Some Notes on Classical Mechanics

Alberto S. Cattaneo

Institute of Mathematics, University of Zurich

Last update: November 30, 2021

Abstract. This is a collection of some random notes on topics related to classical mechanics that I prepared for classes of the same title.

Contents

CHAPTER 1

Introduction

Mechanical problems are described using different mathematical formalisms: notably, Newton's equations, Lagrangian mechanics, Hamiltonian mechanics.

1.1. Newton's equations

This is the most basic but also most general way of describing a mechanical problem. The underlying mathematical formalism is just the theory of ODEs. The experimental fact that initial position and velocity are enough to determine the motion is encoded in the usage of second-order ODEs. Newton's equations are used in a variety of mechanical problems as they allow for all kinds of forces including, e.g., friction. On the other hand, they lack in general methods for finding and describing solutions.

1.2. Lagrangian mechanics

Based on variational calculus, Lagrangian mechanics is available only for certain mechanical problems, which however includes all fundamental problems. Lagrangian mechanics makes the change of coordinates and the study of constrained systems routine. Via Noether's theorem it connects configuration space symmetries to conservation laws. The Lagrangian formalism is also at the basis of Feynman's approach to quantum mechanics via path integrals.

1.3. Hamiltonian mechanics

It is the most flexible way of studying mechanical systems as one has symplectomorphisms at one's disposal. Noether's theorem becomes easier and at the same time more powerful. The possibility of looking for coordinates in which the system gets easier is of tremendous importance in the applications. The Hamiltonian formalism is also at the basis of the operator approach to quantum mechanics. With respect to Lagrangian mechanics, Hamiltonian mechanics has the drawback of not being directly available for systems in which the Lagrangian is degenerate.

1.4. Field theory

The higher dimensional generalization of mechanics is called field theory and is based on PDEs. The unknown functions are now defined on a domain in $\mathbb{R}^n$ (or on an n-dimensional manifold). Lagrangian and Hamiltonian methods are available. The Hamiltonian formalism has one further drawback in that it makes an explicit choice of the time direction. As a consequence, fundamental field theories, where special relativity is also taken into account, are usually described in the Lagrangian formalism. Theories with degenerate Lagrangian (e.g. electrodymanics) are also quite common, so the Legendre transform is not directly available.

CHAPTER 2

Symplectic Integrators

The Hamiltonian flow preserves the symplectic structure; an approximation, such as for numerical integration, in general does not, which can be expected to miss some important features of the system under study. In this note, we briefly recall how to improve the Euler method in such a way that the symplectic structure is preserved.

2.1. The Euler method

Let X be a vector field on an open subset O of $\mathbb{R}^n$ and $\phi_{X,t}$ its flow (at time t). Recall that $\phi_{X,t}(x_0)$, $x_0 \in U$, is by definition the evaluation at time t of the unique solution to the Cauchy problem

$$\begin{cases} \dot{x}(t) = X(x), \\ x(0) = x_0. \end{cases} \tag{2.1}$$

The time t must belong to the maximal interval of definition of the solution, which in turn in general depends on x_0. By uniqueness we have the fundamental property

$$\phi_{X,t} \circ \phi_{X,s} = \phi_{X,t+s},$$

for t, s and $t+s$ in the maximal interval. This implies that, for t in the maximal interval and any integer N, we have

$$\phi_{X,t} = \underbrace{\phi_{X,\frac{t}{N}} \circ \cdots \circ \phi_{X,\frac{t}{N}}}_{N \text{ times}}.$$

The Euler method consists in approximating $\phi_{X,\tau}$, $\tau = \frac{t}{N}$, by a truncation of its Taylor expansion in τ. Let us work it out up to $O(\tau^2)$. A path $x(t)$ may be expanded around $t = 0$ as

$$x(\tau) = x(0) + \tau\dot{x}(0) + O(\tau^2).$$

If it is the solution to (2.1), we then have

$$x(\tau) = x_0 + \tau X(x_0) + O(\tau^2).$$

This yields

$$\phi_{X,\tau}(x_0) = x_0 + \tau X(x_0) + O(\tau^2). \tag{2.2}$$

The Euler method, at this order, consists in replacing $\phi_{X,\tau}$ by its truncation

$$\widetilde{\phi}_{X,\tau}(x_0) := x_0 + \tau X(x_0),$$

getting the approximate solution

$$\phi^{\text{Euler}}_{X,t} := \underbrace{\widetilde{\phi}_{X,\frac{t}{N}} \circ \cdots \circ \widetilde{\phi}_{X,\frac{t}{N}}}_{N \text{ times}}.$$

2.2. Hamiltonian systems

Let H be a Hamiltonian function (on $W \times \mathbb{R}^d$ with W an open subset of $\mathbb{R}^d$) and X_H its Hamiltonian vector field. We want to compute its flow $\phi_{X_H,t}$. (To match the notation with that of other chapters, we set $n = 2d$ and $O = W \times \mathbb{R}^d$.) The problem is that in general the truncation $\widetilde{\phi}_{X_H,\frac{t}{N}}$ does not preserve the symplectic form, nor does so the ensuing Euler approximation $\phi^{\text{Euler}}_{X_H,t}$.

The idea of a symplectic integrator, in this context, consists in choosing a different approximation of $\phi_{X_H,\tau}$ that is equal to $\widetilde{\phi}_{X_H,\tau}$ up to $O(\tau^2)$ but preserves the symplectic form.

Suppose that $H = H_1 + H_2$ (we will see below that this is interesting for practical purposes if we can compute the Hamiltonian flows of H_1 and H_2 exactly). We then have $X_H = X_{H_1} + X_{H_2}$. By (2.2) we have

$$\begin{aligned}\phi_{X_{H_1},\tau}(x_0) &= x_0 + \tau X_{H_1}(x_0) + O(\tau^2),\\ \phi_{X_{H_2},\tau}(x_0) &= x_0 + \tau X_{H_2}(x_0) + O(\tau^2).\end{aligned}$$

This implies

$$\begin{aligned}(\phi_{X_{H_1},\tau} \circ \phi_{X_{H_2},\tau})(x_0) &= x_0 + \tau(X_{H_1}(x_0) + X_{H_2}(x_0)) + O(\tau^2)\\ &= x_0 + \tau X_H(x_0) + O(\tau^2),\end{aligned}$$

so

$$\phi_{X_H,\tau} = \phi_{X_{H_1},\tau} \circ \phi_{X_{H_2},\tau} + O(\tau^2).$$

The idea is now to replace $\phi_{X_H,\tau}$ by

$$\widehat{\phi}_{X_H,\tau} := \phi_{X_{H_1},\tau} \circ \phi_{X_{H_2},\tau} \tag{2.3}$$

getting the approximate solution

$$\boxed{\phi^{\text{SI}}_{X_H,t} := \underbrace{\widehat{\phi}_{X_H,\frac{t}{N}} \circ \cdots \circ \widehat{\phi}_{X_H,\frac{t}{N}}}_{N \text{ times}}} \tag{2.4}$$

Notice that $\phi_{X_{H_1},\tau}$ and $\phi_{X_{H_2},\tau}$ preserve the symplectic structure and hence so does $\phi^{\text{SI}}_{X_H,t}$.

