

$$\mathcal{U}_* = \{ y \in C^1[a,b] \mid y(a) = A, y(b) = B \}$$

$$f(y) = \int_a^b F(x,y,y') \, dx$$

where $F$ is a smooth (at least, $C^2$) function of three variables, and where we use the usual illogical calculus convention of writing $y = y(x)$. Thus our set of functions is the set of $C^1$ curves connecting $(a,A)$ and $(b,B)$, and our functional assigns to each curve the integral of some quantity concocted using that curve.
Theorem: (Euler-Lagrange equation) A curve $y \in \mathcal{U}$ is a critical point of $f$ if and only if $y$ satisfies the second order ordinary differential equation

$$ F_y = \frac{d}{dx} (F_{y'}) . $$

Proof. In applying the condition for a critical point, it suffices to consider variations of the form $y_t = y + tu$, where $h \in C^1[a,b]$ vanishes at $a$ and $b$. (This is because $\mathcal{U}$ is an affine space, as will be explained in more detail in lecture 3.) The condition says that

$$ \frac{d}{dt} \int_a^b F(x,y+tu,y'+tu') \, dx \bigg|_{t=0} = 0 $$

i.e.

$$ \int_a^b F_y(x,y,y')u + F_{y'}(x,y,y')u' \, dx = 0 $$

for all such $u$. The proof is finished by integrating the second term of the integrand by parts, and observing that if $w$ is a function for which $\int_a^b wu \, dx = 0$ for all such $u$, then $w = 0$. QED

The general solution to the Euler-Lagrange equation contains in general two arbitrary parameters, which, if we are lucky, may be adjusted to obtain solutions satisfying the boundary conditions $y(a) = A, y(b) = B$. If we are especially lucky, the desired maxima or minima will be amongst these solutions. Of course we are not always lucky; but in many "good" problems we are.

Solution to problem A. Use the following diagram, which shows the path of the particle from time $0$ to time $t$.

\[ \text{Diagram showing the path of the particle from time } 0 \text{ to time } t. \]
The gain in kinetic energy by time $t$ is equal to the loss in potential energy, so one has

$$(1/2)mv^2 = mgy \quad (\text{where } m \text{ is the mass, } g \text{ is acceleration due to gravity, and } v = ds/dt \text{ is the speed after time } t).$$

Hence $v = (2gy)^{1/2}$, and the time of descent is given by

$$T(y) = \int_0^1 dt/ds \frac{ds}{dx} \quad dx = (2g)^{-1/2} \int_0^1 y^{-1/2}(1 + y^2)^{1/2} \quad dx.$$  

The Euler-Lagrange equation can now be written down, but it is a dreadful mess. A shortcut is provided by the following lemma.

Lemma. Suppose that the $F(u,v,w)$ is independent of $u$. Then the Euler-Lagrange equation implies the equation

$$y' F_{y'} - F = C$$

for some constant $C$.

Proof. The Euler-Lagrange equation, on applying the chain rule, may be written $F_y = F_{yy}y' + F_{yy}y''$ (using the assumption that $F_x = 0$). On multiplying by $y'$, one has (miraculously) $d/dx (y'F_{y'} - F) = 0$. QED

Applying the lemma, one quickly arrives at the equation $y(1 + y^2) = -1/C^2 = D \quad (\text{say})$. Now let $dy/dx = \tan \theta$. Then $y = D \cos^2 \theta = D(1 + \cos 2\theta)/2$. Hence $dy/d\theta = -D \sin 2\theta$, and $dx/d\theta = dx/dy \ dy/d\theta = -2D \cos 2\theta = -D(1 + \cos 2\theta)$. So $x = -D(2\theta + \sin 2\theta)/2 + E$ for some constant $E$. These equations for $x$ and $y$ are parametric equations for a cycloid.

Solution to problem B. The surface area of revolution is $2 \int_0^1 y(1+y^2)^{1/2} \quad dx$. We are assuming that $y(x)$ is nonnegative here, but let's proceed under the assumption that this slight modification to the space $\mathbb{U}$ is not important. The "integrated" Euler-Lagrange equation is $-y(1+y^2)^{-1/2} = C$. Hence $y = C \cosh (x/C + D)$ for constants $C, D$. To satisfy the boundary conditions we need $D = 0$ and $a = C \cosh (1/C)$. This is possible providing $a$ is greater than approximately 1.51. (For the hanging chain problem, see later.)

Solution to problem C. The volume of revolution is $\pi \int_{-1}^1 y^2 \quad dx$, which leads to the
equation $-y^2 = C$. The boundary conditions are never satisfied (unless $a = 0$).
Lecture 2: Generalizations and further examples

The special functional $f(y) = \int_a^b F(x,y,y') \, dx$ of lecture 1 can be generalized by allowing more "independent" variables, more "dependent" variables, and more derivatives. The method is an obvious generalization of the previous method.

**First generalization.** Take $\mathcal{U}$ and $f$ of the following form:

$$\mathcal{U} = \{ (y_1,\ldots,y_n) \mid y_i \in C^1[a,b], \ y_i(a) = A_i, \ y_i(b) = B_i \}$$

$$f(y_1,\ldots,y_n) = \int_a^b F(x,y_1,\ldots,y_n,y_1',\ldots,y_n') \, dx$$

where $F$ is a smooth function of $2n+1$ variables. As in lecture 1 we obtain an Euler-Lagrange equation by considering a variation of the form $Y_t = Y + t\xi$, where $Y = (y_1,\ldots,y_n)$. We obtain in fact the following system of equations:

$$F_{y_1} = d/dx (F_{y_1'}), \ldots, F_{y_n} = d/dx (F_{y_n'}) \, .$$

As in the lemma of lecture 1, if $F(u,v,w,\ldots)$ is independent of $u$, then a "first integral" of this system may be obtained, namely

$$y_1' F_{y_1'} + \ldots + y_n' F_{y_n'} - F = C \, .$$

If $n > 1$, however, it may still be quite difficult to solve the system.

**Example 1.** The functional of lecture 1 may be put in "parametric form" by introducing a parameter $t$ ($t \in [0,1]$) for the curve $y = y(x)$, so that $x = x(t)$, $y = y(t)$. Thus the functional is $g(x,y) = \int_0^1 F(x,y,y'/x) \, dx$. As $t$ does not appear explicitly in the integrand, a first integral may be written down immediately. (This does not necessarily represent a simplification of the original problem, however; it merely accounts for the redundancy from introducing a second dependent variable.)

**Example 2.** What are the geodesics in the cylinder $x^2 + y^2 = r^2$ in $\mathbb{R}^3$ (i.e. the extremals for the length functional on curves lying on the cylinder)? Using coordinates $z, \theta$ for the cylinder, where $x = r \cos \theta$, $y = r \sin \theta$, we may represent any curve parametrically as $c(t) = (\theta(t), z(t))$, $t \in [0,1]$. The length functional is given by $L(c)$
The system of Euler-Lagrange equations gives easily the conclusion that \( \frac{dz}{d\theta} \) (or \( \frac{d\theta}{dz} \)) must be constant.

Second generalization. Take \( \mathcal{U} \) and \( f \) of the following form:

\[
\mathcal{U} = \{ y \in C^n[a,b] \mid y^{(i)}(a) = A_i, \; y^{(i)}(b) = B_i, \; i=0,\ldots,n-1 \}
\]

\[f(y) = \int_a^b F(x,y,y',\ldots,y^{(n)}) \, dx\]

where \( F \) is a smooth function of \( n+1 \) variables. The Euler-Lagrange equation (obtained by taking a variation of the form \( y_t = y + \theta h \), and performing \( n \) integrations by parts in the final step) is:

\[F_y - \frac{d}{dx}(F_y') + (\frac{d}{dx})^2(F_y'(2)) - \ldots - (-1)^n(\frac{d}{dx})^n(F_y(n)) = 0.\]

Third generalization. Take \( \mathcal{U} \) and \( f \) of the following form:

\[
\mathcal{U} = \{ z \in C^1(D) \mid z|_c = u \}
\]

where \( D \) is a region of the xy-plane whose boundary is a simple smooth closed curve \( c \), and where \( u \in C^1(c) \) is fixed, and

\[f(z) = \int_D F(x,y,z,z_x,z_y) \, dx \, dy\]

where \( F \) is a smooth function of five variables. Consider a variation of \( z \) of the form \( z_t = z + \theta h \), where \( h \in C^1(D) \) and \( h|_c = 0 \). The condition for \( z \) to be a critical point of \( f \) is then

\[
\int_D F_z h + F_{z_x} h_x + F_{z_y} h_y \, dx \, dy = 0
\]

for all such \( h \). This may be rewritten as

\[
\int_D F_z h + \frac{\partial}{\partial x}(hF_{z_x}) - h\frac{\partial}{\partial x}(F_z) + \frac{\partial}{\partial y}(hF_{z_y}) - h\frac{\partial}{\partial y}(F_z) \, dx \, dy = 0
\]

where the partial derivatives \( \frac{\partial}{\partial x} \) and \( \frac{\partial}{\partial y} \) operate on functions of the two variables \( x \) and \( y \) (and not on functions of the five variables \( x, y, z, z_x, z_y \)). The second and fourth
terms together are a "divergence", so Green's theorem may be applied to their integral, which is easily seen to be zero (since \( h \) is zero on the curve \( c \)). Hence the usual argument gives the following Euler-Lagrange equation:

\[
F_z - \partial_\theta x(F_{z\theta}) - \partial_\phi y(F_{z\phi}) = 0.
\]

**Example 3.** For \( \mathcal{U} \) as above, let \( f \) be the "energy" functional

\[
E(z) = \int_D (z_x)^2 + (z_y)^2 \, dxdy.
\]

The Euler-Lagrange equation is very simple in this case:

\[
z_{xx} + z_{yy} = 0.
\]

This is called Laplace's equation, and its solutions (the critical points of the energy functional) are called harmonic functions.

For comments on "first integrals" of the variational problems of the second and third generalizations, see lecture 12.
Lecture 3: Some remarks

1. The relation between critical points and local extrema.

The definition of a local extremum depends on choosing a topology for $\mathcal{U}$.

**Definition:** A function $y$ is said to be a local maximum for $f : \mathcal{U} \to \mathbb{R}$ (with respect to a choice of topology on $\mathcal{U}$) if and only if there exists a neighbourhood $U$ of $y$ in such that $f(z) \leq f(y)$ for all $z \in U$. A local minimum is defined in the obvious way, and local maxima or minima are called local extrema.

In this section we shall always take $\mathcal{U}$ and $f : \mathcal{U} \to \mathbb{R}$ as in lecture 1. This will enable us to use a familiar topology on $\mathcal{U}$. The conclusions will generalize easily to the examples of lecture 2, however. We begin by listing various properties of $\mathcal{U}$.

(a) The space $\mathcal{U}$ is an affine subspace of the vector space $C^1[a,b]$. (In other words, it is a translate of a vector subspace.)

(b) The space $\mathcal{U}$ acquires a metric (and hence a topology) from the metric $d$ on $C^1[a,b]$ defined by $d(y,z) = \|y-z\|$, where $y = \sup\{y(x)\} + \sup\{|y'(x)|\}$. (It is well known that the topology on $\mathbb{R}^n$ obtained from a norm on $\mathbb{R}^n$ is independent of which norm is used; this is definitely not true for infinite dimensional vector spaces such as $C^1[a,b]$ !)

(c) The concept of "smooth variation", introduced without explanation in lecture 1, may be defined with respect to the above choice of norm. We consider a variation $\{y_t\}$ of a function $y \in \mathcal{U}$ to be a map $v : (-\epsilon, \epsilon) \to \mathcal{U}$, with $v(0) = y_0 = y$. We say that the variation is continuous if the function $v$ is continuous, and that the variation is differentiable if the function $v$ is differentiable (in the sense that, for each $c \in (-\epsilon, \epsilon)$, $v(t) = v(c) + (t-c)C + O((t-c)^2)$ for some $C \in C^1[a,b]$ - we then write $C = v'(c)$, as in ordinary calculus). A smooth variation means a variation that is infinitely differentiable.

