A TASTE OF CALCULUS OF VARIATIONS

LIVIU I. NICOLAESCU

ABSTRACT. These are notes for a talk at Arlo Caine's MPPM Lectures Series, University of Notre Dame. We survey through examples some of the basic problems, principles and results of calculus of variations.

CONTENTS

INTRODUCTION

The calculus of variations can be thought of as a sort of calculus in infinitely many variables. The first problems of calculus of variations appeared immediately after the inception of calculus and attracted the attention of all the classics of mathematics.

They first dealt with the description of paths in an Euclidean space that are extrema of certain natural problems.

- *Geodesics.* The shortest paths on a given surface connecting a pair of given points.
- *Brachistochones.* Given two points P_0, P_1 in a vertical plane, find a path connecting P_0 to P_1 inside that plane such that a bead sliding along the path under its own weight will travel from P_0 to P_1 in the shortest amount of time.
- *Dido's Problem.* Find the closed plane curves of given length that surround the largest area.

The paths in each of the above examples are extrema of certain problems. We say that they are defined by a *variational principle*. This imposes severe restrictions on their behavior: they must be solutions of certain (nonlinear) second order differential equations called the *Euler-Lagrange equations*. Often, these suffice to determine the extrema, although many of the solutions of these equations are not extrema.

Surprisingly, most of the equations of classical mechanics are of Euler-Lagrange type, and Lagrange used this point of view to lay the foundations of what is now commonly referred to as *analytical*[1] *mechanics*. For a beautiful introduction to the Lagrangian approach to classical mechanics the nice lecture by R. Feynman [FLS] is still the best source.

We can clearly think of several variables functions satisfying variational principles: minimal (area) surfaces, multidimensional Dido (or isoperimetric) problem etc. The physical theory of classical fields is about certain several variables functions satisfying variational principles. In this case, the

Last modified on November 14, 2008.

[1] In the view of this author, the attribute *analytical* is a misnomer. The term *geometric* seems more appropriate. Arnold's classic [A] is a strong argument in favor of the geometric point of view.

Euler-Lagrange equations are nonlinear second order *partial* differential equations, and often are very difficult to solve explicitly. Even existence of solutions to these equations is far from obvious.

In this case we can turn things on their heads, and use the variational principles to prove the existence of solutions of very complicated partial differential equations.

In these notes we would like to present through examples some of the basic problems, principles and applications of calculus of variations. For more details on the classical aspects of calculus of variations we refer to Gelfand's notes [GF].

1. One-dimensional Euler-Lagrange Equations

Let us start with a simple motivating example. Consider a surface of revolution in the Euclidean space $\mathbb{R}^3$ described in cylindrical coordinates (r, θ, z) by the equation

$$S = \{ (r, \theta, z); r = f(z) \}, \tag{1.1}$$

where f is a smooth positive function defined on an interval $a < z < b$.

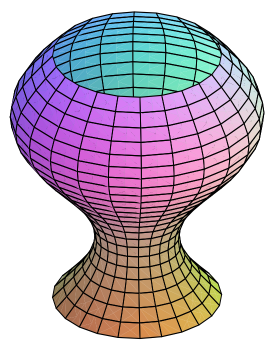

FIGURE 1. *The* **Maple** *generated plot of the surface of revolution* $r = 2 + \sin z$, $-\pi < z < \pi$.

We consider two points $p_i = (r_i, \theta_i, z_i)$, $i = 0, 1$, and we seek a shortest smooth path on S that connects p_0 to p_1. In Riemannian geometry, such a path is called a *geodesic*. How do we go about finding such a geodesic?

A smooth path γ on S can be described by indicating the position of the point $\gamma(t)$ in cylindrical coordinates. Due to (1.1) it suffices to know only the coordinates z and θ. Thus a path on S is described by a smooth path in the plane with coordinates (θ, z)

$$\gamma : [0, 1] \to \mathbb{R}^2_{(\theta,z)}, \;\; t \mapsto \big(\theta(t), z(t) \big)$$

We obtain a path on S using the defining equation of S

$$r = f(z(t)), \;\; \theta = \theta(t), z = z(t).$$

To find the length of this path we need to express it in Euclidean coordinates

$$x = r\cos\theta, \;\; y = r\sin\theta, \;\; z = z.$$

We regard x, y, z as functions depending on the time variable t. A dot will denote a t-derivative. The length of γ is then

$$\ell(\gamma) = \int_0^1 |\dot{\gamma}|dt = \int_0^1 \sqrt{|\dot{x}|^2 + |\dot{y}|^2 + |\dot{z}|^2}dt = \int_0^1 \sqrt{|\dot{r}|^2 + r^2|\dot{\theta}|^2 + |\dot{z}|^2}dt,$$

so that

$$\ell(\gamma) = \int_0^1 \underbrace{\sqrt{f(z)^2|\dot{\theta}|^2 + (1 + |f'(z)|^2)|\dot{z}|^2}}_{=:L}\, dt. \tag{1.2}$$

The integrand is a function L depending on 4-variables $L = L(\theta, z, \dot{\theta}, \dot{z})$. We denote the first pair of variables by $\boldsymbol{x}$ and the second pair of variables by $\dot{\boldsymbol{x}}$. We can now describe the length of γ as the integral

$$\ell(\gamma) = \int_0^1 L\big(\gamma(t), \dot{\gamma}(t)\big)dt.$$

If we denote by $\mathcal{X}_{p_0,p_1}$ the set of smooth paths on S that travel from p_0 to p_1 in one second, we see that the length of the shortest paths is given by

$$\min\left\{ \int_0^1 L\big(\gamma(t), \dot{\gamma}(t)\big)dt;\ \ \gamma \in \mathcal{X}_{p_0,p_1}, \right\}.$$

As is often the case, it will be convenient to solve a more general problem, and then specialize to the case at hand.