The method is applicable if we can compute $\phi_{X_{H_1},\tau}$ and $\phi_{X_{H_2},\tau}$ exactly. The typical case is that of a Hamiltonian of the form $H(q,p) = T(p)+U(q)$. In this case, we set $H_1(q,p) = T(p)$ and $H_2(q,p) = U(q)$.[1] To be even more specific (even though this is not needed), let us assume that $T(p) = \frac{||p||^2}{2m}$. Let us now compute the corresponding flows. In the case of H_1 we have to solve

$$\begin{aligned}\dot{q} &= \frac{p}{m},\\ \dot{p} &= 0,\end{aligned}$$

with initial conditions at $t = 0$ given by q_0 and p_0. This yields

$$q(\tau) = q_0 + \tau\,\frac{p_0}{m}, \qquad p(\tau) = p_0.$$

Hence

$$\phi_{X_{H_1},\tau}(q_0,p_0) = \left(q_0 + \tau\,\frac{p_0}{m}, p_0\right).$$

In the case of H_2 we have to solve

$$\begin{aligned}\dot{q} &= 0,\\ \dot{p} &= -\nabla U(q),\end{aligned}$$

with the same initial conditions. This yields

$$q(\tau) = q_0, \qquad p(\tau) = p_0 - \tau\nabla U(q_0).$$

Therefore,

$$\phi_{X_{H_2},\tau}(q_0,p_0) = (q_0, p_0 - \tau\nabla U(q_0)).$$

We finally have by (2.3) that

$$\boxed{\widehat{\phi}_{X_H,\tau}(q_0,p_0) = \left(q_0 + \tau\,\frac{p_0 - \tau\nabla U(q_0)}{m}, p_0 - \tau\nabla U(q_0)\right)} \tag{2.5}$$

Notice that this agrees up to $O(\tau^2)$ with the first-order Taylor approximation

$$\widetilde{\phi}_{X_H,\tau}(q_0,p_0) = \left(q_0 + \tau\,\frac{p_0}{m}, p_0 - \tau\nabla U(q_0)\right)$$

but is different from the second-order Taylor approximation of $\phi_{X_H,\tau}$ (which in general does not preserve the symplectic structure).

If we had chosen instead $H_1(q,p) = U(q)$ and $H_2(q,p) = T(p) = \frac{||p||^2}{2m}$, we would have got

$$\widehat{\phi}_{X_H,\tau}(q_0,p_0) = \left(q_0 + \tau\,\frac{p_0}{m}, p_0 - \tau\nabla U\left(q_0 + \tau\,\frac{p_0}{m}, p_0\right)\right),$$

[1]Observe that the Hamiltonian H_2 is not hyperregular as a function of p, so it certainly does not arise as the Legendre transform of a Lagrangian. This shows one more reason why the Hamiltonian formalism is often preferable.

which yields a different approximation that also preserves the symplectic structure.

CHAPTER 3

The Noether Theorem

A symmetry is an invertible transformation that preserves some properties. In geometry, e.g., one considers transformations that preserve "shape" (in Euclidean geometry this leads to translations, rotations and reflections as symmetries). In mechanical systems, symmetries are transformations that preserve the equations of motions. One also considers stricter symmetries that preserve more, e.g., the action functional.

Symmetry transformations often arise as flows of vector fields. The latter are then called infinitesimal symmetries. A fundamental result in mechanics is Noether's theorem which relates infinitesimal symmetries to constants of motion.

3.1. Symmetries in Lagrangian mechanics

Let J_L be the action functional associated to a Lagrangian function on $U \times \mathbb{R}^d \times I$:

$$J_L[x] = \int_a^b L(x(t), \dot{x}(t), t)\, \mathrm{d}t,$$

where x is a path $[a,b] \to U$, $[a,b] \subset I$. A symmetry, in a strict sense, is a diffeomorphism ϕ of U such that

$$J_L[\phi \circ x] = J_L[x] \tag{3.1}$$

for each path x on U. Notice that in particular a symmetry maps extremal paths (with end points x_a, x_b) to extremal paths (with end points $\phi(x_a), \phi(x_b)$).

A vector field X on U is called an infinitesimal symmetry if its flow ϕ_s is a symmetry for all s (where it is defined). We then have

$$0 = \frac{\partial}{\partial s}\Big|_{s=0} J_L[\phi_s \circ x] = \frac{\delta J_L}{\delta x}[x, X],$$

where we regard the path $t \mapsto X(x(t))$ as a variation. Recall the general formula

$$\frac{\delta J_L}{\delta x}[x, \delta x] = EL_L[x, \delta x] + \left(\sum_{i=1}^d \frac{\partial L}{\partial v^i}\, \delta x^i \right) \Big|_a^b,$$

with

$$EL_L[x, \delta x] = \int_a^b \sum_{i=1}^d \left(\frac{\partial L}{\partial q^i}(x(t), \dot{x}(t), t) - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial v^i}(x(t), \dot{x}(t), t) \right) \delta x^i(t)\, \mathrm{d}t.$$

If X is an infinitesimal symmetry and x is an extremal path (i.e., it satisfies the Euler–Lagrange equations), then we get

$$0 = \left(\sum_{i=1}^d \frac{\partial L}{\partial v^i} X^i \right) \Big|_a^b.$$

This shows that the quantity in brackets is the same at the initial time a and at the final time b. Since the choice of the end times was irrelevant for the derivation, this shows that the quantity in brackets does not change and is therefore a constant of motion.[1] We can summarize this as follows:

Definition 3.1. The Noether 1-form associated to the Lagrangian L is the 1-form α_L on $U \times \mathbb{R}^d \times I$ defined by

$$\alpha_L := \sum_{i=1}^d \frac{\partial L}{\partial v^i}(q, v, t)\, \mathrm{d}q^i.$$

Theorem 3.2 (Noether's Theorem). *If X is an infinitesimal symmetry of J_L, then*

$$\boxed{I_X := \iota_X \alpha_L}$$

is a constant of motion.