(d) It is convenient to introduce a concept of differentiability for functions $k : \mathcal{U} \to \mathbb{R}$:

**Provisional definition:** A function $k : \mathcal{U} \to \mathbb{R}$ is smooth if and only if the functions $k \cdot v : (-\epsilon, \epsilon) \to \mathbb{R}$ are smooth (in the usual sense) for all smooth functions $v$. 


The smoothness of $F$ guarantees that $f(y) = \int_a^b F(x,y,y') \, dx$ defines a smooth function $f : \mathcal{U} \to \mathbb{R}$, in the sense of this definition.

This concludes our review of the properties of $\mathcal{U}$.

**Proposition:** Let $\mathcal{U}$ and $f$ be as in lecture 1. If $y$ is a local extremum of $f$, then $y$ is a critical point of $f$.

**Proof.** Assume that $y$ is a local extremum of $f$. Let $v$ be a smooth variation of $y$. Since $v$ is smooth, it is continuous (by the usual argument in calculus), so there exists some open interval $I$ with $0 \in I \subseteq (-\varepsilon, \varepsilon)$ and with $t = 0$ a local extremum of $f \cdot v[I]$. But $f \cdot v[I]$ is smooth (in the usual sense), so we must have $d/dt(f(v(t)))|_{t=0} = 0$. Hence $y$ is a critical point of $f$. QED

2. **Types of variation.**

In the classical literature, a continuous variation is referred to as a *weak* variation. A variation which is only continuous with respect to the topology given by the norm $\| y \| = \sup x \in I y(x)$ is referred to as a *strong* variation. We shall only be concerned with smooth variations (these are automatically weak). In lectures 1 and 2, we used an even more restricted kind of variation, namely a "linear" variation

$$v(t) = y + \theta$$

(determined by a fixed $h \in C^1[a,b]$ with $h(a) = h(b) = 0$). This is justified by the following result.

**Proposition:** Let $\mathcal{U}$ and $f$ be as in lecture 1. A function $y \in \mathcal{U}$ is a critical point of $f$ if and only if $d/dt(y + \theta h)|_{t=0} = 0$ for all $h \in C^1[a,b]$ with $h(a) = h(b) = 0$.

**Proof.** The variation $v(t) = y + \theta h$ is certainly smooth (with $v'(c) = h$, $v(0) = 0$ for $i > 1$), so if $y$ is a critical point, then the condition in the proposition is satisfied. Conversely, given the condition, we must show that $d/dt(f(v(t)))|_{t=0} = 0$ for any smooth variation $v$.

To do this, write $v(t) = y + \theta v'(0) + O(\theta^2) = y + \theta (h + u)$, where $h = v'(0)$ and $u = O(\theta)$. Then $d/dt(f(v(t)))|_{t=0}$

$$= \lim_{t \to 0} \frac{1}{t} \int_a^b F(x, y + \theta (h + u), y' + \theta (h' + u')) - F(x, y, y') \, dx$$
\[ \lim_{t \to 0} \int_a^b (h+u)F_y(x,y+t[h+u],y'+t[h'+u']) + (h'+u')F_y(x,y+t[h+u],y'+t[h'+u']) \, dx \]

\[ = \int_a^b hF_y(x,y,y') + h'F(x,y,y') \, dx . \]

\[ = \frac{d}{dt}(f(y+th))|_{t=0} \]

and this is zero by hypothesis. QED

3. Constant functionals.

If \( F(x,y,y') = \frac{d}{dx} G(x,y) = G_x(x,y) + y'G_y(x,y) \), then the functional \( f(y) = \int_a^b F(x,y,y') \, dx \) obviously takes the constant value \( G(a,A) + G(b,B) \) for every \( y \). The Euler-Lagrange equation reduces to \( 0 = 0 \). This rather trivial observation can be important in more complicated variational problems, as it may be difficult to tell whether a given integrand is a "total derivative". The constancy of the functional often indicates that it represents some fundamental geometrical, topological or physical invariant.

4. Parametric integrals.

If we introduce a parameter \( t \) and write \( x = x(t), y = y(t) \), for \( t \in [0,1] \), then we may write the functional \( f(y) = \int_a^b F(x,y,y') \, dx \) in "parametric form" (as in example 1 of lecture 2). This gives a slightly different variational problem of the form \( g(x,y) = \int_0^1 G(t,x,y,\dot{x},\dot{y}) \, dt \). (Not every solution of the new problem defines a function \( y = y(x) \).)

However, it is intuitively clear that this new problem is left unchanged by introducing a further change of parameter by means of a formula \( t = t(s) \). This is true, and is a special case of the following result.

**Proposition:** Assume that the smooth function \( G \) satisfies the identity \( G(u,v,\lambda w,\lambda x) = \lambda G(u,v,w,x) \) for all \( \lambda > 0 \). Then the functional defined by \( g(x,y) = \int_0^1 G(x,y,\dot{x},\dot{y}) \, dt \) is independent of any smooth change of parameter.

**Proof.** By a smooth change of parameter we mean a smooth map \( m:[0,1] \to [p,q] \) such that \( m^{-1} \) exists and is smooth, and such that \( m(0) = p, m(1) = q \). We write \( s = s(t) = m(t) \) (and \( t = t(s) = m^{-1}(s) \)). Changing variables from \( t \) to \( s \), we have:

\[ \int_0^1 G(x,y,\dot{x},\dot{y}) \, dt = \]
\[ \int_p^q G(x,y,\dot{s}(dx/ds),\dot{s}(dy/ds))(1/\dot{s}) \, ds \]
\[ = \int_p^q G(x,y,dx/ds,dy/ds) \, ds \]
as required. \textbf{QED}

In the classical literature, functionals which are unchanged by a re-parametrization of the integrand are called \textit{parametric integrals}. This "invariance" property expresses a certain "redundancy" in the variational problem, which manifests itself in the "first integral" \( \dot{x}G_x + \dot{y}G_y - G = C \) of the system of Euler-Lagrange equations (as we pointed out in lecture 1). (Of course this is not usually equivalent to the system of Euler-Lagrange equations, rather, it is an identity satisfied by all solutions of the system.) Note that in the situation of the proposition, this equation (with \( C = 0 \)) is an immediate consequence of Euler's identity for homogeneous functions.
Lecture 4: Variational problems with constraints

The prototype for this kind of problem in calculus is to find the critical points of a smooth function \( f : U \to \mathbb{R} \) (where \( U \) is an open subset of \( \mathbb{R}^n \)) subject to the "constraint" \( g(x_1, \ldots, x_n) = c \) (where \( g : U \to \mathbb{R} \) is smooth). In other words, the critical points of \( f|_D \) are sought, where \( D = \{ x = (x_1, \ldots, x_n) \in U \mid g(x) = c \} \). The statement that "\( x \) is a critical point of \( f|_D \)" needs some clarification: a reasonable definition is that \( \frac{d}{dt}(f(g(t))) \bigg|_{t=0} = 0 \) for any smooth curve \( t \mapsto y(t) \) such that (1) \( y(0) = x \), (2) \( g(y(t)) = c \), and (3) \( \dot{y}(0) = 0 \). This condition may be put in the following more useful form. Let \( T_x = \{ y \in \mathbb{R}^n \mid y = \dot{y}(0) \text{ for } y \text{ as above} \} \cup \{0\} \). Then \( \nabla g(x) \cdot y = 0 \) for any \( y \in T_x \) (by applying the chain rule to (2)). If \( \nabla g(x) \neq 0 \), then the implicit function theorem shows that \( \dim T_x \) must be at least \( n-1 \), so we obtain a characterization of \( T_x \) as \( \{ y \in \mathbb{R}^n \mid y \cdot \nabla g(x) = 0 \} \).

The formula \( \frac{d}{dt}(f(g(t))) \bigg|_{t=0} = 0 \) gives \( \nabla f(x) \cdot y = 0 \) for all \( y \in T_x \), hence

\[
\nabla f(x) = \lambda \nabla g(x)
\]

for some constant \( \lambda \). Conversely, any \( x \in U \) for which this formula holds is obviously a critical point. The formula gives \( n \) equations, which, together with the equation \( g(x) = c \), are (in an ideal world) precisely what is needed to determine the \( n+1 \) quantities \( x_1, \ldots, x_n, \lambda \). This method of finding the critical points of \( f|_D \) is known as the method of Lagrange multipliers.

Essentially the same method generalizes to the calculus of variations. Imposing a "constraint" on a functional \( f : \mathcal{U} \to \mathbb{R} \) means replacing the domain \( \mathcal{U} \) by some subset \( \mathcal{V} \).

Of course, the nature of the subset \( \mathcal{V} \) is of crucial importance (and a source of great trouble). In the classical theory, two kinds of constraints are considered: those which involve an integral of the dependent variable(s), and those which do not.

Integral constraints.

Let \( \mathcal{U} \) and \( f \) be as in lecture 1. Let \( g(y) = \int_a^b G(x,y,y') \, dx \) where \( G \) is a smooth function of three variables. Define \( \mathcal{V} \) by

\[
\mathcal{V} = \{ y \in \mathcal{U} \mid g(y) = c \}
\]

where \( c \) is a constant. By a critical point of \( f|_\mathcal{V} \) we mean a function \( y \in \mathcal{V} \) such that \( \frac{d}{dt} f(y(t)) \bigg|_{t=0} = 0 \) for all smooth variations \( \{ y_t \} \) of \( y \) which have the property that \( g(y_t) \)
= c for all \( t \) near 0. (Warning: it is not immediately clear that such "constrained variations" exist. In fact they do, but we need some extension of the implicit function theorem to prove this. We shall ignore this point for the moment.)

**Theorem 1:** Assume that the function \( G_y(x,y,y') - \frac{d}{dx} G_y(x,y,y') \) when evaluated at a certain function \( y \), is not the zero function (in \( C^1[a,b] \)). Then \( y \) is a critical point of \( f \) (as defined above) if and only if there exists a constant \( \lambda \) such that

\[
F_y - \frac{d}{dx}(F_y') = \lambda (G_y - \frac{d}{dx}(G_y)) \quad \text{(at } (x,y,y') \text{).}
\]

**Proof.** As usual (see lecture 3, remark 2), it suffices to take \( y_t \) of the form \( y_t = y + th \).

The usual argument shows that \( \int_a^b (F_y - \frac{d}{dx}(F_y))h = 0 \) for all \( h \in C^1[a,b] \) such that \( h(a) = h(b) = 0 \) and such that \( \int_a^b G(x,y+th,y'+th') = c \) for all \( t \) near to 0. The condition in italics is equivalent to \( \int_a^b (G_y - \frac{d}{dx}(G_y))h = 0 \). The argument is completed by the following lemma, whose proof we omit.

**Lemma:** If \( p, q \), are smooth functions satisfying \( \int_a^b ph = 0 \) for all \( h \in C^1[a,b] \) such that \( h(a) = h(b) = 0 \) and such that \( \int_a^b qh = 0 \), and if \( q \) is not the zero function, then there exists a constant \( \lambda \) such that \( p = \lambda q \).

**QED**

Roughly speaking, the expression \( F_y - \frac{d}{dx}(F_y) \) plays the role of the "gradient" of \( f \). If the inner product \( \langle p, q \rangle = \int_a^b pq \) is used on \( C^1[a,b] \), then the argument of the theorem becomes analogous to that of the Lagrange multiplier rule above. In particular, the role of the "tangent space" \( T_x \) is played by \( \{ h \in C^1[a,b] \mid \langle G_y - \frac{d}{dx}(G_y), h \rangle = 0 \} \).

The theorem generalizes easily to the more general variational problems of lecture 2. The equation obtained is called the Euler-Lagrange equation (for the variational problem).

**Example 1:** Consider the problem of finding the curve \( y = y(x) \) of length \( L \) which maximizes the area between the curve, the x-axis, and the lines \( x = a, x = b \). The functional here is \( f(y) = \int_a^b y \, dx \), but we have the constraint \( \int_a^b (1 + y'^2)^{1/2} \, dx = L \). The Euler-Lagrange equation (from the theorem) is

\[
1 = \lambda(0 - \frac{d}{dx}(y'(1+y'^2)^{-1/2}))
\]
which reduces to

\[-1/\lambda = y''/\left(1 + y'^2\right)^{3/2} \]

Hence the candidates for maxima are the curves of constant non-zero curvature.