Consider the finite dimensional vector space $E = \mathbb{R}^n$. Viewed as a smooth manifold, its tangent bundle TE can be identified with the product $E \times E$, where the tangent space T_pE is be identified with $\{p\} \times E$. We will denote by $\boldsymbol{x} = (x^1, \dots, x^n)$ the coordinates on E by $\boldsymbol{v} = (v^1, \dots, v^n)$ the coordinates of a tangent vector. A *Lagrangian* on E is then a smooth function

$$L : TE \to \mathbb{R},\ \ L = L(\boldsymbol{x}, \boldsymbol{v}).$$

Any C^1 path $\gamma : [0, 1] \to E$ defines a continuous path on TE,

$$[0, 1] \ni t \mapsto \big(\gamma(t), \dot{\gamma}(t)\big) \in TE.$$

If $\mathcal{X}$ denotes the space of all C^1 paths $[0, 1] \to E$, then the Lagrangian function L defines an *action functional*

$$S_L : \mathcal{X} \to \mathbb{R},\ \ \mathcal{X} \ni \gamma \mapsto \int_0^1 L\big(\gamma(t), \dot{\gamma}(t)\big)dt.$$

The number $S_L(\gamma)$ is called the *action* of the path γ with respect to the Lagrangian L. We seek least action paths connecting two given points $p_0, p_1 \in E$. We denote by $\mathcal{X}_{p_0,p_1}$ the set of paths in $\mathcal{X}$ connecting p_0 to p_1.

Theorem 1.1 (The Least Action Principle). *If $\gamma : [0, 1] \to E$ is a C^2 path connecting two given points p_0, p_1 and*

$$S_L(\gamma) \le S_L(\varphi),\ \ \forall \varphi \in \mathcal{X}_{p_0,p_1}$$

then γ satisfies the Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \boldsymbol{v}}(\gamma, \dot{\gamma}) = \frac{\partial L}{\partial \boldsymbol{x}}(\gamma, \dot{\gamma}). \tag{$\boldsymbol{EL}$}$$

More precisely, γ satisfies the system of second order differential equations

$$\frac{d}{dt}\frac{\partial L}{\partial v^i}\big(\gamma(t), \dot{\gamma}(t)\big) = \frac{\partial L}{\partial x^i}\big(\gamma(t), \dot{\gamma}(t)\big),\ \ i = 1, \dots, n.$$

If we write $\gamma(t) = (x^1(t), \dots, x^n(t))$ then the above system can be rewritten as

$$\sum_{j=1}^n \Big(\frac{\partial^2 L}{\partial x^j \partial v^i}(\gamma, \dot{\gamma})\dot{x}^j + \frac{\partial^2 L}{\partial v^j \partial v^i}(\gamma, \dot{\gamma})\ddot{x}^j \Big) = \frac{\partial L}{\partial x^i}\big(\gamma(t), \dot{\gamma}(t)\big), \;\; i = 1, \dots, n.$$

Proof. Denote by $C_0^\infty([0,1], E)$ the space of smooth maps $\delta : [0,1] \to E$ such that $\delta(t) = 0$ in a neighborhood of 0 and 1. For any smooth path $\delta \in C_0^\infty([0,1], E)$ and any $\gamma \in \mathcal{X}_{p_0,p_1}$ we get a one parameter family of paths $\gamma_s \in \mathcal{X}_{p_0,p_01}$ (see Figure 2)

$$\gamma_s(t) = \gamma(t) + s\delta(t), \;\; s \in \mathbb{R}, \;\; t \in [0,1].$$

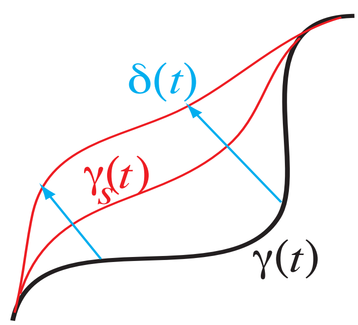

FIGURE 2. *Deforming the path $\gamma(t)$ using a displacement $\delta(t)$.*

We say that the family γ_s is a *deformation* of the path $\gamma(t)$. Define the smooth function

$$f : \mathbb{R} \to \mathbb{R}, \;\; f(s) = S_L(\gamma_s).$$

If γ is a least action path, then

$$f(0) = S_L(\gamma) \leq S_L(\gamma_s) = f(s), \;\; \forall s \in \mathbb{R}.$$

Hence $f'(0) = 0$ so that

$$\frac{d}{ds}|_{s=0} \int_0^1 L(\gamma + s\delta, \dot{\gamma} + s\dot{\delta})dt = 0.$$

We express $\delta(t)$ as a collection of paths $t \mapsto \delta^i(t)$, $1 \leq i \leq n$ and we deduce

$$0 = \int_0^1 \sum_{i=1}^n \left(\frac{\partial L}{\partial x^i}(\gamma, \dot{\gamma})\delta^i + \frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma})\dot{\delta^i} \right) dt, \;\; \forall \delta \in C_0^\infty([0,1], E) \tag{1.3}$$

for any path $\gamma \in \mathcal{X}_{p_0,p_1}$ such that

$$S_L(\gamma) = \min_{\varphi \in \mathcal{X}_{p_0,p_1}} S_L(\varphi).$$

If γ is actually a C^2 path, then an intgration by parts shows

$$\int_0^1 \frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma})\dot{\delta}^i dt = \left(\frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma})\delta^i(t) \right) \Big|_{t=0}^{t=1} - \int_0^1 \frac{d}{dt}\frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma})\delta^i dt = -\int_0^1 \frac{d}{dt}\frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma})\delta^i dt.$$