3.1.1. Symmetries and the Lagrangian function. A symmetry of J_L may actually be recognized directly on L. Recall that we defined the tangent lift $\check{\phi}$ of ϕ by

$$\begin{aligned} \check{\phi}\colon\ U \times \mathbb{R}^d \times [a, b] &\to U \times \mathbb{R}^d \times [a, b] \\ (q, v, t) &\mapsto (\phi(q), \mathrm{d}\phi(q)v, t) \end{aligned}$$

Lemma 3.3. *A diffeomorphism ϕ of U is a symmetry of J_L, as expressed by equation* (3.1)*, if and only if $L \circ \check{\phi} = L$.*

Proof. Recall that we observed that for any map ϕ we have $J_L[\phi \circ x] = J_{L\circ\check{\phi}}[x]$. This immediately shows that ϕ is a symmetry if $L \circ \check{\phi} = L$.

On the other hand, assume that ϕ is a symmetry. Then, by the above observation, we have that $J_{L\circ\check{\phi}}[x] = J_L[x]$ for every path x. If we define $\widetilde{L} := L \circ \check{\phi} - L$, we then have $J_{\widetilde{L}}[x] = 0$ for every path x. We

[1]Recall that a function f on $U \times \mathbb{R}^d \times I$ is a called a constant of motion (a.k.a. conserved quantity or first integral) if $f(x(t), \dot{x}(t), t)$ is constant in t for every solution x of the Euler–Lagrange equations.

want to show that $\widetilde{L}$ vanishes identically. Assume on the contrary that there is a point (q, v, τ) in $U \times \mathbb{R}^d \times I$ such that $\widetilde{L}(q, v, \tau) \neq 0$. Then consider the path $x(t) = q + (t - \tau)v$ on the interval $[\tau, \tau + \epsilon]$. We then have

$$J_{\widetilde{L}}[x] = \int_{\tau}^{\tau+\epsilon} \widetilde{L}(q + (t - \tau)v, v, t)\, \mathrm{d}t = \widetilde{L}(q, v, \tau)\epsilon + O(\epsilon^2).$$

By choosing ϵ appropriately, we then see that $J_{\widetilde{L}}[x]$ cannot vanish for all paths, and this is a contradiction. □

We now want to move on to infinitesimal symmetries. First we need the following

LEMMA 3.4. *Let X be a vector field on U and ϕ_s its flow. Then the tangent lift $\check{\phi}_s$ is the flow of the vector field $\check{X}$ on $U \times \mathbb{R}^d \times I$ given by*

$$\check{X}(q, v, t) = \sum_{i=1}^{d} X^i(q) \frac{\partial}{\partial q^i} + \sum_{i,j=1}^{d} \frac{\partial X^i}{\partial q^j}(q)\, v^j\, \frac{\partial}{\partial v^i},$$

called the **tangent lift** *of X.*

Notice that the tangent lift of X defined above does not depend on t and does not have a component in the t direction, so it can be regarded (and this is usually done) as a vector field on $U \times \mathbb{R}^d$.

PROOF. From $\phi_{s+s'} = \phi_s \circ \phi_{s'}$ and $\phi_0 = \mathrm{id}$, we get, by the definition of the tangent lift, that $\check{\phi}_{s+s'} = \phi_s \circ \phi_{s'}$ and $\check{\phi}_0 = \mathrm{id}$. Therefore, $\check{\phi}_s$ is also a flow. To compute the corresponding vector field, we just have to derive $\check{\phi}_s$ with respect to s at $s = 0$, and this gives the explicit expression for $\check{X}$ in the lemma. □

We now have the

COROLLARY 3.5. *A vector field X is an infinitesimal symmetry of J_L if and only if*

$$\boxed{\check{X}(L) = 0}$$

3.1.2. Examples.

EXAMPLE 3.6. Suppose that the system is invariant under translations in the $i$the direction, i.e., $X^i = \frac{\partial}{\partial q^i}$ is an infinitesimal symmetry. This implies that the ith component of the generalized momentum,

$$p_i = \frac{\partial L}{\partial v^i} = \iota_{X^i} \alpha_L,$$

is a constant of motion. Since $\check{X}^i = \frac{\partial}{\partial q^i}$, we see that this occurs if and only if $\frac{\partial L}{\partial q^i} = 0$.

EXAMPLE 3.7. Suppose that $d = 3$ and the system is invariant under rotations around the ith axis. In this case, the infinitesimal symmetry is given by the vector field $X^i(q) = \sum_{j,k=1}^{3} \epsilon_{ijk}\, q^j \frac{\partial}{\partial q^k}$. The corresponding constant of motion is

$$J^i = \iota_{X^i}\alpha_L = \sum_{j,k=1}^{3} \epsilon_{ijk}\, q^j \frac{\partial L}{\partial v^k}$$

which is called the (generalized) angular momentum. If we denote $\frac{\partial L}{\partial v^k}$ by p_k, we then have $J^i = (\boldsymbol{q} \times \boldsymbol{p})^i$. One can easily compute $\check{X} = \sum_{j,k=1}^{3} \epsilon_{ijk}\, q^j \frac{\partial}{\partial q^k} + \sum_{j,k=1}^{3} \epsilon_{ijk}\, v^j \frac{\partial}{\partial v^k}$. This describes an infinitesimal rotation acting simultaneously on the q-space and on the v-space. A Lagrangian L of the form $L(\boldsymbol{q}, \boldsymbol{v}, t) = T(\boldsymbol{v}) - U(\boldsymbol{q})$ is then invariant if both functions T and U are invariant under rotations (around the ith axis). If T and U are invariant under the whole group $SO(3)$, then all components of the vector $\boldsymbol{J} = \boldsymbol{q} \times \boldsymbol{p}$ are conserved. Finally, if $T(\boldsymbol{v}) = \frac{1}{2}m||\boldsymbol{v}||^2$, we then have $\boldsymbol{J} = \boldsymbol{q} \times m\boldsymbol{v}$ which is the usual expression for the angular momentum.