**Example 2:** Consider the "isoperimetric problem" (perhaps the oldest problem in the calculus of variations, originating from long before the discovery of calculus), of finding the curve in the plane of length \( L \) which encloses the largest area. In terms of polar coordinates, we have to maximize \( f(r) = 1/2 \int_0^2 r^2 \, d\theta \), subject to the constraint \( \int_0^2 r \, d\theta = L \) (we assume \( r = r(\theta) > 0 \)). The Euler-Lagrange equation is

\[
2r - \frac{d}{d\theta}(0) = \lambda (1 - 0)
\]

i.e. \( 2r = \lambda \). The only solution of this, which also satisfies the constraint, is the circle of length \( L \).

**Non-integral constraints.**

Let \( \mathcal{U} \) and \( f \) be as in lecture 1. A constraint of the form \( G(x,y,y') = c \) is simply a first order differential equation for \( y \), and so the imposition of such a constraint reduces \( \mathcal{U} \) to a very small subspace indeed (a subspace with only "one degree of freedom" in general). This is of little interest.

A less trivial example is provided by the functional

\[
f(y_1,y_2) = \int_a^b F(x,y_1,y_2,y'_1,y'_2) \, dx
\]

on the space \( \mathcal{U} \) of pairs of \( C^1 \) functions taking fixed values at \( a \) and \( b \). Let us impose a constraint of the form \( S(y_1,y_2) = 0 \), where \( S \) is a smooth function of two variables. In other words, we take

\[\nabla = \{ (y_1,y_2) \in \mathcal{U} | S(y_1,y_2) = 0 \}\]

Of course, this cuts down \( \mathcal{U} \) quite drastically, but the resulting space \( \nabla \) is still "infinite dimensional".
Theorem 2: Assume that $(S_{y_1}, S_{y_2})$ is not the zero function (when evaluated at a function $y$). Then $y$ is a critical point of $f \mid_{\mathcal{V}}$ if and only if there exists a function $x \mapsto \lambda(x)$ such that

$$F_{y_1} - \frac{d}{dx}(F_{y_1}) = \lambda S_{y_1}, \quad F_{y_2} - \frac{d}{dx}(F_{y_2}) = \lambda S_{y_2} \quad \text{(at } (y_1, y_2)).$$

Proof. As in the proof of theorem 1, we proceed under the (as yet unjustified) assumption that linear variations of the form $(y_1)_t = y_1 + th, \ (y_2)_t = y_2 + tk$ exist with the property that $S(y_1 + th, y_2 + tk) = 0$ for all $t$ near 0. The condition for a critical point is that $\int_a^b (F_{y_1} - \frac{d}{dx}(F_{y_1}))h + (F_{y_2} - \frac{d}{dx}(F_{y_2}))k = 0$ for all those $h, \ k$ satisfying the condition $hS_{y_1} + kS_{y_2} = 0$. The desired equation follows on eliminating $h$ or $k$.

QED

Example 3. Let $\mathcal{V}$ consist of (smooth) functions $y = (y_0, \ldots, y_m) : \mathbb{R}^n \to \mathbb{R}^{m+1}$ subject to the condition $y_0^2 + \ldots + y_m^2 = 1$. Let $f$ be the energy functional (c.f. example 3 of lecture 2) defined by $f(y) = \int_B \sum (\partial y_j / \partial x_j)^2 \ dv$ where $B$ is the unit ball in $\mathbb{R}^n$. The Euler-Lagrange equation is the system

$$-\partial / \partial x_1 \left( 2 \partial y_0 / \partial x_1 \right) - \ldots - \partial / \partial x_n \left( 2 \partial y_0 / \partial x_n \right) = \lambda \partial y_0$$

$$-\partial / \partial x_1 \left( 2 \partial y_m / \partial x_1 \right) - \ldots - \partial / \partial x_n \left( 2 \partial y_m / \partial x_n \right) = \lambda \partial y_m.$$

This reduces to the condition $\Delta y = (\text{constant}) \times y$.
Lecture 5: Classical mechanics

Classical mechanics deals with the behaviour of "systems" (e.g. particles, rigid bodies, vibrating strings) under the influence of "forces" (e.g. gravity, electromagnetism). In general, a system is specified by giving certain coordinates \( q_1, ..., q_n \), and these vary with time as the force takes effect. The object is to predict the behaviour of the system.

*Hamilton's Principle* is a general law which facilitates this prediction, in situations where the system possesses a "kinetic energy function" \( T(t,q,\dot{q}) \) and a "potential energy function" \( V(t,q) \). The principle states that the actual motion of the system (namely, the function \( t \mapsto q(t) \)) is a critical point of the functional

\[
f(q) = \int_a^b T - V.
\]

In the prototype situation where the system consists of a particle moving in \( \mathbb{R}^3 \) under the influence of a force \( F: \mathbb{R}^3 \rightarrow \mathbb{R}^3 \), with \( F = -\nabla V \) for some function \( V: \mathbb{R}^3 \rightarrow \mathbb{R} \), we have a kinetic energy function defined by

\[
T: (x,y,z) \mapsto (1/2)m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)
\]

and we use

\[
V: (x,y,z) \mapsto V(x,y,z)
\]
as the potential energy function. Hamilton's Principle says that the motion of the particle is then given by the Euler-Lagrange equation of the functional

\[
f(x,y,z) = \int_a^b (1/2)m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(x,y,z) \, dt,
\]
i.e. \( -\frac{\partial V}{\partial x} = m\ddot{x}, \quad -\frac{\partial V}{\partial y} = m\ddot{y}, \quad -\frac{\partial V}{\partial z} = m\ddot{z} \). This is just Newton's Third Law, \( F = m(\ddot{x}, \ddot{y}, \ddot{z}) \).

Classically, Hamilton's Principle is usually deduced from Newton's Laws, for whatever particular system is under consideration. It is therefore nothing new. However, it is interesting *theoretically* because it shows that Newton's Laws are essentially the Euler-Lagrange equations for a variational problem (not every second order system of differential equations is!), and it is useful *practically* because it sometimes provides an easy way of

17
obtaining the equations of motion of a system.

Before illustrating these two points, we pause to mention some classical terminology. The set of all possible states (or positions) of a system is called the "configuration space" of the system. Mathematically, the configuration space is the set of all possible values of \( q_1, \ldots, q_n \), the "parameters" describing the system. The function \( L = T - V \) is called the "Lagrangian" of the system, and the Euler-Lagrange equations of the functional \( \int_a^b L \) are called the "equations of motion" or "Lagrange's equations of motion". If the potential energy \( V(t, q) \) does not contain the variable \( t \) explicitly, the force is said to be "conservative". The "total energy" of the system is the function \( T + V \).

Here is a theoretical application of Hamilton's Principle:

**Proposition: (The Principle of Conservation of Energy)** Suppose a system moves under the influence of a conservative force, and that \( T = T(q, \dot{q}) \) is a homogeneous expression of degree 2 in the variables \( \dot{q}_1, \ldots, \dot{q}_n \) which does not contain \( t \) explicitly. Then the total energy \( T + V \) is independent of \( t \).

**Proof.** Since \( T - V \) does not contain \( t \) explicitly, we have the first integral \( \int \dot{q}_i \frac{\partial T}{\partial \dot{q}_i} - (T-V) = C \) for some constant \( C \). But by Euler's identity for homogeneous functions, we have \( \sum \dot{q}_i \frac{\partial T}{\partial \dot{q}_i} = 2T \). Hence \( T + V = C \). \( \Box \)

Note that \( \frac{\partial}{\partial t}(T+V) \) is obviously zero, since \( t \) does not appear explicitly; the proposition says that in fact, for a critical point, \( d/dt(T+V) = 0 \) as well. If we restrict attention to those \((q_1, \ldots, q_n)\) for which \( T+V \) is constant, a critical point of \( \int_a^b T - V \) is the same as a critical point of \( \int_a^b T \) (or \( \int_a^b V \)). Classically, the functional \( \int_a^b T \) is called the "action", and Hamilton's Principle is called the "Principle of Least Action" (even though it involves critical points, not necessarily minimum points).

Now for some practical examples:

**Example 1.** Consider a rigid body moving in \( \mathbb{R}^3 \) under the influence of some force. The position of the body may be specified by six variables, namely the coordinates of the centre of mass \((x, y, z)\) and the "Euler angles" \((\theta, \phi, \gamma)\) (which specify the position of a system of coordinate axes \( O'x', O'y', O'z' \) fixed relative to the body, with respect to the standard coordinate axes \( Ox, Oy, Oz \)). Thus there are six parameters \( q_1, \ldots, q_6 \) in this situation.
Example 2. Consider a system of two jointed rods, fixed at one end at the origin in $\mathbb{R}^2$, fixed together but free to rotate, and with the other end free. The position of this system may be specified by two parameters $q_1, q_2$, namely the angles $\theta, \phi$ shown in the diagram.

Example 3. In this example we shall obtain the equations of motion. Consider a "spherical pendulum" of unit length, moving under gravity, with "bob" of mass $m$.

The potential energy function is given by $V = mgz = mg\cos\Theta$ (taking the plane $Oxy$ as the "reference level"), and the kinetic energy function is given by $(1/2)m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = \frac{1}{2}m(\dot{\theta}^2 + \dot{\phi}^2\sin^2\Theta)$ (using cylindrical polar coordinates). By Hamilton's Principle, the equations of motion are the Euler-Lagrange equations of the functional.
\[ m \int_a^b g \cos \theta - \left(1/2\right) \left(\dot{\theta}^2 + \phi^2 \sin^2 \theta \right) \, dt. \]

We obtain the two equations

\[ g \sin \theta + \phi^2 \sin \theta \cos \theta - \ddot{\theta} = 0 \]

\[ \dot{\phi} \sin^2 \theta = C \]

for some constant \( C \). A first integral of these equations can be written down, namely the conservation of energy equation.

**Example 4.** Consider a string, fixed at \((0,0)\) and \((0,1)\) and allowed to vibrate in the plane \(Oxy\). Let the mass of the string per unit length be \( m \), and let Hooke’s constant be \( k \).

Assume that the effect of the weight of the string is negligible compared to the effect of the tension in the string. This example is more general than the previous ones, as the configuration space is "infinite dimensional": the system is specified by a function \( q(x,t) \) rather than a point \( q_1(t), \ldots, q_n(t) \) (for each value of \( t \)).

The potential energy function is \( V = \int k(ds-dx) = k \int_0^1 \left(1 + q_x^2\right)^{1/2} \, dx \) (using only the tension, and ignoring gravity). To simplify matters even further, let us approximate the integrand here by \( (1/2)q_x^2 \) (using the binomial theorem). The kinetic energy function is \( T = (1/2) \int_0^1 m q_t^2 \, dx \). Thus Hamilton’s Principle gives the equations of motion as the Euler-Lagrange equations of the functional

\[ \int_a^b \int_0^1 \left( m/2 \right) q_t^2 - \left( k/2 \right) q_x^2 \, dx \, dt, \]

i.e. the usual wave equation \( q_{xx} = (m/k) q_{tt} \).
Lecture 6: Smooth manifolds

Here are some of the difficulties we have encountered so far:

1. What is \( \mathfrak{U} \)?

2. What is the best definition of critical point?

3. What constraints are allowed (or, what is \( \mathfrak{U} \))?

4. What is a "configuration space"?

The concept of a manifold provides a way of dealing with these difficulties.

**Definition:** Let \( M \) be a topological space. A smooth atlas for \( M \) is a family \( \{(U_\alpha, f_\alpha)\}_{\alpha \in A} \) where \( U_\alpha \) is an open subset of \( M \) and \( f_\alpha: U_\alpha \to \mathbb{R}^n \) is a homeomorphism onto an open subset \( V \) of \( \mathbb{R}^n \), such that

1. \( \bigcup_{\alpha \in A} U_\alpha = M \)

2. for all \( \alpha, \beta \in A \), \( f_\beta \circ f_\alpha^{-1} \) (on the largest permissible domain) is smooth with a smooth inverse (the inverse being \( f_\beta \circ f_\alpha^{-1} \)).

Thus a smooth atlas provides a "good parametrization" for \( M \). The functions \( f_\alpha \) are called *charts* of the atlas. Two atlases are equivalent if their union is an atlas. A smooth structure for \( M \) is an equivalence class of atlases. A smooth manifold is a topological space \( M \) together with a choice of equivalence class of smooth atlases. Usually we omit the word "smooth".