Hence, if γ is a C^2 least action path then

$$\int_0^1 \sum_{i=1}^n \left(\frac{\partial L}{\partial x^i}(\gamma, \dot{\gamma}) - \frac{d}{dt}\frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma}) \right) \delta^i dt = 0,$$

for any $\delta \in C_0^\infty([0,1], E)$. Because δ is arbitrary we conclude

$$\frac{\partial L}{\partial x^i}(\gamma, \dot{\gamma}) - \frac{d}{dt}\frac{\partial L}{\partial v^i}(\gamma, \dot{\gamma}) = 0, \quad \forall i = 1, \dots, n.$$

$\square$

Definition 1.2. A path $\gamma : (a,b) \to E$ satisfying the Euler-Lagrange equations (***EL***) with respect to the Lagrangian function $L : TE \to \mathbb{R}$ is called an *extremal* of the Lagrangian L. $\square$

Before we discuss several concrete applications of the least action principle let us comment on some issues raised by the above proof. These were eloquently and vigorously raised by K. Weierstrass more than a century and a half ago and had the effect of casting a shadow on this line of reasoning.

Remark 1.3 (Weierstrass critique)**. A. Existence** The above proof assumes a priori that a least action path exists. This requires a rigorous proof.
B. Regularity. Even if a minimizer exists, it might only be a C^1 path, so the last step in the proof fails. Thus one needs to prove that a minimizer, which a priori is only C^1, must have better regularity, namely that it is actually a C^2-path.

These two issues are even more severe for multi-dimensional variational calculus to be discussed in the next section. David Hilbert considered these to be fundamental issues and included them as problems 19, 20 in the famous list of Hilbert problems. $\square$

Example 1.4. Suppose $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a conservative force filed in $\mathbb{R}^3$, i.e., there exists a smooth function $U : \mathbb{R}^3 \to \mathbb{R}$ called *potential* such that

$$\boldsymbol{F} = -\nabla U.$$

The motion of a particle particle of mass m affected by this force field is governed by *Newton's equations*

$$m\ddot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) = -\nabla U(\boldsymbol{x}) \Longleftrightarrow m\ddot{x}^i = -\frac{\partial U}{\partial x^i}(x^1, x^2, x^3), \;\; i = 1, 2, 3..$$

Consider the Lagrangian $L = L_U : T\mathbb{R}^3 \to \mathbb{R}$

$$L(\boldsymbol{x}, \boldsymbol{v} = \frac{1}{2}m|\boldsymbol{v}|^2 - U(\boldsymbol{x}) = \frac{m}{2}\big(\,(v^1)^2 + (v^2)^2 + (v^3)^2\,\big) - U(x^1, x^2, x^3).$$

Then

$$\frac{\partial L}{\partial v^i} = mv^i, \quad \frac{\partial L}{\partial x^i} = -\frac{\partial U}{\partial x^i}.$$

If the path $t \mapsto \boldsymbol{x}(t)$ satisfies the Euler-Lagrange equations (***EL***) with respect to this Lagrangian function then

$$\frac{d}{dt}\frac{\partial L}{\partial \boldsymbol{v}^i}(\boldsymbol{x}, \dot{\boldsymbol{x}}) = m\frac{d}{dt}\dot{x}^i = \ddot{x}^i = \frac{\partial L}{\partial x^i} - \frac{\partial U}{\partial x^i}(\boldsymbol{x}).$$

We see that Newton's equation of motion are none other than the Euler-Lagrange equations for this Lagrangian function. $\square$

Example 1.5 (Geodesics on a surface of revolution. Part 1)**.** Consider the Lagrangian $L : T\mathbb{R}^2 \to \mathbb{R}$ we encountered in when we sought geodesics on a surface of revolution

$$L(\theta, z, \dot{\theta}, \dot{z}) = \sqrt{f(z)^2|\dot{\theta}|^2 + (1 + |f'(z)|^2)|\dot{z}|^2}$$

For simplicity we set

$$h(z) := \sqrt{(1 + |f'(z)|^2)}$$

so that we can write

$$L(\theta, z, \dot{\theta}, \dot{z}) = \sqrt{f(z)^2|\dot{\theta}|^2 + h(z)^2|\dot{z}|^2}.$$

We observe that

$$\frac{\partial L}{\partial \theta} = 0, \quad \frac{\partial L}{\partial z} = \frac{1}{L}\big(f(z)f'(z)|\dot{\theta}|^2 + h(z)h'(z)|\dot{z}|^2 \big)$$

$$\frac{\partial L}{\partial \dot{\theta}} = \frac{1}{L} f(z)^2\dot{\theta}, \quad \frac{\partial L}{\partial \dot{z}} = \frac{1}{L} h(z)^2 \dot{z}.$$

Thus the Euler-Lagrange equations take the form

$$\begin{cases} \frac{1}{L} f(z)^2\ddot{\theta} + \frac{2}{L} f(z)f'(z)\dot{z}\dot{\theta} + \frac{d}{dt}\left(\frac{1}{L}\right) f(z)^2\dot{\theta} & = & 0 \\ \frac{1}{L} h(z)^2\ddot{z} + \frac{2}{L} h(z)h'(z)\dot{z}^2 + \frac{d}{dt}\left(\frac{1}{L}\right) h(z)^2\dot{\theta} & = & \frac{1}{L}\big(f(z)f'(z)|\dot{\theta}|^2 + h(z)h'(z)|\dot{z}|^2 \big). \end{cases} \tag{1.4}$$

They seem hopeless, don't they! At this point, we need to rely on the geometric nature of this problem to make further progress.