3.1.3. Generalized symmetries. Condition (3.1) is a bit too strong since it involves the action functional also away from extremal paths. One way to weaken it, in such a way that a symmetry still preserves the Euler–Lagrange equations, is to assume that the restriction of

$$J_{L,\phi}[x] := J_L[\phi \circ x] - J_L[x]$$

on each path space $\mathcal{P}_{x_a,x_b}^{[a,b]} := \{x \colon [a,b] \to U : x(a) = x_a,\ x(b) = x_b\}$ is constant. This can also be characterized as in the following

LEMMA 3.8. *Assume that the restriction of $J_{L,\phi}$ on each $\mathcal{P}_{x_a,x_b}^{[a,b]}$ is constant. Then there is a function F on U such that*

$$J_{L,\phi}[x] = F(x_b) - F(x_a)$$

for each $x \in \mathcal{P}_{x_a,x_b}^{[a,b]}$ for each a, b, x_a, x_b.

REMARK 3.9. Notice that the function F is defined up to an additive constant on each connected component of U.

PROOF. Let us first consider the case when U is connected. Fix a point y in U. Then $J_{L,\phi}$ will take the same value on all paths in $\mathcal{P}_{y,x_b}^{[a,b]}$. We then define $F(x_b)$ as such value.

Next we want to show that $F(y) = 0$. For $x_b = y$, we also have the constant path $x(t) = y\ \forall t \in [a,b]$ at our disposal. This can be considered for arbitrary a and b. Since $J_{L,\phi}$ is defined as an integral, its value vanishes if b tends to a. So $F(y) = 0$.

Consider now a path x in $\mathcal{P}^{[a,b]}_{x_a,y}$ with the property that all its derivatives at $t = a$ vanish. Next define $\underline{x} \in \mathcal{P}^{[a,b]}_{y,x_a}$ by $\underline{x}(t) = x(a + b - t)$. Finally, define $\widetilde{x} \in \mathcal{P}^{[a,b]}_{y,y}$ by

$$\widetilde{x}(t) = \begin{cases} \underline{x}(2t - a) & \text{if t} \in \left[a, \frac{a+b}{2}\right] \\ x(2t - b) & \text{if } t \in \left[\frac{a+b}{2}, b\right] \end{cases}$$

Notice that the vanishing condition on the derivatives of x at its initial point makes this path smooth. We have $J_{L,\phi}[\widetilde{x}] = F(y) = 0$. On the other hand, since $J_{L,\phi}$ is defined as an integral, we have $J_{L,\phi}[\widetilde{x}] = J_{L,\phi}[\underline{x}] + J_{L,\phi}[x]$. Therefore, $J_{L,\phi}[x] = -J_{L,\phi}[\underline{x}] = -F(x_a)$, since $\underline{x} \in \mathcal{P}^{[a,b]}_{y,x_a}$. By the constancy of $J_{L,\phi}$, we conclude that $J_{L,\phi}[x] = -F(x_a)$ for every $x \in \mathcal{P}^{[a,b]}_{x_a,y}$

Finally, consider a path x in $\mathcal{P}^{[a,b]}_{x_a,x_b}$ that passes through y at some time $\tau \in (a, b)$. Denote by $\widetilde{x}$ the restriction of this path to the interval $[a, \tau]$ and by $\bar{x}$ its restriction to $[\tau, b]$. Since $J_{L,\phi}$ is defined as an integral, we have $J_{L,\phi}[x] = J_{L,\phi}[\widetilde{x}] + J_{L,\phi}[\bar{x}]$. Since $\bar{x} \in \mathcal{P}^{[\tau,b]}_{y,x_b}$, we know that $J_{L,\phi}[\bar{x}] = F(x_b)$ and, since $\widetilde{x} \in \mathcal{P}^{[a,\tau]}_{x_a,y}$, we know that $J_{L,\phi}[\widetilde{x}] = -F(x_a)$. We conclude that $J_{L,\phi}[x] = F(x_b) - F(x_a)$.

If U is not connected, we just apply the above procedure to each connected component. □

We then define a **generalized symmetry** as a pair (ϕ, F), where ϕ is a diffeomorphism of U and F is a function on U, such that

$$J_L[\phi \circ x] = J_L[x] + F(x_b) - F(x_a)$$

for each $x \in \mathcal{P}^{[a,b]}_{x_a,x_b}$ for each a, b, x_a, x_b. Notice that we can write

$$F(x_b) - F(x_a) = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t} F(x(t))\, \mathrm{d}t = \int_a^b \sum_{i=1}^d \dot{x}^i(t)\, \frac{\partial F}{\partial q^i}(x(t))\, \mathrm{d}t.$$

If we define $\widetilde{F}(q, v) := \sum_{i=1}^d v^i \frac{\partial F}{\partial q^i}(q)$, then, by repeating the arguments in the proof of Lemma 3.3, we conclude that (ϕ, F) is a generalized symmetry if and only if

$$L \circ \check{\phi} = L + \widetilde{F}. \tag{3.2}$$

The infinitesimal version of a generalized symmetry is a vector field X such that its flow ϕ_s together with a family F_s of functions is a generalized symmetry for all s (where it is defined):

$$J_L[\phi_s \circ x] = J_L[x] + F_s(x_b) - F_s(x_a) \tag{3.3}$$

for all s. Define $f = \frac{\partial F_s}{\partial s}|_{s=0}$. We then say that the pair (X, f) defines a **generalized infinitesimal symmetry**. By differentiating equations (3.2) and (3.3) with respect to s at $s = 0$ we get the following

PROPOSITION 3.10. *The pair (X, f) defines a generalized infinitesimal symmetry of J_L if and only if*

$$\boxed{\check{X}(L) = \widetilde{f}}$$

with $\widetilde{f}(q, v) := \sum_{i=1}^d v^i \frac{\partial f}{\partial q^i}(q)$. *In this case*

$$\boxed{I_X := \iota_X \alpha_L - f}$$

is a constant of motion.