**Lemma:** If \( M \) is path connected in the above definition, the numbers \( n \) coincide.

**Sketch of proof.** Cover a path connecting any two points by a finite number of charts.

This number is called the *dimension* of the manifold \( M \).

**Definition:** Let \( M, N \) be manifolds. A map \( f: M \to N \) is smooth if, for atlases \( \{(U_\alpha^M, f_\alpha^M)\}_{\alpha \in A} \), \( \{(U_\beta^N, f_\beta^N)\}_{\beta \in B} \), all maps of the form \( f_\beta^N \circ (f_\alpha^M)^{-1} \) are
smooth maps (in the usual sense).

A **diffeomorphism** is a smooth map \( f : M \to N \) of manifolds such that \( f^{-1} \) exists and is also a smooth map. A natural but very difficult problem would be to classify all manifolds, up to diffeomorphism. Unfortunately even very elementary spaces may possess several inequivalent smooth structures (e.g. the seven dimensional sphere, studied by Milnor in the 1950's, or \( \mathbb{R}^4 \), an example which only came to light in the 1980's through the work of Donaldson and Freedman).

**Example 1.** \( \mathbb{R}^n \) (or any open subset of \( \mathbb{R}^n \)) with \( \{ (\mathbb{R}^n, \text{id}) \} \).

**Example 2.** The n-dimensional sphere \( S^n = \{ x \in \mathbb{R}^{n+1} \mid \|x\| = 1 \} \). This has an atlas of the form \( \{(D_+, s_+), (D_-, s_-)\} \) where \( D_+, D_- \) are the northern and southern hemispheres, and \( s_+, s_- \) are stereographic projection. Alternatively, an atlas for \( S^1 \) may be constructed using the map \( t \mapsto (\sin t, \cos t) \).

**Example 3.** Let \( C \) be the set of all positions of the jointed rods in example 2 of lecture 5. This may be given a topology and an atlas by using the natural map \( \mathbb{R}^2 \to C \). In fact, \( C \) is then diffeomorphic to the product \( S^1 \times S^1 \) (in general, the product of two manifolds is in a natural way a manifold).

**Example 4.** Real projective space \( \mathbb{R}P^n \) is defined to be the set of all lines through the origin in \( \mathbb{R}^{n+1} \). It acquires an atlas via the map \( S^n \to \mathbb{R}P^n \).

**Example 5.** Let \( C \) be the set of all positions of a rigid body in \( \mathbb{R}^3 \) (see example 1 of lecture 5). This has a natural manifold structure, which makes it diffeomorphic to \( \mathbb{R}^3 \times \mathbb{R}P^3 \).

**Example 6.** It is easy to classify all path connected manifolds of dimension one. Any such manifold is diffeomorphic to \( \mathbb{R} \) or to \( S^1 \).
Lecture 7: Submanifolds

Definition: Let $M$ be a manifold of dimension $m$. A subset $N \subseteq M$ is called a \textit{submanifold of} $M$ if there exists an atlas $\{(U_{\alpha}, f_{\alpha})\}_{\alpha \in A}$ of $M$ such that any $x \in N$ is contained in some $U_{\alpha}$ with:

1. $f_{\alpha}(U_{\alpha}) = V_{\alpha} = V_{1} \times V_{2}$ where $V_{\alpha}, V_{1}, V_{2}$ are all open subsets of Euclidean spaces, and

2. $f_{\alpha}(U_{\alpha} \cap N) = V_{1} \times \{y_{2}\}$ (where we write $f_{\alpha}(x) = (y_{1}, y_{2})$).

Proposition: A submanifold of $M$ has the structure of a smooth manifold, whose underlying topology is the induced topology from $M$.

\textit{Sketch of proof.} An atlas is given by $\{(U_{\alpha} \cap N, p_{1}f_{\alpha}|_{\alpha \in N})\}$ where the sets $U_{\alpha}$ are those taken from the definition of submanifold, and where $p_{1} : V_{1} \times V_{2} \to V_{1}$ is the projection.

Thus a submanifold of $M$ is a subset of $M$ which is a manifold in its own right, and which is compatible with the manifold structure of $M$. The dimension of the submanifold (in the definition above), assuming it to be path connected, is the dimension of each set $V_{1}$.

Warning: There are other (not quite equivalent) definitions of a submanifold!

For example, $S^{n}$ is a submanifold of $\mathbb{R}^{n+1}$.

Proposition: Let $U$ be an open subset of $\mathbb{R}^{n}$, and let $f : U \to \mathbb{R}$ be smooth. Take $a \in U$ such that $\frac{\partial f}{\partial x_{n}}(a) \neq 0$. Let $f(a) = c$. Then there exists an open subset $U_{a}$ of $\mathbb{R}^{n}$ containing $a$, and a diffeomorphism $f : U_{a} \to V_{1} \times V_{2}$, with $f(a) = (a_{1}, \ldots, a_{n-1}, c)$, where $V_{1}$ is an open subset of $\mathbb{R}^{n-1}$ containing $(a_{1}, \ldots, a_{n-1})$ and $V_{2}$ is an open subset of $\mathbb{R}$ containing $c$, such that $f(U_{a} \cap f^{-1}(c)) = V_{1} \times \{c\}$.

\textit{Sketch of proof.} Apply the inverse function theorem to the function $(x_{1}, \ldots, x_{n}) \mapsto (x_{1}, \ldots, x_{n-1}, f(x))$. 

23
In other words, $U_a \cap f^{-1}(c)$ is a submanifold of $\mathbb{R}^n$ (of dimension $n-1$). More generally, the same argument proves:

**Theorem:** Let $U$ be an open subset of $\mathbb{R}^n$. Let $f: U \to \mathbb{R}$ be smooth. Assume that $(\nabla f)_a \neq 0$ for all $a \in f^{-1}(c)$ for some $c \in \mathbb{R}$. Then $f^{-1}(c)$ is a submanifold of $\mathbb{R}^n$ of dimension $n-1$.

**Example 1.** $S^n = f^{-1}(1)$ where $f(x_1, \ldots, x_{n+1}) = \sum x_i^2$.

**Example 2.** $SL_n(\mathbb{R}) = \det^{-1}(1)$ where $\det$ is the determinant function on the open subset $GL_n(\mathbb{R})$ of the vector space of all real $n \times n$ matrices.

The theorem is in fact a reformulation of a special case of the implicit function theorem. The full implicit function theorem is equivalent to the following theorem, whose proof is analogous to that of the previous theorem.

**Theorem:** Let $U$ be an open subset of $\mathbb{R}^n$. Let $f: U \to \mathbb{R}^m$ be smooth, with $m \leq n$. Assume that $(\partial f/\partial x_j)_a$ has maximal rank for all $a \in f^{-1}(c)$, for some $c \in \mathbb{R}^m$. Then $f^{-1}(c)$ is a submanifold of $\mathbb{R}^n$ of dimension $n-m$.

**Example 3.** $O_n = f^{-1}(I)$ where $f(A) = AA^t$. 

24
Lecture 8: The derivative and the tangent bundle

For a smooth map \( f : U \rightarrow V \) of open subsets of Euclidean spaces (of dimensions \( m, n \) respectively), the "derivative of \( f \)" is usually thought of as the Jacobian matrix

\[
(\partial f_i / \partial x_j)
\]

of \( f \). Alternatively, and more canonically, the derivative may be defined as that function

\[
U \rightarrow \text{Hom}(\mathbb{R}^m, \mathbb{R}^n), \quad a \mapsto Df_a
\]

where \( Df_a \) is the linear transformation uniquely defined by the formula

\[
f(a+v) = f(a) + Df_a(v) + o(|v|).
\]

It turns out to be convenient to define the derivative of \( f \) to be the map

\[
Df : U \times \mathbb{R}^m \rightarrow V \times \mathbb{R}^n \quad \text{given by} \quad Df(a,v) = (f(a), Df_a(v)).
\]

**Example 1.** If \( n = m = 1 \), then \( Df = (f,f') \). We have \( Df_a(1) = f'(a) \).

**Example 2.** If \( m = 1 \), \( Df = (f,f') \) and \( Df_a(1) = f'(a) \). We call \( Df_a(1) \) the velocity vector of the curve \( f \) at \( a \).

If now \( f : M \rightarrow N \) is a smooth map of manifolds, we should like to define \( Df \) to be the collection of all those functions \( D(f^N_{\mathcal{B}} f^M_{\mathcal{A}})^{-1}) \), where the functions \( f^M_{\mathcal{A}}, f^N_{\mathcal{B}} \) vary over all charts of \( M \) and \( N \). We should like this collection to constitute a map from some object \( M' \) to another object \( N' \), where \( M', N' \) are obtained by taking all of the sets \( U \times \mathbb{R}^m, V \times \mathbb{R}^n \) respectively.

**Definition:** Let \( M \) be a manifold. For any \( x \in M \), the tangent space to \( M \) at \( x \) is defined to be the set

\[
T_x M = \{ \text{smooth curves } \gamma : (-\varepsilon, \varepsilon) \rightarrow M \mid \gamma'(0) = x \}/\sim,
\]

where the equivalence relation \( \sim \) is given by: \( \gamma_1 \sim \gamma_2 \) if and only if \( f_{\mathcal{A}} \gamma_1 \) and \( f_{\mathcal{A}} \gamma_2 \) have the same velocity vector at 0, for some chart \((U, f_{\mathcal{A}})\) at \( x \).
Definition: The tangent bundle of a manifold $M$ is the set $TM = \bigcup_{x \in M} T_x M$ (disjoint union).

Definition: Let $f : M \to N$ be a smooth map. The derivative of $f$ is the map $Df : TM \to TN$ defined as follows. Let $v \in T_x M$. Then $Df(v)$ is the equivalence class of the curve $f \circ \gamma$ for any curve $\gamma \in v$.

Note that $Df(T_x M) \subseteq T_{f(x)} N$. We write $Df_x$ for the map $T_x M \to T_{f(x)} N$ obtained by restricting $Df$. Note that there is a natural "projection" map

$$\pi : TM \to M, \quad v \mapsto \gamma(0) \text{ (for any } \gamma \in v).$$

This map is onto, and $\pi^{-1}(x) = T_x M$ for any $x \in M$.

Theorem: 1) Let $M$ be a manifold. Then $TM$ may be given the structure of a smooth manifold such that $\pi$ is a smooth map.
2) Let $f : M \to N$ be a smooth map. Then $Df : TM \to TN$ is also a smooth map.

Sketch of proof. A suitable atlas for $TM$ may be constructed from an atlas $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ of $M$ by taking $\{(\pi^{-1}(U_\alpha), Df_{\phi_\alpha})\}_{\alpha \in A}$, where $Df_{\phi_\alpha}$ is suitably identified.

Observe that $TM$ is a manifold whose dimension is twice that of $M$. 

26
Lecture 9: Vector Fields

We begin with two further remarks on the tangent bundle.

1. Let $f : M \to N$ be a smooth map between manifolds $M, N$. We say $f$ is an *immersion* if $Df_x$ is injective for all $x \in M$, a *submersion* if $Df_x$ is surjective for all $x \in M$, and an *embedding* if it is an immersion and a homeomorphism from $M$ to $f(M)$ (with the induced topology on $M$).

If $f : M \to N$ is an immersion, we say that $M$ is *immersed* in $N$. If $f : M \to N$ is an embedding, we say that $M$ (or $f(M)$) is *embedded* in $N$; in this case $f(M)$ is diffeomorphic to $M$, and it is a submanifold of $N$. A fundamental problem in differential topology is to decide when a given manifold can be immersed or embedded in another manifold. For example, it is known that every two-dimensional manifold embeds in $\mathbb{R}^4$, but not necessarily in $\mathbb{R}^3$.