Observe that if $\gamma : [0,1] \to S$ is a shortest length path between p_0 and p_0 then for very smooth, increasing function $\alpha : [0,a] \to [0,1]$ such that $\alpha(0) = 0$, $\alpha(a) = 1$ the composition

$$[0,1] \ni s \mapsto \alpha(s) \mapsto \gamma(\alpha(s)) \in S,$$

is also a shortest length path because the length of a path is independent of the reparametrization of a path.

Now consider the arclength function along $\gamma(t)$,

$$s(t) = \int_0^t |\dot{\gamma}(\tau)| d\tau = \int_0^t L(\gamma(\tau), \dot{\gamma}(\tau)\,) d\tau.$$

Hence $ds = Ldt$ so that

$$\frac{d}{dt} = L\frac{d}{ds} \ \text{ and in particular, } \ \dot{\gamma} = \frac{d\gamma}{dt} = L\frac{d\gamma}{ds}.$$

We deduce

$$\frac{\partial L}{\partial \dot{\theta}}(\gamma, \dot{\gamma}) = \frac{1}{L} f(z)^2 \frac{d\theta}{dt} = f(z)^2\frac{d\theta}{ds}.$$

Similarly

$$\frac{\partial L}{\partial \dot{z}} = \frac{1}{L} h(z)^2 \frac{dz}{dt} = h(z)^2 \frac{dz}{ds}$$

The Euler-Lagrange equations become

$$0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = L\frac{d}{ds}\Big(f(z)^2\frac{d\theta}{ds} \Big) \tag{1.5a}$$

$$\frac{\partial L}{\partial z} = \frac{d}{dt}\frac{\partial L}{\partial \dot{z}} = L\frac{d}{ds}\Big(h(z)^2\frac{dz}{ds} \Big) \tag{1.5b}$$

From (1.5a) we deduce that $f(z)^2\frac{d\theta}{ds}$ is independent of s. Hence, there exists a constant c such that

$$\frac{d\theta}{ds} = \frac{c}{f(z)^2}. \tag{1.6}$$

Observing that

$$\frac{\partial L}{\partial z} = \frac{1}{L}\big(f(z)f'(z)|\dot{\theta}|^2 + h(z)h'(z)|\dot{z}|^2 \big) = L\left(f(z)f'(z)\left|\frac{d\theta}{ds}\right|^2 + h(z)h'(z)\left|\frac{dz}{ds}\right|^2 \right)$$

we deduce from (1.5b) that

$$\frac{d}{ds}\Big(h(z)^2\frac{dz}{ds}\Big) = f(z)f'(z)\left|\frac{d\theta}{ds}\right|^2 + h(z)h'(z)\left|\frac{dz}{ds}\right|^2,$$

or equivalently

$$h(z)\frac{d}{ds}\Big(h(z)\frac{dz}{ds}\Big) = f(z)f'(z)\left|\frac{d\theta}{ds}\right|^2 = \frac{c^2 f'(z)}{f(z)^3}.$$

From the last equality we can determine z explicitly as a function of s. Then using (1.6) we can also determine θ explicitly as a function of s. The exact expressions may not be as illuminating. However the equality

$$f(z)^2\frac{d\theta}{ds} = c$$

has a nice geometric interpretation.

Our computations show that the tensor

$$g = f(z)^2 d\theta^2 + h(z)^2 dz^2$$

is the metric on the surface of revolution induced by the Euclidean metric. Then we can write

$$\dot\gamma = \dot\theta\partial_\theta + \dot z\partial_z,$$

and we have

$$|\partial_\theta|^2 = g(\partial_\theta,\partial_\theta) = f(z)^2,\;\; g(\dot\gamma,\partial_\theta) = \dot\theta f(z)^2.$$

Observe that ∂_θ is tangent to the parallels $z = const$. If we denote by $\varphi = \varphi(t)$ the angle between γ and the parallel that it intersects at time t we have

$$|\partial_\theta|\cdot|\dot\gamma|\cos\varphi = g(\dot\gamma,\partial_\theta) = \dot\theta f(z)^2$$

If we divide by the length of $\dot\gamma$ we deduce

$$r\cos\varphi = f(z)\cos\theta = |\partial_\theta|\cos\varphi = \frac{1}{L}\dot\theta f(z)^2 = f(z)^2\frac{d\theta}{ds} = const \tag{1.7}$$

so that the quantity $r\cos\varphi$ is constant along a geodesic. The last equality is known as *Clairaut's theorem*. It has the following geometric interpretation. The quantity r measures the distance from a point on the geodesic to the axis of revolution, while φ is the angle between the geodesic and the parallel that it intersects at a given moment.

Observe that if $\gamma(t)$ is a geodesic, i.e., a solution of (1.4), then so is the reparametrized curve $t\mapsto\gamma(ct)$, where c is a positive constant. Since the length of the velocity vector along a geodesic is constant, we may assume after a possible affine reparametrization that this length is 1, i.e., the geodesic is arclength parametrized.

If we define a meridian on S to be the intersection of the surface with a plane containing the axis of revolution, and a parallel to be the intersection of the surface with a horizontal plane $z = const$, we see that the meridians intersect the parallels orthogonally, and thus, along the meridians the quantity $r\cos\varphi$ is constant, equal to 0. This shows that the meridians are geodesics on a surface of revolutions.