REMARK 3.11. The even more general case when the restriction of $J_{L,\phi}$ on each $\mathcal{P}^{[a,b]}_{x_a,x_b}$ is only locally constant[2] on each path space $\mathcal{P}^{[a,b]}_{x_a,x_b} := \{x \colon [a, b] \to U : x(a) = x_a,\ x(b) = x_b\}$ can be treated with a bit more of work. We enunciate the results for completeness. A diffeomorphism ϕ of U has this property if and only there is a closed 1-form $\Theta = \sum_{i=1}^d \Theta_i \, \mathrm{d}q^i$ such that

$$J_L[\phi \circ x] = J_L[x] + \int_x \Theta,$$

and this occurs if and only if $L \circ \check{\phi} = L + \widetilde{\Theta}$ with $\widetilde{\Theta}(q, v) = \sum_{i=1}^d v^i \, \Theta_i(q)$. A vector field X defines the infinitesimal version of this if and only if there is a closed 1-form $\theta = \sum_{i=1}^d \theta_i \, \mathrm{d}q^i$ such that $\check{X}(L) = \widetilde{\theta}$ with $\widetilde{\theta}(q, v) = \sum_{i=1}^d v^i \, \theta_i(q)$. This does not lead to a conservation law if θ is not exact, but only to the statement that the closed 1-form $\iota_X \alpha_L - \theta$ integrates to zero along every orbit. Notice that if the closed form θ is exact, $\theta = \mathrm{d}f$, then we are in the case described above, and we have indeed a constant of motion as in Proposition 3.10. By Poincaré Lemma we know that this necessarily happens if U is star shaped. Otherwise, we may always restrict our attention to a star-shaped neighborhood $V \subset U$ of the initial conditions. As long as the orbit lies in V, we have a constant of motion. We may also take a later point of this orbit as a new initial condition and choose a new star-shaped neighborhood. In general, we may patch the whole orbit by star-shaped neighborhoods V_i, and in each of them we have a constant of motion. In each V_i we have indeed a function f_i such that $\theta|_{V_i} = \mathrm{d}f_i$ and a constant of motion $I_{X,i} := \iota_X \alpha_L - f_i$ defined on V_i. Notice however that, if V_i and V_j have

[2]This means that each such restriction does not change under continuous deformations of the path. On the other hand, if the path space is not connected, the restriction may take a different value on different connected components.

a nonempty intersection, the restriction of f_i and f_j to this intersection will not be equal in general but will differ by a constant. Therefore, in this setting, a constant of motion exists, but only locally, and it may be considered globally only up to locally defined constants. In particular, it is not a globally defined function.

REMARK 3.12. Notice that we have a linear map

$$\begin{array}{rcl} T\colon & \Omega^1(U) & \to C^\infty(U\times\mathbb{R}^d) \\ & \sum_{i=1}^d \theta_i(q)\,\mathrm{d}q^i & \mapsto \sum_{i=1}^d v^i\,\theta_i(q). \end{array}$$

The function $\widetilde{\theta}$ used in the previous remark is then $T\theta$. The function $\widetilde{f}$ used above is $T\mathrm{d}f$. Notice that

$$T\phi^*\theta = \check{\phi}^*T\theta$$

for every smooth map $\phi\colon U\to V$ and for every $\theta\in\Omega^1(V)$. More asbtracly, we may write

$$\boxed{T\phi^* = \check{\phi}^*T} \tag{3.4}$$

as an equality of linear maps $\Omega^1(V)\to\Omega^1(U)$.

3.1.3.1. *Equivalent Lagrangians.* Let L be a Lagrangian on $U\times\mathbb{R}^d\times\mathrm{I}$, U open in $\mathbb{R}^d$, and let G be a function on U. Define

$$\widetilde{L} = L + T\mathrm{d}G$$

(i.e., $\widetilde{L}(q,v,t) = L(q,v,t) + \sum_{i=1}^d v^i\frac{\partial G}{\partial q^i}$). We then have $J_{\widetilde{L}}[x] = J_L[x] + G(x_b) - G(x_a)$ for all $x\in\mathcal{P}^{[a,b]}_{x_a,x_b}$. This implies that L and $\widetilde{L}$ have the same extremal paths and are therefore equivalent from the point of view of Euler–Lagrange equations. The Noether 1-forms are related by

$$\alpha_{\widetilde{L}} = \alpha_L + \mathrm{d}G.$$

Also notice that, if (ϕ,F) is a generalized symmetry for L, then by (3.4) we get that $(\phi, F+\phi^*G)$ is a generalized symmetry for $\widetilde{L}$. Hence, if (X,f) is a generalized infinitesimal symmetry for L, we see that $(X, f+X(G))$ is a generalized infinitesimal symmetry for $\widetilde{L}$. We then have that $\iota_X\alpha_{\widetilde{L}} - f - X(G) = \iota_X\alpha_L - f$, so the constant of motion corresponding to the generalized symmetry (X,f) of L is equal to the constant of motion corresponding to the generalized symmetry $(X, f+X(G))$ of $\widetilde{L}$.

Notice that a strict infinitesimal symmetry X of L is only a generalized one if one uses the equivalent Lagrangian $\widetilde{L}$ and $X(G)\neq 0$. Therefore, the notion of strict symmetry really depends on the choice of Lagrangian. Since one cannot be sure to have chosen the "right" Lagrangian (and there might be no Lagrangian that is "right" for all symmetries in case there are more at hand), the correct framework to use is that of generalized symmetries.

Also observe that, if f can be written as $-X(G)$ for some function G, then we may transform the generalized infinitesimal symmetry (X, f) of L into an infinitesimal symmetry, in the strict sense, of the equivalent Lagrangian $\widetilde{L}$. Notice however that, even if this may be possible, different infinitesimal symmetries may require different equivalent Lagrangians to be made strict.

3.1.3.2. *Example: constant magnetic field.* Consider a charged particle moving in a constant magnetic field of magnitude $B \neq 0$. The system is clearly invariant under translations in every direction. On the other hand, the Lagrangian depends on a vector potential generating this field, and this cannot be translation invariant, otherwise it would produce the zero magnetic field. This provides an example of generalized symmetry.

Suppose for definiteness that B points in x^3-direction and the particle has mass m and charge e. Newton's equation then read (setting the speed of light c to 1)

$$\begin{aligned} m\ddot{x}^1 &= eB\dot{x}^2, \\ m\ddot{x}^2 &= -eB\dot{x}^1, \\ m\ddot{x}^3 &= 0, \end{aligned}$$

and they are clearly invariant under translations $\boldsymbol{x}(t) \mapsto \boldsymbol{x}(t) + \boldsymbol{a}$, where $\boldsymbol{a}$ is an arbitrary vector, since they only depend on the first and second time derivatives of $\boldsymbol{x}$. On the other hand, the Lagrangian of the system is

$$L(\boldsymbol{x}, \boldsymbol{v}) = \frac{1}{2} m ||\boldsymbol{v}||^2 + e\boldsymbol{v} \cdot \boldsymbol{A}(\boldsymbol{x}).$$

Notice that $\boldsymbol{A}$ cannot be translation invariant (i.e., constant) since $\boldsymbol{B} = \nabla \times \boldsymbol{A} \neq 0$. For example we may choose $A_1 = -Bx^2$, $A_2 = A_3 = 0$ getting