2. If $M$ is a submanifold of $\mathbb{R}^n$, the inclusion map $i : M \to \mathbb{R}^n$ is an embedding, and one has $Di_x : T_xM \to \mathbb{R}^n$ for each $x \in M$. Obviously $\text{Im } Di_x$ is just
$$\{ v \in \mathbb{R}^n \mid v = \dot{\gamma}(0) \text{ for some smooth curve } \gamma : (-\epsilon, \epsilon) \to M \text{ with } \gamma(0) = x \}.$$ 

If $M = g^{-1}(c)$ for some smooth function $g : \mathbb{R}^{m+1} \to \mathbb{R}$ which satisfies the condition $(\nabla g)_x \neq 0$ for all $x \in M$ (see lecture 7), then $\text{Im } Di_x$ is
$$\{ v \in \mathbb{R}^n \mid v \perp (\nabla g)_x \}.$$ 

As an application of this, we can now give a clear proof of the "Lagrange Multiplier Rule" of calculus (c.f. lecture 4). Let $g : \mathbb{R}^{m+1} \to \mathbb{R}$ be smooth, and let $M$ be $g^{-1}(c)$ (i.e. the set of points of $\mathbb{R}^{m+1}$ subject to the constraint $g(x) = c$). Assume that $(\nabla g)_x \neq 0$ for all $x \in M$. Let $f : \mathbb{R}^{m+1} \to \mathbb{R}$ be another smooth function. If $x \in M$ is a critical point of the function $f|_M$, then there exists some constant $\lambda$ such that $(\nabla f)_x = \lambda (\nabla g)_x$.

**Definition:** A (smooth) vector field on a manifold $M$ is a smooth map $X : M \to TM$ such that $\pi \circ X = I_M$.

An important property of vector fields is that they operate on functions (this is a generalization of "directional derivative"): 

27
Definition: Let $X$ be a vector field on $M$. Let $f : M \to \mathbb{R}$ be a smooth function. The function $Xf : M \to \mathbb{R}$ is defined by $Xf(m) = \frac{d}{dt}(f \cdot \gamma)|_{t=0}$ where $\gamma$ is any curve in the class $X(m)$.

It is easy to check that $Xf$ is well defined, and is a smooth function. Moreover, it follows from the usual properties of $\frac{d}{dt}$ that $X$ is a linear operator, and that it is a "derivation", i.e. $X(fg) = fXg + gXf$ for any functions $f$, $g$.

Lemma: Let $X$, $Y$ be vector fields on $M$. Let $m \in M$. Then $X(m) = Y(m)$ if and only if $Xf = Yf$ for all smooth functions defined on open subsets of $M$ containing $m$.

Thus the operation on functions actually determines the vector field. Some authors use this to define a vector field as a first order linear differential operator. The function $Xf$ should be thought of as the "directional derivative of $f$ in the direction of $X$".

An important question in differential topology is to determine whether a manifold possesses a vector field which is never zero (i.e. $X(m) \not= 0$ for all $m \in M$). For example, odd-dimensional spheres do, but even-dimensional spheres do not.

Vector fields on an open subset $U$ of $\mathbb{R}^n$ are easy to describe explicitly. First, given the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, one defines the standard vector fields $\partial/\partial x_i$ as follows. For any $x \in U$, the value of $\partial/\partial x_i$ at $x$ is the equivalence class of the curve $t \mapsto x + te_i$. Note that the operation of $\partial/\partial x_i$ on functions $f : U \to \mathbb{R}$ is given by $f \mapsto \partial f/\partial x_i$, i.e. the usual partial derivative; this explains the choice of notation. Any vector field on $U$ may now be written in the form $X = \sum F_i \partial/\partial x_i$, for some smooth functions $F_1, \ldots, F_n : U \to \mathbb{R}$. (To obtain the functions $F_i$, just compose $X : U \to TU$ with the identification $TU \cong U \times \mathbb{R}^n$ of lecture 8.)

A similar "local expression" may be given for a vector field on an arbitrary manifold $M$. If $f : U \to V , \ldots, x_1, \ldots, x_n$ is a chart for $M$, then $X \cdot f_i^{\perp}$ is a vector field on $V$, so it may be written in the form $\sum F_i \partial/\partial x_i$, where the $x_i$ are now coordinates in the vector space containing $V$. This is called "expressing $X$ in local coordinates".
Lecture 10: Differential Forms

Definition: The cotangent bundle (or dual of the tangent bundle) of a manifold $M$ is the set $T^*M = \bigcup_{x \in M}(T_xM)^*$ (disjoint union).

As for the tangent bundle, $T^*M$ is itself a manifold and one has an analogous smooth map $\pi: T^*M \to M$. (It may be more informative to call this map $\pi^*$, but to simplify notation we shall not do so.) More generally, one may define $\otimes^k TM$, $\Lambda^k(TM)$, $S^k TM$, etc., by using the appropriate operations on the vector spaces $T_xM$. All such constructions produce manifolds, each of which has a smooth "projection" map $\pi$ to $M$.

Definition: A Riemannian metric on a manifold $M$ is a smooth map $g: M \to \mathbb{R}^2 T^*M$ such that $\pi \circ g = I_M$ and $g(x)$ is a positive definite (symmetric bilinear form) for all $x \in M$.

Recall that an element of the vector space $\mathbb{R}^2 V^*$ (for any vector space $V$) may be identified with a symmetric bilinear form on $V$. Thus, a Riemannian metric is simply a smooth assignment of an inner product to each of the vector spaces $T^*M$.

Definition: A (smooth) $k$-form on a manifold $M$ is a smooth map $\alpha: M \to \Lambda^k T^*M$ such that $\pi \circ \alpha = I_M$.

Thus, a $k$-form is a smooth assignment of an alternating multilinear form (of $k$ variables) to each of the vector spaces $T^*M$. It follows from the definitions that if $\alpha$ is a 1-form and $X$ is a vector field, then we can form $\alpha(X)$, and that this is a smooth function from $M$ to $\mathbb{R}$. If $g$ is a Riemannian metric on $M$, we have a one to one correspondence between vector fields $X$ and 1-forms $\alpha$ given by $g(x)(X,Y) = \alpha(Y)$ (for all vector fields $Y$). This generalizes the usual correspondence in linear algebra between vectors and linear functionals, in the presence of an inner product.

We define the standard 1-forms $dx_1, \ldots, dx_n$ on any open subset $U$ of $\mathbb{R}^n$ by requiring that $dx_i(\partial/\partial x_j) = \delta_{ij}$. Then, as for vector fields, any 1-form may be expressed in the form $\sum_i \alpha_i dx_i$, for some smooth functions $\alpha_1, \ldots, \alpha_n: U \to \mathbb{R}$. Similarly, any 1-form on a manifold may be expressed locally in this form.

Two important operations, the wedge product and the exterior derivative, provide much of the justification for studying differential forms:
1. Wedge product

If $\alpha$ is a $a$-form and $\beta$ is an $b$-form, we define the wedge product of $\alpha$ and $\beta$ to be the $(a+b)$-form $\alpha \wedge \beta$ defined by $\alpha \wedge \beta (v_1, \ldots, v_{a+b}) = \sum \text{sign}(\sigma) \alpha(v_{\sigma(1)}, \ldots, v_{\sigma(a)}) \beta(v_{\sigma(a+1)}, \ldots, v_{\sigma(a+b)})$, where the sum is over all permutations $\sigma$ of the numbers $1, \ldots, a+b$ with the property $\sigma(1) < \ldots < \sigma(a)$ and $\sigma(a+1) < \ldots < \sigma(a+b)$. We omit the verification that $\alpha \wedge \beta$ is indeed a $(a+b)$-form. Note however that if $a=b=1$ we get $\alpha \wedge \beta (v, w) = \alpha(v) \beta(w) - \alpha(w) \beta(v)$. Note also that (for general $a$ and $b$) we have $\alpha \wedge \beta = (-1)^{ab} \beta \wedge \alpha$.

Using the wedge product, one can now write any $k$-form on an open subset $U$ of $\mathbb{R}^n$ in the form $\sum \alpha_{i_1 \ldots i_k} dx_{i_1} \ldots dx_{i_k}$ where the sum is over all integers $i_1, \ldots, i_k$ with $i_1 < \ldots < i_k$, for some smooth functions $\alpha_{i_1 \ldots i_k} : U \to \mathbb{R}$. Any $k$-form on a manifold may be expressed locally in this form. Since $dx_i dx_j = -dx_j dx_i$ (for $i \neq j$) and $dx_i dx_i = 0$, we see that any $k$-form on a manifold of dimension $n$ is necessarily the zero $k$-form, if $k > n$.

2. Exterior derivative.

The exterior derivative $d$ associates to a $k$-form $\alpha$ a $(k+1)$-form $d\alpha$ in such a way that the following properties hold:

(a) For any smooth function $f : M \to \mathbb{R}$ (i.e. a 0-form on $M$), $df$ satisfies $df(X) = Xf$ for any vector field $X$ on $M$.

(b) If $\alpha$ is an $a$-form and $\beta$ is a $b$-form, then $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^a \alpha \wedge d\beta$.

Clearly, if we take (a) to be the definition of $d$ on 0-forms, then (b) gives an inductive procedure for defining $d$ on $k$-forms. (We omit the verification that $d$ is in fact a $(k+1)$-form.)

Proposition: $d \circ d$ (i.e. $d^2$) = 0.

The notation $dx_i$ for the $i$-th standard 1-form can now be explained. For if $x_i$ is the $i$-th coordinate function on any open subset $U$ of $\mathbb{R}^n$, then the exterior derivative $dx_i$ satisfies $dx_i(\partial / \partial x_j) = (\partial / \partial x_j) x_i = \delta_{ij}$. 

36
Definition: Let $g$ be a Riemannian metric on a manifold $M$, and let $f: M \rightarrow \mathbb{R}$ be smooth. The gradient of $f$ is the vector field $\nabla f$ which corresponds to the 1-form $df$ via $g$.

Example: Let $U$ be any open subset of $\mathbb{R}^3$. There exist non-zero $k$-forms only for $k=0,1,2,3$. A 0-form is a function $f: \mathbb{R}^3 \rightarrow \mathbb{R}$, and $df = (\partial f/\partial x_1)dx_1 + (\partial f/\partial x_2)dx_2 + (\partial f/\partial x_3)dx_3$. Any 1-form may be written $\alpha = f_1dx_1 + f_2dx_2 + f_3dx_3$, and (using the defining properties of $d$ above) we find $d\alpha = (-\partial f_1/\partial x_2 + \partial f_2/\partial x_1)dx_1 \wedge dx_2 + (\partial f_1/\partial x_3 - \partial f_3/\partial x_1)dx_1 \wedge dx_3 + (-\partial f_2/\partial x_3 + \partial f_3/\partial x_2)dx_2 \wedge dx_3$. Finally, any 2-form may be written $\alpha = f_1dx_2 \wedge dx_3 + f_2dx_3 \wedge dx_1 + f_3dx_1 \wedge dx_2$, and we find $d\alpha = [\partial f_1/\partial x_1 + \partial f_2/\partial x_2 + \partial f_3/\partial x_3]dx_1 \wedge dx_2 \wedge dx_3$. Thus the exterior derivative in this situation gives the familiar operations grad, curl, div.

One of the most important features of $k$-forms on a manifold $M$ is that they can be integrated over $k$ dimensional submanifolds of $M$. In fact, they can be integrated more generally over submanifolds with boundary. Denoting the boundary of such a submanifold $N$ by $\partial N$, one then has a generalization of Green's theorem, Stokes' theorem etc., which may be written simply as: $\int_{\partial N} \alpha = \int_{\partial N} \alpha$, for any $(k-1)$-form $\alpha$. Unfortunately we shall not have time to go into the details of this.
Lecture 11: Applications of global analysis in the calculus of variations

We shall discuss two applications of the ideas of lectures 6-10 to the material of lectures 1-5. In this lecture, we shall show that the terminology of manifolds and derivatives is an appropriate language for the calculus of variations, the use of which helps to avoid some of the "difficulties" encountered in the classical theory. In future lectures we shall discuss the way in which the theory of manifolds may be used to advance the classical theory.

We begin by re-phrasing some results of ordinary calculus in terms of the theory of (finite dimensional) manifolds. Let \( M \) be a manifold, and let \( f : M \rightarrow \mathbb{R} \) be a smooth function. A fundamental problem is to study the local extrema of \( f \).

**Definition:** A point \( x \in M \) is **critical** for \( f \) if and only if \( \frac{d}{dt}(f(\gamma(t)))|_{t=0} = 0 \) for all smooth curves \( \gamma : (-\epsilon, \epsilon) \rightarrow M \) such that \( \gamma(0) = x \).