The case of the round sphere is very special. A sphere admits infinitely many axes of symmetry. If $\gamma(t)$ is a geodesic on the sphere parametrized by arclength, and p_0 is a point on the geodesic, then there exists a unique plane through the center of the sphere and p_0 that is orthogonal to $\gamma(t)$ at p_0. Choose as axis of symmetry the diameter orthogonal to this plane. With respect to this axis of symmetry the quantity $r\cos\varphi$ must be zero along this geodesic, so that this geodesic must be an arc of a meridian on the sphere.

□

Remark 1.6. The above discussion has avoided one important issue: how do we decide if an extremal for a given Lagrangian is indeed a least action path? This investigation has been carried out in many instances, including the case of geodesics on manifolds. This has lead to many important developments in Riemannian geometry and topology; see e.g. [F, N].

In particular, one can show that on a compact Riemann manifold, for any pair of points, there exists a minimal geodesic connecting them. If moreover, the points are not too far apart, then this minimal geodesic is unique. The uniqueness does not extend to relatively far apart points. For example there exist infinitely many minimal geodesics on the round sphere connecting two opposite poles. □

Consider again the Lagrangian L in Example 1.4

$$L(\boldsymbol{x},\boldsymbol{v}) = \frac{1}{2}m|\boldsymbol{v}|^2 - U(x).$$

Following the terminology of classical mechanics we will refer to the quantities

$$p_i : T\mathbb{R}^3 \to \mathbb{R},\ \ p_i = mv^i$$

as the *momenta*.

$$H : T\mathbb{R}^3 \to \mathbb{R},\ \ H(\boldsymbol{x},\boldsymbol{v}) = \frac{1}{2}m|\boldsymbol{v}|^2 + U(x)$$

is called the *total energy* = kinetic energy + potential energy. Observe that

$$p_i = \frac{\partial L}{\partial v^i},\ \ H = \Bigl(\sum_{i=1}^3 v^i p_i\Bigr) - L.$$

This justifies the following terminology.

Definition 1.7. Suppose $L : TE \to \mathbb{R}$ is a Lagrangian function on the vector space $E = \mathbb{R}^n$. The *momenta* of L are the functions

$$p_i = p_i(\boldsymbol{x},\boldsymbol{v}) = \frac{\partial L}{\partial v^i},\ \ i = 1,\dots,n.$$

The *total* energy of L is the function

$$H_L = H_L(\boldsymbol{x},\boldsymbol{v}) = \Bigl(\sum_{i=1}^n v^i p_i\Bigr) - L.$$ □

Observe that the Euler-Lagrange equations can be rewritten as

$$\dot{p}_i = \frac{\partial L}{\partial x^i},\ \ i = 1,\dots,n. \tag{1.8}$$

Theorem 1.8 (Conservation of Energy)**.** *Consider a Lagrangian L on the space $E = \mathbb{R}^n$. Then the total energy of L is conserved along any extremal of L. In other words, if $t \mapsto \boldsymbol{x}(t) \in E$ is an extremal of L then the function*

$$t \mapsto H_L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)\,)$$

is independent of t. □

Proof. We set $H(t) = H_L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)\,)$. We have to prove that $\dot{H} = 0$. We have

$$\dot{H} = \frac{d}{dt}\Bigl(\sum_{i=1}^n v^i p_i\Bigr) - \frac{d}{dt}L(\boldsymbol{x}(t),\dot{\boldsymbol{x}}(t)\,) = \sum_{i=1}^n\Bigl(\dot{v}^i p_i + v^i\dot{p}^i\Bigr) - \sum_{i=1}^n\Bigl(\frac{\partial L}{\partial x^i}\dot{x}^i + \frac{\partial L}{\partial v^i}\ddot{x}^i\Bigr)$$

($v^i = \dot{x}^i$ along the extremal)

$$= \sum_{i=1}^n v^i\Big(\dot{p}^i - \frac{\partial L}{\partial x^i}\Big) + \sum_{i=1}^n \dot{v}^i\Big(p_i - \frac{\partial L}{\partial v^i}\Big).$$

The first sum is zero due to the Euler-Lagrange equations (1.8), while the second sum is zero due to the definition of the momenta p_i. □

Theorem 1.9 (Conservation of Momentum)**.** *Suppose L is a Lagrangian on $\mathbb{R}^n$ such that for some $i = 1, \dots m$*

$$\frac{\partial L}{\partial x^i} = 0.$$

Then the momentum p_i is conserved along any extremal of L.

Proof. This is an immediate consequence of (1.8). □

We want to conclude this section with a more in depth look at the two conservation laws we presented above.

Suppose L is a Lagrangian on the vector space E, $L : TE \to \mathbb{R}$. Following the terminology dear to physicists, we define a *symmetry* of E to be a 1-parameter group of diffeomorphisms of E,

$$\Phi : \mathbb{R} \times E \to E,\ \ \mathbb{R} \times E \ni (t, \boldsymbol{x}) \mapsto \Phi^t(\boldsymbol{x}) \in E,\ \ \Phi^0 = \mathbb{1}_E,\ \ \Phi^t \circ \Phi^s = \Phi^{t+s}.$$

This defines a vector field on E,

$$\boldsymbol{F} : E \to E,\ \ \boldsymbol{F}(\boldsymbol{x}) := \frac{d}{dt}|_{t=0}\Phi^t(\boldsymbol{x}).$$

We denote by $F^1, \dots, F^n$ the components of $\boldsymbol{F}$,

$$\boldsymbol{F}(\boldsymbol{x}) = \big(\, F^1(\boldsymbol{x}), \dots, F^n(\boldsymbol{x})\,\big).$$