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

We then have

$$\alpha_L = \sum_{i=1}^{3} m v^i \mathrm{d}q^i - eB\, x^2 \mathrm{d}x^1.$$

Notice that this Lagrangian is invariant under translations in the x^1 and in the x^3 direction. So we have the integrals of motion

$$\begin{aligned} P_1 &:= \iota_{\frac{\partial}{\partial x^1}} \alpha_L &= m v^1 - eB\, x^2, \\ P_3 &:= \iota_{\frac{\partial}{\partial x^3}} \alpha_L &= m v^3. \end{aligned}$$

Under a translation in the x^2 direction we have, on the other hand,

$$\frac{\check{\partial}}{\partial x^2}L = \frac{\partial}{\partial x^2}L = -eB\,v^1 = \widetilde{f}$$

with $f(\boldsymbol{x}) = -eB\,x^1$. We then get the integral of motion

$$P_2 := \iota_{\frac{\partial}{\partial x^2}}\alpha_L - f = mv^2 + eB\,x^1.$$

Exercise 3.13. Show that rotations around the x^3 axis are also generalized symmetries and compute the corresponding integral of motion.

Exercise 3.14. Show that the integrals of motions computed above do not change if the vector potential $\boldsymbol{A}$ is changed to $\boldsymbol{A} + \nabla\lambda$.

Exercise 3.15. Show that one can choose $\boldsymbol{A}$ such that $\boldsymbol{v}\cdot\boldsymbol{A}$ is invariant under rotations around the x^3 axis, so that rotations around the x^3 axis become symmetries in the strict sense.

Exercise 3.16. Show that changing the vector potential $\boldsymbol{A}$ to $\widetilde{\boldsymbol{A}} = \boldsymbol{A} + \nabla\lambda$ produces an equivalent Lagrangian.

3.2. From the Lagrangian to the Hamiltonian formalism

Suppose L is a hyperregular Lagrangian on $U\times\mathbb{R}^d\times I$. Denote by ψ_L its associated Legendre mapping and by H its Legendre transform. The first observation is that the factors $\frac{\partial L}{\partial v^i}$ appearing in the Definition 3.1 of the Noether 1-form are just the generalized momenta p_i, so we have

$$\alpha_L = \psi_L^*\alpha,$$

where

$$\alpha := \sum_{i=1}^d p_i \mathrm{d}q^i$$

is the Liouville 1-form (a.k.a. the Poincaré 1-form or the tautological 1-form). Notice that the dependency on the Lagrangian of the Noether 1-form comes only through the Legendre mapping: the Liouville 1-form is independent of L. Also notice that the canonical symplectic form ω is actually $\mathrm{d}\alpha$. Now suppose that X is an infinitesimal symmetry. Then $I_X = \iota_X\alpha_L$ is a constant of motion. We clearly have $I_X = \psi_L^* F_X$ with

$$F_X = \iota_X\alpha = \sum_{i=1}^d X^i(q)\,p_i.$$

Since this is a constant of motion, we then have $X_H(F_X) = 0$. But this is equivalent to

$$\widetilde{X}(H) = 0, \tag{3.5}$$

where $\widetilde{X}$ is the Hamiltonian vector field of F_X:

$$\iota_{\widetilde{X}}\omega = -\mathrm{d}F_X. \tag{3.6}$$

This is a consequence of the following basic

LEMMA 3.17. *Let f and g be functions, and let X_f and X_g be their Hamiltonian vector fields. Then*

$$\boxed{X_f(g) = -X_g(f) = \iota_{X_g}\iota_{X_f}\omega}$$

We will se that equations (3.5) and (3.6) characterize symmetries in the Hamiltonian formalism. The present case has however two peculiarities. The first is that F_X is linear in the p variables. The second is that we have $F_X = \iota_{\widetilde{X}}\alpha$, which in turn implies $\mathrm{L}_{\widetilde{X}}\,\alpha = 0$.

A simple computation shows that

$$\widetilde{X} = X - \sum_{i,j=1}^{d} \frac{\partial X^i}{\partial q^j}(q)\, p_i\, \frac{\partial}{\partial p_j},$$

which is called the cotangent lift of X.

If (X, f) is a generalized inifnitesimal symmetry, then we have $I_X = \iota_X\alpha_L - f = \psi_L^* F_{X,f}$ with $F_{X,f} = \sum_{i=1}^d X^i(q)\, p_i - f(q)$. If we denote by $\widetilde{X}_f = \widetilde{X} - X_f$ the Hamiltonian vector field of $F_{X,f}$, we then get $\widetilde{X}_f(H) = 0$ and $\iota_{\widetilde{X}_f}\omega = -\mathrm{d}F_{X,f}$, which have the same form as (3.5) and (3.6). Notice that in this case $F_{X,f}$ is at most linear in the p variables and that $\mathrm{L}_{\widetilde{X}_f}\,\alpha = \mathrm{d}f$.

3.3. Symmetries in Hamiltonian mechanics

Let $H \in C^\infty(V)$ be a time-independent Hamiltonian on an open subset V of $\mathbb{R}^d \times \mathbb{R}^d$. Notice that we can assume H to be time independent without loss of generality.[3]

[3] If $H(q, p, t)$ were time dependent, we could replace it by the time-independent Hamiltonian $\widetilde{H}(q, \tau, p, p_\tau) = H(q, p, \tau) + p_\tau$ on extended phase space.

Recall that Hamilton's equation for H can be obtained as Euler–Lagrangian equations for the Lagrangian[4]

$$\widetilde{L}_H(q,p,v_q,v_p) = \sum_{i=1}^d p_i v_q^i - H(q,p) = T\alpha - H,$$

where we have used the notations of Remark 3.12 with $\alpha = \sum_{i=1}^d p_i\,\mathrm{d}q^i$ the Liouville 1-form. Notice that we also have $\alpha_{\widetilde{L}_H} = \alpha$.

A symmetry is then a diffeomorphism ϕ of the phase space V such that $\check{\phi}^*\widetilde{L}_H = \widetilde{L}_H$. By Remark 3.12, we have $\check{\phi}^*\widetilde{L}_H = T\phi^*\alpha - \phi^*H$. Since H does not depend on the velocities, we then have that ϕ is a symmetry if and only if

$$\phi^*H = H \quad \text{and} \quad \phi^*\alpha = \alpha.$$

Notice that the second equation also implies $\phi^*\omega = \omega$. A vector field Y on V is then an infinitesimal symmetry if and only if

$$Y(H) = 0 \quad \text{and} \quad L_Y\alpha = 0,$$

whereas the corresponding constant of motion is $\iota_Y\alpha$.