**Proposition 1:** If \( x \) is a local extremum of \( f \), then \( x \) is a critical point of \( f \).

**Proof:** If \( x \) is a local extremum of \( f \), then 0 is a local extremum of \( f \circ \gamma \), for any curve \( \gamma \) of the above form. Since \( f \circ \gamma \) is a smooth real function of a real variable, the result follows. \( \text{QED} \)

**Proposition 2:** A point \( x \) is critical for \( f \) if and only if \( Df_x = 0 \).

**Proof:** According to the definition, \( x \) is critical if and only if \( Df_x \gamma_0(1) = 0 \) for all curves \( \gamma \) of the above form. But \( T_x M = \{ D\gamma_0(1) | \gamma \text{ is a curve of the above form} \} \), hence the result. \( \text{QED} \).

In the (often encountered) case where \( M \) is a submanifold of some Euclidean space \( \mathbb{R}^n \), we can simplify the condition for a critical point:

**Proposition 3:** Assume \( M \) is a submanifold of \( \mathbb{R}^n \). Identify \( T_x M \) with a linear subspace of \( \mathbb{R}^n \) in the usual way. Then \( x \) is critical for \( f \) if and only if \( (d/dt)f(x+tv)|_{t=0} = 0 \) for all \( v \in T_x M \).

**Proof:** The condition \( (d/dt)f(x+tv)|_{t=0} = 0 \) for all \( v \) is equivalent to \( Df_x(v) = 0 \) for all \( v \), i.e. \( Df_x = 0 \). This is the condition for a critical point of \( f \), by proposition 2. \( \text{QED} \)
Finally, the following definition generalizes a well known concept from calculus:

**Definition:** Let \( f : M \to \mathbb{R} \) be a smooth function, as above. Let 
\( s : M \to S^2(T^*M) \) be a Riemannian metric on \( M \). Then the gradient of \( f \) is the function \( \nabla f : M \to TM \) defined by: \( s(\nabla f(x), v) = Df_x(v) \) for all \( v \in T_xM \).

It is easy to check that \( \nabla f \) is a smooth function such that \( \tau \circ \nabla f = \text{Id}_M \), i.e., it is a vector field on \( M \). In fact, if we consider \( Df \) as a differential form (in which case it is precisely \( df \)), then \( \nabla f \) is the vector field which is dual to this 1-form. Obviously, the "zeros" of this vector field are precisely the critical points of \( f \). It is important to note that the definition of \( \nabla f \) depends on a choice of a Riemannian metric for \( M \).

All this generalizes directly when \( M \) is replaced by an infinite dimensional manifold. This is the situation of interest in the calculus of variations, where we take the infinite dimensional manifold \( \text{Map}(M, N) \) consisting of all smooth maps \( f : M \to N \) (where \( M \) and \( N \) are arbitrary finite dimensional manifolds). Now, we have not actually discussed infinite dimensional manifolds, although it is clear from lectures 6-10 how the theory of such manifolds should be set up: instead of using Euclidean space as a "model space", we should use an appropriate infinite dimensional vector space. Thus, we define an infinite dimensional manifold in the same way as a finite dimensional manifold, except that the charts \( f : U \to V \) are such that \( V \) is an open subset of a fixed infinite dimensional vector space \( E \). Unfortunately there are many inequivalent types of infinite dimensional spaces (e.g. Banach spaces, Hilbert spaces,...) and hence many inequivalent theories. As we do not have time to get into such matters, we shall in this lecture merely assume that \( \text{Map}(M, N) \) may be given a suitable manifold structure, and just re-phrase some of the classical results from chapters 1-5 in such a way that it becomes blindingly obvious that there is a theory of calculus on infinite dimensional manifolds such as \( \text{Map}(M, N) \) which generalizes everything concerning finite dimensional manifolds in lectures 6-10, and of which the results of lectures 1-5 are a special case. For the details of this theory, we refer to the book of Palais and the article of Eells listed in the references.

The general calculus of variations problem is to study the local extrema of a smooth function \( f : \mathcal{U} \to \mathbb{R} \), where \( \mathcal{U} \) is a submanifold of \( \text{Map}(M, N) \).

In the simple situation of lecture 1, \( \mathcal{U} \) is an affine subspace of the vector space \( \text{Map}([a, b], \mathbb{R}) \). So \( \text{Map}([a, b], \mathbb{R}) \) is certainly an infinite dimensional manifold, and \( \mathcal{U} \) is a submanifold. Moreover, the function defined by \( f(y) = \int_{a}^{b} F(x, y, y') \, dx \) is a smooth function (because \( F \) is smooth). We have

\[
f(y + h) = f(y) + \int_{a}^{b} (F_y - (d/dx)F_y) h + o(h)
\]

33
and so the derivative of $f$ is given by

$$Df(y)(h) = \int_a^b (F_y - (d/dx)F_y)h = \langle \nabla f(y), h \rangle$$

where we use the inner product $\langle h, k \rangle = \int_a^b h(x)k(x)dx$ to obtain the gradient. (An inner product on a vector space $E$ defines a Riemannian metric on the manifold $E$ in an obvious way. The gradient of a smooth function taking values in $E$ is a priori a smooth map taking values in $TE$, but since $TE$ may be identified with $E \times E$, we obtain again a map taking values in $E$ itself.) Thus, the gradient $\nabla f_y$ of $f$ at $y$ is just the function $F_y - (d/dx)F_y$ (evaluated at $(x, y, y')$), and this vanishes if and only if $y$ is a solution of the Euler-Lagrange equation.

Classically, the map $\nabla f : y \mapsto F_y - (d/dx)F_y$ is called the first variation of $f$, and is written $\delta f$.

In all the examples of lecture 2, the Euler-Lagrange equations can similarly be interpreted as the equation $\nabla f_y = 0$. In lecture 4 we discussed "constraints"; these can be interpreted as follows.

First, consider the integral constraint $g(y) = c$, where $g(y) = \int_a^b G(x, y, y')dx$, in the situation of lecture 1. In terms of gradients, the Euler-Lagrange equations may be written in the form $\nabla f(y) = \lambda \nabla g(y)$ (for some constant $\lambda$). This applies under the assumption that $\nabla g(y) \neq 0$ (see lecture 1). But now we see that this is precisely the analogue of the Lagrange Multiplier method for functions on finite dimensional manifolds (see lecture 8). The strange condition $\nabla g(y) \neq 0$ is the condition we expect which is needed to ensure that $g^{-1}(c)$ is (at least, locally) a submanifold of $\mathcal{U}$. We emphasize, however, that this is merely an analogy; to prove an extension of the Lagrange Multiplier method to infinite dimensional manifolds, we would need to have an extension of the implicit function theorem.

The situation for non-integral constraints is easier. A typical example of such a constraint is to choose a submanifold $N'$ of $N$, and impose the condition on the maps $y : M \to N$ that $y(M) \subseteq N'$. It can be shown that the subset $\text{Map}(M, N')$ is in general a submanifold of $\text{Map}(M, N)$, so one expects the subset $\mathcal{V} = \text{Map}(M, N')$ to be a submanifold of $\mathcal{U}$. The problem is then to study the critical points of the restriction of $f : \mathcal{U} \to \mathbb{R}$ to $\mathcal{V}$. The example of lecture 5 is a simple example of this type. Here $N = \mathbb{R}^2$, and $N' = \{(u, v) \in \mathbb{R}^2 \mid S(u, v) = 0\}$. The Euler-
Lagrange equations were derived under the assumption that \((S_{y_1}S_{y_2}) \neq 0\) at \((y_1,y_2)\).

This condition is now seen to be very sensible: it is the condition that \(N\) be (locally, near each point \((y_1,y_2)\)) a submanifold of \(\mathbb{R}^2\).

Finally we mention a general framework for classical mechanics (c.f. lecture 5; see the book of Arnold for further information). The starting point is

1) a manifold \(M\) (the "configuration space" of the mechanical system), and

2) a smooth function \(L : TM \to \mathbb{R}\) (the "Lagrangian function").

The function \(L\) may be assumed to be of the form \(T - \mathbb{U}\), where \(T(v) = (1/2)S(v,v)\) for some Riemannian metric \(s\) on \(M\) (the "kinetic energy"), and \(U : M \to \mathbb{R}\) is any smooth function (the "potential energy"). The problem is then to find critical points of the functional

\[
\int_a^b L(D\gamma(t)) \, dt
\]

which is defined on smooth paths in \(M\) (i.e. smooth functions \(\gamma : [a,b] \to M\) with prescribed values at \(a\) and \(b\)). Such critical points represent "motions of the mechanical system", and the Euler-Lagrange equations are the "equations of motion".

35
Lecture 12: Noether's theorem

Consider the variational problem with functional \( f(y) = \int_a^b F(x,y,y') \, dx \), as in lecture 1. The Euler-Lagrange equation is \( F_y = d/dx(F_y') \), i.e., a second order ordinary differential equation for \( y \). In lecture 1, we saw how useful it was to have a "first integral" of the Euler-Lagrange equation. Such first integrals can be obtained in the special cases where \( F \) "does not contain one of the variables explicitly". If \( y \) or \( y' \) is missing, the Euler-Lagrange equation simplifies in an obvious way, as the corresponding partial derivative is zero. If \( x \) is missing, we obtained the following first integral (see the lemma of lecture 1):

\[
y''F_y' - F = C
\]

where \( C \) is constant. This is certainly not obvious, and we might ask whether other such first integrals exist in certain circumstances. This is potentially of great practical value, since it is usually difficult to solve the Euler-Lagrange equation directly. Noether's theorem provides a way of finding such first integrals.

The essential ingredient is the concept of "invariance" (or "symmetry") of a variational problem. Consider a (smooth, invertible) change of variable from \((x,y)\) to \((X,Y)\), so that we may write \( X = X(x,y) \), \( Y = Y(x,y) \). We say that the variational problem is \emph{invariant} under this transformation if

\[
\int_a^b F(x,y,y') \, dx = \int_{u^*}^{v^*} F(X,Y,Y') \, dX
\]

for all \( u,v \in [a,b] \), where \( Y' \) means \( dY/dX \). and we write \( u^* = X(u,y(u)) \), \( v^* = X(v,y(v)) \).

**Theorem (Noether):** Suppose that the variational problem above is invariant under each of a smooth family of transformations \( (x,y) \mapsto (x_t,y_t) \), where \((x_0,y_0) = (x,y)\). Then the following equation is satisfied by any extremal \( y \) of the variational problem:

\[
F_y' \neq F + (F - yF_y') \bar{x} = C
\]

where \( \bar{x} = dx_t/dt \bigg|_{t=0} \) and \( \bar{y} = dy_t/dt \bigg|_{t=0} \), and where \( C \) is some constant.

**Proof.** Since \( y \) is an extremal, we know that the usual Euler-Lagrange equation \( F_y - \)
\( \frac{d}{dx}(F_y) = 0 \) is satisfied. Since this equation does not depend on the interval \([a,b]\), the restriction of \( y \) to any sub-interval \([u,v]\) is an extremal for the corresponding variational problem.