The symmetry Φ^t of E defines a symmetry $\hat{\Phi}^t$ of TE

$$\hat{\Phi}^t(\boldsymbol{x}, \boldsymbol{v}) = \big(\, \Phi^t(\boldsymbol{x}), D_{\boldsymbol{x}}\Phi^t(\boldsymbol{v})\,\big),$$

where $D_{\boldsymbol{x}}G$ denotes the differential at $\boldsymbol{x}$ of a smooth map $G : E \to E$ defined by

$$D_{\boldsymbol{x}}G(\boldsymbol{v}) = \frac{d}{ds}|_{s=0}G(\boldsymbol{x} + s\boldsymbol{v}).$$

We say that the Lagrangian L is invariant under the symmetry Φ^t if

$$L \circ \hat{\Phi}^t = L,\ \ \forall t \in \mathbb{R}.$$

This is equivalent to the condition

$$\frac{d}{dt}L(\hat{\Phi}^t(\boldsymbol{x}, \boldsymbol{v})) = 0,\ \ \forall t \in \mathbb{R},\ \ \boldsymbol{x}, \boldsymbol{v} \in \mathbb{R},$$

which further translates into

$$\sum_{i=1}^n F^i \frac{\partial L}{\partial x^i} + \sum_{j,k=1}^n v^j p_k \frac{\partial F^k}{\partial x^j} = 0. \tag{1.9}$$

Example 1.10. Consider for example the symmetry

$$\Phi^t(x^1, \dots, x^n) = (x^1 + t, x^2, \dots, x^n).$$

The associated vector field is

$$\boldsymbol{F} = (1, 0, \dots, 0),$$

and a Lagrangian is invariant with respect to this symmetry if an only if

$$\frac{\partial L}{\partial x^1} = 0,$$

i.e., L is independent of the variable x^1. □

Example 1.11. Consider the symmetry of $\mathbb{R}^3$

$$\Phi^t(x^1, x^2, x^3) = (x^1 \cos t - x^2 \sin t, x^1 \sin t + x^2 \cos t, x^3)$$

In other words, Φ^t is the counterclockwise rotation of angle t about the x^3-axis. Then

$$\boldsymbol{F} = (-x^2, x^1, 0),$$

i.e.,

$$F^1 = -x^2, \ \ F^2 = x^1, \ \ F^3 = 0.$$

A lagrangian L is invariant under this symmetry if

$$-x^2 \frac{\partial L}{\partial x^1} + x^1 \frac{\partial L}{\partial x^2} = v^2 \frac{\partial L}{\partial v^1} - v^1 \frac{\partial L}{\partial v^2}.$$

If we use cylindrical coordinates

$$x^1 = r \cos\theta, \ \ x^2 = r \sin\theta, \ \ x^3 = x^2.$$

then

$$v^1 = \dot{x}^1 = \dot{r} \cos\theta - (r \sin\theta)\dot{\theta}, \ \ v^2 = \dot{x}^2 = \dot{r} \sin\theta + (r \cos\theta)\dot{\theta}$$

and we can express a lagrangian $L(x^1, x^2, x^3, \dot{x}^1, \dot{x}^2, \dot{x}^3)$ in terms of $(r, \theta, x^3, \dot{r}, \dot{\theta}, \dot{x}^3)$.

In the cylicndrical coordinates the above symmetry has the simple expression

$$\Phi^t(r, \theta, x^3) = (r, \theta + t, x^3)$$

and we deduce that L is invariant under this symmetry if

$$\frac{\partial}{\partial \theta}(r, \theta, x^3, \dot{r}, \dot{\theta}, \dot{x}^3) = 0.$$

For example, if

$$L(\boldsymbol{x}, \boldsymbol{v}) = \frac{1}{2} m |v|^2 - U(\boldsymbol{x})$$

then

$$|\boldsymbol{v}|^2 = |\dot{x}^1|^2 + |\dot{x}^2|^2 + |\dot{x}^3|^2 = |\dot{r}|^2 + r^2 |\dot{\theta}|^2 + |\dot{x}^3|^2,$$

and

$$L = \frac{m}{2}(|\dot{r}|^2 + r^2 |\dot{\theta}|^2 + |\dot{x}^3|^2)^2 - U(r, \theta, x^3).$$

Thus L is invariant under this symmetry if and only if $\frac{\partial U}{\partial \theta} = 0$, i.e., the function U is roationally symmetric with respect to the rotations about the axis x^3. □

To any symmetry Φ^t of E with associated vector field $\boldsymbol{F}$, and any Lagrangian L on E we associate the *generalized momentum*

$$p_\Phi := \sum_{i=1}^n p_i F^i.$$

Theorem 1.12 (Emy Noether's Conservation Principle). *If the Lagrangian L is invariant under the symmetry Φ, then the generalized momentum p_Φ is constant along any extremal of L.* □

The proof is a simple application of the Euler-Lagrange equations and the invariance condition (1.9). For more details an applications we refer to [A].

2. The direct method in the calculus of variations

In the previous section we investigated functions depending on a single variable that are extrema of some action functionals. We can define such action functionals for functions of several variables but the resulting Euler-Lagrange equations are nonlinear, second order *partial* differential equations and they are much more difficult to deal with.