This occurs, e.g., if H is the Legendre transform of L and Y is the cotangent lift of an infinitesimal symmetry of L.

Similarly, we see that (ϕ, F) is a generalized symmetry if and only if

$$\phi^*H = H \quad \text{and} \quad \phi^*\alpha = \alpha + \mathrm{d}F.$$

A pair (Y, g) of a vector field and a function on V is then a generalized infinitesimal symmetry if and only if

$$Y(H) = 0 \quad \text{and} \quad L_Y\alpha = \mathrm{d}g.$$

In this case, the constant of motion is $F = \iota_Y\alpha - g$. This implies

$$\iota_Y\omega = -\mathrm{d}F.$$

Notice that now F can be an arbitrary function on V. We have thus arrived at the

THEOREM 3.18. *A (possibly generalized) infinitesimal symmetry of a Hamiltonian system on V with Hamiltonian function H is a Hamiltonian vector field X_F, $\iota_{X_F}\omega = -\mathrm{d}F$, such that $X_F(H) = 0$.*

[4]Notice that fixing both positions and momenta at endpoints in general yields no solutions. On the other hand, the boundary term in computing the functional derivative of this Lagrangian does not involve the variation of the momenta. One then considers extremal paths with fixed qs at the endpoints, leaving the ps free. The corresponding Euler–Lagrange equations for these extremal paths are then the Hamilton equations.

REMARK 3.19. Notice that by Lemma 3.17 this immediately implies $X_H(F) = 0$, so F is the corresponding constant of motion.

REMARK 3.20. Lemma 3.17 can also be read the other way around. Namely, suppose that F is a constant of motion for a Hamiltonian system with Hamiltonian H (i.e., $X_H(F) = 0$). Then X_F is a symmetry (i.e., $X_F(H) = 0$). In this way, **we have a one-to-one correspondence between symmetries and constants of motion up to an additive constant.**

REMARK 3.21. Suppose that H is the Legendre transform of a Lagrangian L defined on $U \times \mathbb{R}^d =: V$. In the Lagrangian formalism, infinitesimal symmetries are vector fields on U that preserve L (possibly up to a term $T\mathrm{d}f$, with f a function on U). These correspond to infinitesimal symmetries in the Hamiltonian formalism whose constants of motions are at most linear in the p variables. These are just very particular examples of symmetries.

REMARK 3.22. The general case discussed in Remark 3.11 corresponds, in its infinitesimal form, to having a vector field Y and a closed 1-from θ such that $Y(H) = 0$ and $\mathrm{L}_Y\,\alpha = \theta$. Notice that the second equation is equivalent to $\mathrm{L}_Y\,\omega = 0$. Again this does not produce a constant of motion in general. However, locally we may have θ exact and proceed as above.

3.3.1. Symplectic geometry. The above results have a nice, general form in symplectic geometry. There one just assumes to have a closed 2-form ω with the property that for each function f there is a unique vector field X_f, called the Hamiltonian vector field of f, such that $\iota_{X_f}\omega = -\mathrm{d}f$. The latter property may be checked by computing the matrix representing ω at each point in local coordinates and verifying that it is nondegenerate.

A function f satisfying $\iota_X\omega = -\mathrm{d}f$ is called a Hamiltonian function for X and is not uniquely defined (it is defined up to a constant on each connected component). A vector field that is the Hamiltonian vector field for some function is called a Hamiltonian vector field. Not every vector field is Hamiltonian. Notice that a Hamiltonian vector field X automatically satisfies $L_X\omega = 0$. A vector satisfying this property is called a symplectic vector field. In general not every symplectic vector field is Hamiltonian.

Notice that Lemma 3.17 holds in this general setting. Hence, if we are given a Hamiltonian function H, which defines the dynamic of the system, we define a symmetry to be a Hamiltonian vector field X such that $X(H) = 0$. If f is a Hamiltonian function for X, then the lemma implies that $X_H(f) = 0$, so f is a constant of motion. This is the general, simple and beautiful version of Noether's theorem in symplectic geometry.

3.3.2. The Kepler problem. Consider the Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{p}) = \frac{||\boldsymbol{p}||^2}{2m} - \frac{K}{||\boldsymbol{q}||}$$

for a body in a gravitational field, where K is a positive constant. Notice that H is a smooth function on $U \times \mathbb{R}^3$ with $U = \mathbb{R}^3 \setminus 0$. Rotation invariance, in the Lagrangian version, yields conservation of the angular momentum[5]

$$\boldsymbol{J} = \boldsymbol{q} \times \boldsymbol{p},$$

see Example 3.7. Notice that the components of $\boldsymbol{J}$ are linear in $\boldsymbol{p}$ as they come from symmetries in the Lagrangian formalism. Another constant of motion is the Laplace–Runge–Lenz vector (shortly, the Lenz vector)

$$\boldsymbol{A} = \boldsymbol{p} \times \boldsymbol{J} - mK\frac{\boldsymbol{q}}{||\boldsymbol{q}||}.$$

The simplest way to prove that $\boldsymbol{A}$ is conserved is by taking its time derivative along a solution to Hamilton's equations

$$\dot{\boldsymbol{q}} = \frac{\boldsymbol{p}}{m},$$
$$\dot{\boldsymbol{p}} = -K\frac{\boldsymbol{q}}{||\boldsymbol{q}||^3}$$

Using $\dot{\boldsymbol{J}} = 0$ along a solution, we then get

$$\dot{\boldsymbol{A}} = -K\frac{\boldsymbol{q}}{||\boldsymbol{q}||^3} \times \boldsymbol{J} - K\left(\frac{\boldsymbol{q}}{||\boldsymbol{q}||} - \frac{\boldsymbol{p}\cdot\boldsymbol{q}}{||\boldsymbol{q}||^3}\right) = 0.$$

From $X_H(\boldsymbol{A}) = 0$ we then get $X_{\boldsymbol{A}^i}(H) = 0$ for all i.

Notice that $\boldsymbol{A}$ has a quadratic term in the $\boldsymbol{p}$ variables, so it does not come from a symmetry in the Lagrangian setting (nor is $X_{\boldsymbol{A}^i}$ a cotangent lift). On the other hand, the vector fields $X_{\boldsymbol{A}^i}$ are symmetries in the Hamiltonian formalism.