Let us write \( u_t = x_t(u,y(u)), \ v_t = x_t(v,y(v)) \). By invariance,
\( \frac{d}{dt} \int_{u_t}^{v_t} F(x_t,y_t) \, dx_t \bigg|_{t=0} = 0. \) Applying the change of variable \( x_t \mapsto x \) to the integral, we get
\( \frac{d}{dt} \int_{u}^{v} F(x_t,y_t,dy_t/dx_t)(dx_t/dx) \, dx \bigg|_{t=0} = 0. \) Differentiating under the integral sign, and applying the Mean Value Theorem as in lecture 1, we have

\[
\int_{u}^{v} [(F_x(dx/dt) + F_y(dy/dt) + F_{y'}(d/dt)(dy/dx_t))(dx/dx) + F(d/dt)(dx_t/dx)] \, dx \bigg|_{t=0} = 0.
\]

Now, at \( t=0 \), we have \( dx/dt = \overline{x} \) and \( dy/dt = \overline{y} \), and also \( dx/dx = 1 \). Hence we obtain
\( d/dt(dy/dx_t) = d/dt[(dy/dx)(dx/dx)]^{-1} = d/dx(\overline{y}) - y'd/dx(\overline{x}) \) at \( t=0 \). Substituting these values in the integral above, we obtain

\[
\int_{u}^{v} [F_x \overline{x} + F_y \overline{y} + F_{y'}d/dx(\overline{y}) + (F - y'F_y)d/dx(\overline{x})] \, dx = 0.
\]

Integration by parts now gives

\[
\int_{u}^{v} [F_x \overline{x} + F_y \overline{y} - \overline{y}d/dx(F_y) - \overline{x}d/dx(F-y'F_y)] \, dx + \left[ \overline{y}F_y \right]_u^v + \left[ \overline{x}F - y'F_y \right]_u^v = 0.
\]

Using the Euler-Lagrange equation \( F_y - d/dx(F_y) = 0 \), this reduces to

\[
\left[ \overline{y}F_y + \overline{x}(F-y'F_y) \right]_u^v = 0.
\]

Since this is true for all \( u, v \), the quantity \( \overline{y}F_y + \overline{x}(F-y'F_y) \) must be constant on \([a,b]\), as required. QED

This proof illustrates fully the deficiencies of the standard notation \( F_x, F_y \), etc. Like democracy (according to Winston Churchill), it is the worst possible system, except for all the others.

**Example 1.** If \( F \) "does not contain \( x \) explicitly", then the problem is invariant under the family of transformations given by \( x(t) = x + t, \ y(t) = y \). Since \( \overline{x} = 1 \) and \( \overline{y} = 0 \), the theorem gives the first integral \( F - y'F_y = C \), as we have seen already.

**Example 2.** If \( F \) "does not contain \( y \) explicitly", then the problem is invariant under the
family of transformations given by \( x(t) = x, \ y(t) = y + t \). Here, \( \overline{x} = 0, \overline{y} = 1 \), so the theorem gives the first integral \( F_y = C \), which we have already noted as an immediate consequence of the Euler-Lagrange equation.

Noether's theorem extends to variational problems where the functional is of the form \( f(y_1, \ldots, y_n) = \int_a^b F(x, y_1, \ldots, y_n, y_1', \ldots, y_n') \, dx \). If such a problem is invariant under a smooth family of transformations \( (x, y_1, \ldots, y_n) \mapsto (x_t, (y_1)_t, \ldots, (y_n)_t) \), of which \( t = 0 \) gives the identity transformation, then one obtains the first integral

\[
\sum F_{y_i} \overline{y_i} + (F - \sum y_i' F_{y_i}) \overline{x} = C
\]

where \( \overline{x} = \frac{dx}{dt}|_{t=0} \) and \( \overline{y} = \frac{dy_i}{dt}|_{t=0} \), for some constant \( C \).

**Example 3.** Consider the problem of a particle of mass \( m \) moving in \( \mathbb{R}^n \) under the influence of a force given by a (conservative) potential function \( V : \mathbb{R}^n \rightarrow \mathbb{R} \). We shall denote the path of the particle by \( t \mapsto (q_1(t), \ldots, q_n(t)) \) (as is traditional in mechanics). According to Hamilton's Principle (see lecture 5), the path is an extremal of the functional \( f(q_1, \ldots, q_n) = \int_a^b T - V \), where \( T \) is the kinetic energy function \( T(t) = (1/2)m \sum q_i^2 \). As \( T - V \) does not contain \( t \) explicitly, we have the first integral \( T + V = C \) for some constant \( C \) (see lectures 5 and example 1 above). In other words, for the actual motion of the particle, the total energy is conserved. If \( V \) does not contain \( q_i \) explicitly, i.e. the force has no component in the \( q_i \)-direction, we obtain the first integral \( (T - V)q_i = C \) for some constant \( C \) (see example 2 above). This says that the quantity \( m \dot{q}_i \) is conserved during the motion ("conservation of the \( q_i \)-component of momentum").

These first integrals have already been obtained earlier. Here is a new example:

**Example 4.** Let \( n = 3 \) in example 3. Assume that the potential function \( V \) is rotationally symmetric about the \( q_3 \)-axis, i.e. that \( V(q_1, q_2, q_3) = V(q_1 \cos(s) + q_2 \sin(s), q_3 \sin(s) + q_2 \cos(s), q_3) \) for any \( s \). (For example, this would be the case if the force were always directed towards the \( q_3 \)-axis.) Then the variational problem is invariant under the family of transformations given by \( t = t, \ (q_1)_s = q_1 \cos(s) + q_2 \sin(s), \ (q_2)_s = -q_1 \sin(s) + q_2 \cos(s), \ (q_3)_s = q_3 \). We can apply the theorem to this situation, with \( \overline{T} = 0, \overline{q_1} = q_2, \overline{q_2} = -q_1, \overline{q_3} = 0 \). Thus we obtain the first integral \( q_1 \ddot{q}_2 - q_1q_2 = C \), for some constant \( C \). In other words, angular momentum about the \( q_3 \)-axis is conserved during the motion.

(Recall that the total angular momentum of the particle is defined as the vector \( m(q_1, q_2, q_3) \times (\dot{q}_1, \dot{q}_2, \dot{q}_3) \).)
Noether's theorem generalizes to variational problems with higher derivatives and with several independent variables (in particular, to the second and third generalizations of the simplest problem, as described in lecture 2). For example, if the functional $f(z) = \int_{\mathcal{D}} F(x,y,z,z_x,z_y) dx dy$ is invariant under a family of transformations $(x,y,z) \mapsto (x_1,y_1,z_1)$, then one obtains by the same argument the equation

$$\text{div} \left( VF_z + F_xV_{z_x} + F_yV_{z_y} \right) = 0$$

where $V = z - z_x \frac{\partial}{\partial x} - z_y \frac{\partial}{\partial y}$, which is satisfied whenever $z$ is a critical point. If $z$ is in fact a function of $x$ alone, then $\frac{\partial}{\partial y}$, $z_y$, and $F_{z_y}$ are all zero, and one obtains the first integral stated earlier. Note, however, that in general one does not obtain an expression which is constant on any extremal, merely one whose divergence vanishes. For more general formulae, including the case where the functional contains higher derivatives of the dependent variables, we refer to the book of Logan.
Lecture 13: Remarks on Noether's theorem

1. **Infinitesimal invariance.**

Obviously the proof of Noether's theorem in lecture 12 only requires

$$\frac{d}{dt}\int_{a}^{t} F(x_{t},y_{t},\frac{dy_{t}}{dx_{t}})dx_{t}\bigg|_{t=0} = 0$$

rather than invariance. This property is called *infinitesimal invariance*. Note that invariance is equivalent to the equation

$$F(x,y,y') = \int F(x_{t},y_{t},y_{t}')dx_{t}/dx$$
on [a,b]. Infinitesimal invariance is equivalent to this equation holding up to $o(t)$.

2. **Sources of invariance.**

Let $M$, $N$ be manifolds, and let $\mathcal{U}$ be a submanifold of $\text{Map}(M,N)$. Consider a functional of the form $f(y) = \int_{M} L(Dy)$. As we shall use this problem only to make a general point, we shall avoid giving precise details; the reader may think of the example at the end of lecture 11, where $M$ is replaced by an interval $[a,b]$. Examples of families of transformations under which the variational problem is invariant arise from families of "symmetries" of the domain $M$, i.e. smooth maps $K:(-\epsilon,\epsilon) x M \rightarrow M$, such that $m \mapsto K(t,m)$ is a diffeomorphism of $M$ and such that $t=0$ gives the identity map of $M$. Symmetries of the target $N$ give examples in a similar fashion. All the examples of lecture 12 are of this form. The most general transformations considered in lecture 12 allow "mixing" of the domain and target manifolds, however.

3. **Generalized invariance.**

There is no reason why one should not consider transformations of the form $X = X(x,y,y')$, $Y = Y(x,y,y')$ for the functional $f(y) = \int_{a}^{b} F(x,y,y')dx$. Similarly, for the more general variational problems, one may consider transformations which involve derivatives of the dependent variables. Invariance under a family of transformations of this type is known as *generalized invariance*. Noether's theorem extends to variational problems which possess this kind of invariance.
4. The inverse problem (see sections 3.5, 6.4 of the book of Logan).

It is interesting to know when a general second order differential equation (or system of second order equations) is the Euler-Lagrange equation (or system of Euler-Lagrange equations) of some variational problem. For then one may look for families of transformations leaving the problem invariant, each of which gives rise to an identity, by Noether's theorem. For example, suppose we are given a differential equation

\[ y'' = P(x,y,y'). \]

The "inverse problem" is to search for a functional \( f(y) = \int_a^b F(x,y,y')dx \), whose Euler-Lagrange equation is the given equation. The Euler-Lagrange equation is (in expanded form)

\[ F_y - F_{xy} y' - F_{yy} y'' - F_{yy} y^2 = 0. \]

If we substitute \( y'' = P(x,y,y') \) we obtain \( F_y - F_{xy} y' - F_{yy} y'' - F_{yy} y P = 0 \). Differentiation with respect to \( y' \) gives \( F_{yy} y' + y' F_{yy} y'' + F_{yy} y' P + F_{yy} y^2 P = 0 \). Putting \( u = F_{yy} \) we obtain finally an equation

\[ P_y u + u_x + y'u_y + P u_y' = 0 \]

which is a first order linear partial differential equation for the function \( u \) (of three independent variables). Hence, a solution \( u \) may be found, and hence \( F \) may be obtained from the equation \( F_{yy} = u \) (by integrating twice). Thus, the inverse problem may be solved in this case.

**Example (the Korteweg-de Vries equation).**

This famous equation first arose in the theory of solitary waves travelling in a channel. It may be written

\[ u_t - 6uu_{xx} + u_{xxx} = 0. \]

On writing \( u = v_x \) one obtains the equation \( v_{tx} - 6v_{xvxx} + v_{xxxx} = 0 \). One can then conjure up the functional \( f(v) = \int_D F(t,x,v,v_t,v_x,v_{tt},v_{xt},v_{xx})dt \) with \( F(t,x,v,v_t,v_x,v_{tt},v_{xt},v_{xx}) = (1/2)v_x v_t - v_x^3 - (1/2)v_{xx}^2 \). whose Euler-Lagrange equation is this equation. This variational problem has certain obvious invariance properties, since the variables \( t, x, \) and \( v \) are absent. Hence one obtains three identities which are satisfied by any solution of the
KdV equation (see the book of Logan, section 6.4).
Lecture 14: Hamiltonian mechanics and symplectic reduction

Consider the functional \( f(q_1,\ldots,q_n) = \int_a^b F(t,q_1,\ldots,q_n,\dot{q}_1,\ldots,\dot{q}_n)\,dt \), where the functions \( q_1,\ldots,q_n \) are smooth real valued functions on \([a,b]\) taking prescribed values at the endpoints. We shall use the \((t,q)\) notation (and the terminology of mechanics) rather than our usual \((x,y)\) notation as we shall think of the function \( t \mapsto (q_1,\ldots,q_n) \) as defining the path of a particle moving in \( \mathbb{R}^n \) (or, more generally, as the local coordinate representation of the path of a particle moving in an \( n \)-dimensional manifold \( M \)). The theory to be presented in this lecture was originally developed in this situation.

If the Lagrangian function \( F \) does not contain \( t \) explicitly (e.g. if \( F = T - V \) where \( V \) is the potential function for a conservative force), we know that the quantity \( F - \sum q_i F_{q_i} \) is constant during the actual motion of the particle. (In fact, we have seen that this represents the total energy.) This can be used to help solve the system of Euler-Lagrange equations

\[
F_{q_1} - \frac{d}{dt}(F_{\dot{q}_1}) = 0, \ldots, F_{q_n} - \frac{d}{dt}(F_{\dot{q}_n}) = 0
\]

(a system of \( n \) second order equations in \( q_1,\ldots,q_n \), if we make a change of variable.