Example 2.1 (Maxwell's equations)**.** For example, consider the Space-Time $\mathbb{R}^{1,3}$ with coordinates (t, x, y, z). To an electromagnetic field in vacuum we can non-uniquely associate a (scalar) electric potential U and a (vector) magnetic potential $\boldsymbol{A} = A_x\boldsymbol{i} + A_y\boldsymbol{i} + A_z\boldsymbol{k}$. These can be collected in a single 1-form

$$\omega = Udt + A_x dx + A_y dy + A_z dz \in \Omega^1(\mathbb{R}^4).$$

Its exterior differential $F = d\omega$ has a decomposition

$$F = dt \wedge (E_x dx + E_y dy + E_z dz) + B_x dy \wedge dz + B_y dz \wedge dy + B_z dx \wedge dy,$$

where

$$\boldsymbol{E} = E_x\boldsymbol{i} + E_y\boldsymbol{i} + E_z\boldsymbol{k} = -\nabla U$$

is the electric field, and

$$\boldsymbol{B} = B_x\boldsymbol{i} + B_y\boldsymbol{i} + B_z\boldsymbol{k} = \nabla \times \boldsymbol{A}$$

is the magnetic field. The Minkowski energy (density) of F is

$$\mathcal{E}(F) := |\boldsymbol{E}|^2 - |\boldsymbol{B}|^2 = |E_x|^2 + |E_y|^2 + |E_z|^2 - (|B_x|^2 + |B_y|^2 + |B_z|^2).$$

One half of Maxwell's four equations amount to saying that $dF = 0$. More precisely

$$dF = 0 \Longleftrightarrow \frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times \boldsymbol{E}, \ \ \operatorname{div} \boldsymbol{B} = 0.$$

The other half,

$$\frac{\partial \boldsymbol{E}}{\partial t} = \nabla \times \boldsymbol{B}, \ \ \nabla \times \boldsymbol{E} = 0,$$

states that F is an extremal of the action functional

$$\mathcal{S}(\omega) = \int_{\mathbb{R}^4} \mathcal{E}(F) dt\, dx\, dy\, dz.$$ □

Calculus of variations is often used in reverse. For example, if by some means we can conclude that an action functional has minima, we can then conclude that the corresponding Euler-Lagrange equations do have solutions. Fortunately we have such a very general existence principle.

Theorem 2.2 (Fundamental Existence Theorem)**.** *Suppose H is a separable real Hilbert space and $f : H \to \mathbb{R}$ is a* <u>*bounded from below*</u> *function satisfying the following additional conditions.*

- *The function f is* <u>*convex*</u>*, i.e., for any $c \in \mathbb{R}$ the sublevel set $\{f \leq c\}$ is a convex subset of H.*

- *The function f is* lower semicontinuous*, i.e., for any real number c the sublevel set $f \leq c\}$ is a subset of H closed with respect to the norm topology.*
- *The function f is* coercive*, i.e., for any real number c the sublevel set $\{f \leq c\}$ is a bounded subset of H.*

Then there exists $x_0 \in H$ such that

$$f(x_0) \leq f(x), \;\; \forall x \in H. \qquad \square$$

The proof of this theorem is an application of some fundamental principles of functional analysis and we refer to [B, Chap. III] for more details.

Let us explain how to use the above existence theorem to solve a famous problem in geometry, namely the uniformization problem for Riemann surfaces. For details we refer to [N, §10.3.3]. Suppose Σ is a compact, oriented surface of genus ≥ 2 so that the Euler characteristic is negative,

$$\chi(\Sigma) < 0.$$

One possible formulation of the uniformization problem is the following.

For any Riemann metric g on Σ there exists a unique smooth function u on Σ such that the scalar curvature of the metric $g_u = e^u g$ is equal to -1.

A direct computation shows that such a function u must be a solution of the partial differential equation

$$\Delta_g u + e^u = -s(x), \tag{2.1}$$

where $s(x)$ is the scalar curvature[2] of the metric g, and $\Delta_g : C^\infty(\Sigma) \to C^\infty(\Sigma)$ is the Laplace operator determined by the metric g. We can assume without loss of generality that the volume of the initial metric g is 1. Note that the Gauss-Bonnet theorem implies that

$$\int_\Sigma s(x) = 4\pi\chi(\Sigma) < 0. \tag{2.2}$$

Let us look at a more general equation

$$\Delta_g u + f(u) = h(x), \tag{$E_{f,h}$}$$

where $f : \mathbb{R} \to \mathbb{R}$ and $h : \Sigma \to \mathbb{R}$ are C^1-functions. Define the action functional

$$\mathcal{S} = \mathcal{S}_{f,h} : C^1(\Sigma) \to \mathbb{R}, \;\; \mathcal{S}(u) = \int_\Sigma \Big(\frac{1}{2}|du|_g^2 + F(u) - hu \Big) dV_g,$$

where F is an antiderivative of f. Arguing as in the proof of the Least Action Principle we deduce that if $u \in C^1(\Sigma)$ is a minimum of $\mathcal{S}$,

$$\mathcal{S}(u) \leq \mathcal{S}(v), \;\; \forall v \in C^1(\Sigma),$$

then u satisfies

$$\int_M \Big(g(\nabla u, \nabla\varphi) + (f(u) - h)\varphi \Big) dV_g = 0, \;\; \forall \varphi \in C^1(\Sigma). \tag{2.3}$$

If additionally u is twice continuously differentiable, then an integration by parts as in the proof of the Least Action principle shows that u satisfies ($E_{f,h}$).

To apply our existence principle we need first of all a convex functional. The functional $\mathcal{S}_{f,h}$ is convex if and only if $F(u)$ is convex, i.e., the function $f(u)$ is increasing. This is certainly the case if $f(u) = e^u$. In this situation we choose $F(u) = e^u$. Thus, in the sequel we will assume that $f(u)$

[2] A word of warning: the scalar curvature of a Riemann surface is *twice* its sectional curvature, whence the constant 4π in the Gauss-Bonnet formula instead of the usual 2π.

is increasing. The monotonicity of f has an added bonus. More precisely we have the following comparison principle.

Theorem 2.3 (Comparison Principle)**.** [3] *If $u, v \in C^2(\Sigma)$ satisfy*

$$(\Delta_g u)(x) + f(u(x)) \leq (\Delta_g v)(x) + f(v(x)), \quad \forall x \in \Sigma,$$

where f is a strictly increasing C^1-function, then $u(x) \leq v(x)$, $\forall x \in \Sigma$. In particular, the equation ($E_{f,h}$) admits at most one solution. □

To apply the Fundamental Existence Theorem we need a Hilbert space. The correct Hilbert space for our problem is the Sobolev space $L^{1,2}(\Sigma)$ defined as the closure of $C^1(\Sigma)$ in $L^2(\Sigma)$ with respect to the norm

$$\|u\|_{1,2} := \int_\Sigma (|du|_g^2 + u^2) dV_g.$$

Unfortunately, for functions in $u \in L^{1,2}(\Sigma)$ the integral $\int_\Sigma F(u)dV_g$ could be infinite. For certain functions f there is a way out of this quandary.

For every $r \geq 0$ we denote by f_r the unique C^1-function $f_r : \mathbb{R} \to \mathbb{R}$ such that f_r is linear on the interval $[r, \infty)$ while $f_r(u) = e^u$, $\forall u \leq r$; see Figure 3. More explicitly

$$f_r(u) = \begin{cases} e^u, & u \leq r \\ e^r(u - r + 1), & u \geq r. \end{cases}$$

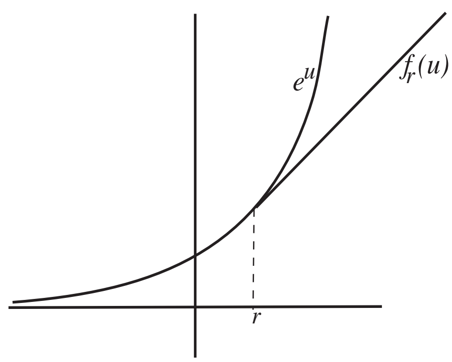

FIGURE 3. *The graphs of e^u and $f_r(u)$*

Since the function e^u is convex, we deduce

$$1 + u \leq f_r(u) \leq e^u, \quad \forall u \in \mathbb{R}. \tag{2.4}$$

Denote by F_r the antiderivative of f_r such that $F_r(0) = e^0 = 1$. Then one can prove that there exists a constant $C_r > 0$ such that

$$|F_r(u)| \leq C_r(|u|^2 + 1), \quad \forall u \in \mathbb{R}.$$

This proves that

$$\int_\Sigma F_r(u)dV_g < \infty, \quad \forall u \in L^{1,2}(\Sigma).$$

We set $\mathcal{S}_{r,h} := \mathcal{S}_{f_r,h}$. We have the following result.

[3]The comparison principle in the above form was first used by *Guido Stampacchia*. We refer to [N, §10.3.3] for a proof which relies on some basic facts about Sobolev spaces.

Theorem 2.4. *Suppose that*

$$\bar{h} := \int_\Sigma h dV_g > 0. \tag{2.5}$$

Then for any $r > \log \bar{h}$ the functional $\mathcal{S}_{r,h} : L^{1,2}(\Sigma) \to \mathbb{R}$ is convex, bounded from below, lower semicountinuous and coercive[4] *and thus it has minima. Moreover, any function $u \in L^{1,2}(\Sigma)$ that minimizes $\mathcal{S}_{r,h}$ is in fact twice differentiable and satisfies the differential equation*

$$\Delta_g u + f_r(u) = h(x) \tag{2.6}$$

□

The proof of this theorem is quite involved and we refer to [N, §10.3.3] for details.

Denote by u_r the unique solution of the equation (2.6). We set $C = \max_{x\in\Sigma} |h(x)|$ and let K be a positive number such that $1 + K > C$. We view K as a constant function on Σ and we deduce

$$\Delta_g K + f_r(K) = f_r(K) \stackrel{(2.4)}{\geq} 1 + K \geq h(x) = \Delta_g u_r + f_r(u_r)$$

From the comparison principle we deduce that $u_r \leq K$. Thus for $r \geq K$ we have

$$f_r(u_r) = e^{u_r}.$$

Hence, if $r > \max(K, \log \bar{h})$ the function u_r is actually a solution of

$$\Delta_g u_r + e^{u_r} = h(x). \tag{2.7}$$

We have thus obtained the following result.

Corollary 2.5. *If h is a smooth function on Σ whose average is positive,*

$$\int_\Sigma h dV_g > 0,$$

then the equation

$$\Delta_g u + e^u = h(x)$$

has a unique smooth solution. □

From (2.2) we deduce that the above corollary is applicable to the function $h(x) = -s(x)$ and we conclude that the equation (2.7) has a unique solution.

References

[A] V.I. Arnold: *The Mathematical Methods of Classical Mechanics*, Graduate Texts in Math., vol. 60, Springer Verlag, 1989.

[B] H. Brézis: *Analyse Fonctionelle. Théorie et Applications*, Dunod, 1999.

[GF] I.M. Gelfand, S.V. Fomin: *Calculus of variations*, Dover Publishing, 2000.

[FLS] R.P. Feynman, R.E. Leighton, M. Sands: *The Feynman Lectures on Physics*, any edition.

[F] T. Frankel: *The Geometry of Physics. An Introduction*, Cambridge University Press,

[N] L.I. Nicolaescu: *Lectures on the geometry of manifolds*, World Scientific, 2007.

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556-4618.
E-mail address: nicolaescu.1@nd.edu
URL: http://www.nd.edu/~lnicolae/

[4]The condition $r > \log \bar{h}$ is needed precisely to ensure coercivity. It will fail without it.