We define \( p_1 = F_{\dot{q}_1}, \ldots, p_n = F_{\dot{q}_n} \), so that we now have a system of \( 2n \) first order equations in the \( 2n \) variables \( q_1,\ldots,q_n, p_1,\ldots, p_n \). It turns out that quantity \( H = F + \sum q_i F_{\dot{q}_i} = -\dot{F} + \sum q_i p_i \) (i.e. minus the total energy) plays the same role as \( F \):

**Proposition:** The function \( F \) satisfies the system of Euler-Lagrange equations if and only if \( H \) satisfies the system of equations

\[
dq_i/dt = \partial H/\partial p_i, \quad dp_i/dt = -\partial H/\partial q_i.
\]

**Proof.** Using (for convenience) the language of differential forms (on \( \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \)) we have \( dH = -dF + \sum p_i dq_i + \sum \dot{q}_i dp_i = -F_t dt - \sum F_{q_i} dq_i - \sum F_{\dot{q}_i} d\dot{q}_i + \sum p_i d\dot{q}_i = -F_t dt - \sum F_{\dot{q}_i} dq_i + \sum \dot{q}_i dp_i \). This allows us to read off the partial derivatives of \( H \) with respect to the variables \( t, q_1,\ldots,q_n, p_1,\ldots,p_n \). First, \( \partial H/\partial p_i = \dot{q}_i \). Next, \( \partial H/\partial q_i = -F_{\dot{q}_i} \), so the Euler-Lagrange equations \( F_{\dot{q}_i} = d/dt(F_{\dot{q}_i}) \) (\( = d/dt(p_i) \)) are now seen to be equivalent to the equation \( \partial H/\partial q_i = -d/dt(p_i) \). QED

The function \( H \) is called the *Hamiltonian* function, and the \( 2n \) equations in
q_1,\ldots,q_n, p_1,\ldots, p_n$ are known as Hamilton's equations. What we have done here is merely apply the standard trick to convert a second order equation into two first order equations by introducing a new variable. However, the choice of the variables $p_1,\ldots, p_n$ resulted in a particularly symmetrical system of first order equations.

We can now describe the principal method of attacking problems in mechanics (there is no guarantee that the attack will be successful). The idea is to choose the variables $q_1,\ldots, q_n$ in such a manner that one of them is absent from $F$ (if possible). Without loss of generality, suppose that $q_1$ is such a variable. (Such a "missing" variable is called a cyclic variable, classically.) Then $q_1$ and $p_1$ may be eliminated entirely from the $2n$-2 equations consisting of Hamilton's equations for $i = 2,\ldots, n$, as follows. First, $q_1$ is missing by hypothesis. Thus $F_{q_1} = 0$, and this implies $H_{q_1} = 0$. But $H_{q_1} = -dp_1/dt$, so $dp_1/dt = 0$ and $p_1$ is constant. This eliminates $p_1$ too. If the system of $2n$-2 first order equations in the $2n-2$ variables $q_2,\ldots, q_n, p_2,\ldots, p_n$ can then be solved, the remaining two quantities $q_1$ and $p_1$ can be found since $p_1$ is constant and $q_1$ satisfies the first order equation $dq_1/dt = \delta H/\delta p_1$. Of course, the aim would be to solve the system of $2n$-2 equations by the same procedure, and proceed inductively.

Noether's theorem has not been used explicitly here, although it is clearly in the background. First, the choice of the function $H$ was motivated by the first integral obtained through invariance of the Lagrangian under time translation. Second, the elimination of the variables $q_1, p_1$ was made possible essentially by the first integral obtained through invariance of the Lagrangian under translation of $q_1$. We are certainly not using the full strength of Noether's theorem in this approach, as we are only considering invariance under families of transformations of the configuration space. However, such transformations are the ones which suggest themselves most naturally in this context.

**Example 1.** Consider a particle of mass $m$, moving on a surface $S$ of revolution in $\mathbb{R}^3$ whose equation is given in cylindrical polar coordinates by $r = r(\theta)$. Assume that there is no applied force, so the potential energy is zero. The Lagrangian is then equal to the kinetic energy: $L(r, \theta, \dot{r}, \dot{\theta}) = (m/2)(\dot{r}^2 + r^2 \dot{\theta}^2 + z^2) = (m/2)(\dot{r}^2 + (1+z^2) + r^2 \dot{\theta}^2)$ (where $z = dz/dr$). The Euler-Lagrange equations are

$$
\frac{d}{dt}(r^2 \dot{\theta}) = 0
$$

$$
\dot{r}^2 z' z'' + r \dot{\theta}^2 - \frac{d}{dt}(f(1+z^2)) = 0.
$$
The variable $\theta$ is cyclic, and the elimination of this variable may be performed by writing \[ r^2 \dot{\theta} = A \] for some constant $A$, and then substituting in the second equation. We obtain a second order equation for $r$, which in principle may be solved to give $r = r(t)$. The first order equation $r^2 \dot{\theta} = A$ may then be solved to give $\theta = \theta(t)$.

Although we did not need to use the Hamiltonian formalism here, this example illustrates the fact that for a two dimensional configuration space, the existence of just one cyclic variable is usually enough to solve the problem completely. In terms of the earlier notation, having eliminated $q_1$ and $p_1$, we are left with a system of two first order equations in $q_2$ and $p_2$ (or, equivalently, a second order equation in $q_2$). Then we need to solve a first order equation in $q_1$, and we are done.

Remarks: (1) The first integral $r^2 \dot{\theta} = A$ may be interpreted as saying that the component of the angular momentum of the particle about the $z$-axis is constant. It also has a geometrical interpretation; the quantity $r \dot{\theta}$ is essentially $\cos \alpha$, where $\alpha$ is the angle between the velocity vector of the particle and the $xy$-plane. Hence $r \cos \alpha$ is constant during the motion of the particle. This statement is known as Clairut's theorem. (Note that the functionals $\int_a^b (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) dt$ and $\int_a^b (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{-1/2} dt$ have the same critical points up to re-parametrization, and hence that the path of the particle is a geodesic on the surface.)

(2) As a computational aid to solving the second order equation for $r$, note that by conservation of total energy we have \[ r^2 (1 + z^2) + r^2 \dot{\psi}^2 = E \] for some constant $E$. Substituting for $\dot{\theta}$ from the first Euler-Lagrange equation, we obtain the first order equation \[ r^2 (1 + z^2) + A^2 / r^2 = E \] for $r$. Differentiation of this equation gives the second Euler-Lagrange equation.

**Example 2.** Consider the special case of example 1 where the surface $S$ is the sphere $S^2$ of radius 1. Here there is additional invariance, which may be seen on using spherical-polar coordinates $\theta, \phi$. The Lagrangian function is just \[ (m/2)(\dot{\theta}^2 + \dot{\phi}^2), \] so both $\theta$ and $\phi$ are cyclic variables, and from the Euler-Lagrange equations we obtain the system of two first order equations $\ddot{\theta} = A, \ddot{\phi} = B$ for constants $A, B$.

**Example 3.** Consider a particle of mass $m$ moving in $M = \mathbb{R}^3$, with no applied force. Then the Lagrangian function is \(F(t, x, y, z, \dot{x}, \dot{y}, \dot{z}) = (m/2)(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)\) and the Euler-Lagrange equations are \(\dot{p}_1 = m \dot{x}, \dot{p}_2 = m \dot{y}, \dot{p}_3 = m \dot{z}\) and \(H(t, x, y, z, p_1, p_2, p_3) = (1/2m)(p_1^2 + p_2^2 + p_3^2)\).
Example 4. Consider a rigid body moving in $\mathbb{R}^3$ with one point of the body fixed at the origin. The configuration space is a manifold $M \cong SO_3 \cong \mathbb{R}P^3$ (where $\cong$ indicates a diffeomorphism). Assume again for simplicity that there is no applied force. Then there are three independent families of rotations preserving $M$, but no local parametrization of $M$ exists in which three variables are cyclic.

It is possible to give a general criterion for an expression $K(q_1, \ldots, q_n, p_1, \ldots, p_n)$ to be a first integral of Hamilton's equations. All that is required is that $d/dt(K(q_1, \ldots, q_n, p_1, \ldots, p_n)) = 0$ for any solution $q_1, \ldots, q_n, p_1, \ldots, p_n$ of the equations. Differentiation gives

$$
\frac{d}{dt}(K) = \sum \frac{\partial K}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial K}{\partial p_i} \frac{dp_i}{dt}
$$

$$
= \sum \left[ \frac{\partial K}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial K}{\partial p_i} \frac{\partial H}{\partial q_i} \right]
$$

This quantity, which depends only on $K$ and $H$, is usually called the Poisson bracket of $K$ and $H$, and written $\{K, H\}$. Thus, $K$ gives a first integral if and only if it "Poisson commutes" with the Hamiltonian $H$. This condition admits another interpretation: it is easy to check that $\{K, H\} = 0$ if and only if $\omega(\nabla K, \nabla H) = 0$, where $\omega$ is the 2-form $\sum dq_i \wedge dp_i$. (This 2-form is called the canonical symplectic form on $T\mathbb{R}P^3$.) We are interested in the set

$$
A_H = \{ K \mid \{K, H\} = 0 \}.
$$

This is clearly a vector space. (In fact, the pairing $\{ , \}$ endows it with a Lie algebra structure.) If $\dim A_H = n$, then we expect to have $n$ cyclic variables, and we expect to be able to reduce the Euler-Lagrange equations to a system of $n$ equations, each of which is a first order equation in one of these cyclic variables. However, this is not quite correct (as example 4 shows); we need in addition some condition to ensure that the $n$ variables are "simultaneously cyclic". It turns out that a suitable condition is that $A_H$ be "Poisson commutative", i.e. that $\{K, K'\} = 0$ for all $K, K' \in A_H$, but we shall not prove this here. If this condition holds, the variational problem is said to be completely integrable.

If $\dim A_H = n-1$ (and $A_H$ is Poisson commutative), the problem is still tractable, as the Euler-Lagrange equations reduce to a system of first order equations in $n-1$ of the variables together with a second order equation in the remaining variable (as in example 1).

We conclude with a general remark on the underlying manifold theory. The basic idea is
that a variable $q_i$ is cyclic when the configuration space $\mathcal{M}$ admits a family of
diffeomorphisms leaving the Lagrangian function $L : \mathcal{T}\mathcal{M} \to \mathbb{R}$ invariant. An
alternative way of saying this is that there is an action of the group $\mathbb{R}$ (or, at least, a subset
of $\mathbb{R}$) on $\mathcal{T}\mathcal{M}$, on whose orbits $L$ is constant. Hence, the problem should reduce to a
problem on the space of orbits $\mathcal{M}/\mathbb{R}$ of $\mathbb{R}$ on $\mathcal{M}$. If we are lucky, $\mathcal{M}/\mathbb{R}$ will be a
manifold (of dimension one less than the dimension of $\mathcal{M}$), and $L$ will induce a real
valued function on $\mathcal{T}(\mathcal{M}/\mathbb{R})$. Thus we have an easier problem to solve. An alternative
point of view is that one has a "momentum function" $p_i : \mathcal{T}\mathcal{M} \to \mathbb{R}$, and that we can
reduce the problem from $\mathcal{T}\mathcal{M}$ to $p_i^{-1}(A)/\mathbb{R}$, whose dimension is two less than the
dimension of $\mathcal{T}\mathcal{M}$. If the manifold $\mathcal{M}$ is acted upon by a Lie group $G$, then each one-
parameter subgroup of $G$ gives rise to a first integral, or momentum function; more
naturally, these momentum functions together give a momentum function
$\mathfrak{g} : \mathcal{T}\mathcal{M} \to \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of $G$. This Lie algebra is a subalgebra of
the Lie algebra $\mathcal{A}_H$. If $G$ is abelian, one can then reduce from $\mathcal{T}\mathcal{M}$ to $\mathcal{A}_H^{-1}(A)/G$. This
process is called symplectic reduction. It can be carried out fairly generally in the
framework of a Lie group acting on a manifold in such a way as to preserve a given
symplectic form on the manifold, and has many interesting applications.
References

1. Classical calculus of variations

C. Fox, *An Introduction to the Calculus of Variations*, Dover (New York) 1987

I. M. Gelfand, S. V. Fomin, *Calculus of Variations*, Prentice Hall (New Jersey) 1963

H. A. Lauwerier, *Calculus of Variations in Mathematical Physics*, Mathematisch Centrum Amsterdam 1966

2. Differentiable manifolds

M. Spivak, *Calculus on Manifolds*, Benjamin (New York) 1965


3. Modern calculus of variations