The conservation of the Lenz vector can be used to derive the Kepler orbits directly (without having to solve the differential equations). First observe that, by the cyclic property of the triple product, we have

$$\boldsymbol{A} \cdot \boldsymbol{q} = ||\boldsymbol{J}||^2 - mK||\boldsymbol{q}||.$$

Recall that adapting the coordinates to the initial conditions we may assume that the motion occurs in the xy plane. Then $\boldsymbol{J}$ points in the

[5]By this we mean that each component of the vector $\boldsymbol{J}$ is a constant of motion.

z direction. So $\boldsymbol{A}$ is also in the xy plane. If we denote by θ the angle between $\boldsymbol{A}$ and $\boldsymbol{q}$, we may then rewrite the above equation as

$$Ar\cos\theta = J^2 - mKr,$$

where we have set $r := ||\boldsymbol{q}||$, $J := ||\boldsymbol{J}||$ and $A := ||\boldsymbol{A}||$. Notice that J and A are obviously also constants of motion. We then get

$$r = \frac{\frac{J^2}{mK}}{1 + \frac{A}{mK}\cos\theta},$$

which shows that the orbits are conic sections with eccentricity $e = \frac{A}{mK}$.

REMARK 3.23. The Hamiltonian vector fields X_{J_i} generate infinitesimal rotations and correspond to the fact that the Hamiltonian is rotation invariant (with rotations extended to phase space by the cotangent lift; i.e, $\mathsf{A} \in SO(3)$ acts by $(\boldsymbol{q}, \boldsymbol{p}) \mapsto (\mathsf{A}\boldsymbol{q}, \mathsf{A}\boldsymbol{p})$). The vector fields X_{A_i} generate additional infinitesimal transformations. One can show that the X_{J_i}'s and the X_{A_i}'s together generate a (nonlinear) action of the group $SO(4)$ on phase space.

CHAPTER 4

The Hamilton–Jacobi Equation

The Hamilton–Jacobi equation is a PDE associated to a Hamiltonian system with which it has a two-way link. On the one hand, it is the PDE satisfied by the action functional evaluated on orbits as a function of the endpoint variables; as such, solving the Hamilton equation provides an effective method of solving the Cauchy problem for the Hamilton–Jacobi equation (method of characteristics). On the other hand, a solution depending on enough parameters (complete integral) provides a generating function for a canonical transformation that trivializes the Hamiltonian, thus allowing one to solve the Hamilton equations.

The method actually only works for very special systems (integrable systems); however, perturbations of integrable systems are more effectively studied in the canonical variables in which the unperturbed Hamiltonian is trivialized.

Finally, the Hamilton–Jacobi equation is related to the asymptotics of the Schrödinger equation in the semiclassical limit.

4.1. The Hamilton–Jacobi equation

Throughout we will denote by U an open subset of $\mathbb{R}^d$, by I an interval and by H a Hamiltonian function on $U \times \mathbb{R}^d \times I$, which we will write as $H(q, p, t)$.

The Hamilton–Jacobi equation for the unknown function S on an open subset of of $U \times I$ is then

$$\boxed{\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right) = 0}$$

where $\frac{\partial S}{\partial q}$ is a shorthand notation for $\frac{\partial S}{\partial q^1}, \dots, \frac{\partial S}{\partial q^d}$.

If the Hamiltonian does not depend on the time variable t, we simply write $H(q, p)$. To it one associates the reduced Hamilton–Jacobi equation in the unknown function S_0 on an open subset of U:

$$\boxed{H\left(q, \frac{\partial S_0}{\partial q}\right) = E}$$

where E is a parameter, called the energy. Notice that if S_0 is a solution of the reduced Hamilton–Jacobi equation at energy E, then $S(q,t) = S_0(q) - Et$ is a solution of the Hamilton–Jacobi equation.

Finally, notice that S and S_0 enter into the equations only through their derivatives; so shifting them by constants yields new solutions.

REMARK 4.1. The Hamilton–Jacobi equation is also related to the asymptotics of the Schrödinger equation which appears in quantum mechanics. For the Hamiltonian $H(q,p) = \sum_{i=1}^d \frac{p_i^2}{2m} + V(q)$, this is the PDE

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = -\sum_{i=1}^d \frac{\hbar^2}{2m}\frac{\partial^2\psi}{(\partial q^i)^2} + V(q)\psi,$$

where the unknown ψ is a time-dependent complex-valued function on $U$and $\hbar$ is a constant. If one writes

$$\psi(q,t) = A(q,t)\,\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\phi(q,t)},$$

where A and ϕ are real valued, and assumes that for $\hbar$ small we have $A = A_0 + O(\hbar)$ and $\phi = \phi_0 + O(\hbar)$, then ϕ_0 solves the Hamilton–Jacobi equation on $W := \{q \in U : A_0(q,t) \neq 0\,\forall t\}$. This gives an indication that in the limit $\hbar \to 0$ quantum mechanics is approximated by classical mechanics.

4.2. The action as a function of endpoints

Let L be a Lagrangian function on $U \times \mathbb{R}^d \times I$ and denote by S its corresponding action functional:

$$S[x] = \int_{t_A}^{t_B} L(x(t), \dot{x}(t), t)\,\mathrm{d}t,$$

where x is a map $[t_A, t_B] \to U$, with $[t_A, t_B] \subseteq I$. Let W be an open subset of $U \times I \times U \times I$ such that for each $(q_A, t_A, q_B, t_B) \in W$ there is a unique extremal path, denoted by $q^*_{q_A,t_A,q_B,t_B}$ (or simply q^*). Define

$$S^*(q_A, t_A, q_B, t_B) := S[q^*_{q_A,t_A,q_B,t_B}].$$

EXAMPLE 4.2. Consider the free particle in one dimension; i.e., $d = 1$, $U = \mathbb{R}$, $L(q,v,t) = \frac{1}{2}mv^2$. The EL equation, $m\ddot{q} = 0$, is easily solved, and we have, with $W = \{(q_A, t_A, q_B, t_B) \in U \times I \times U \times I : t_A \neq t_B\}$,

$$q^*_{q_A,t_A,q_B,t_B}(t) = \frac{(t - t_A)q_B + (t_B - t)q_A}{t_B - t_A}.$$

Hence

$$\dot{q}^*_{q_A,t_A,q_B,t_B}(t) = \frac{q_B - q_A}{t_B - t_A}$$

and

$$S^*(q_A,t_A,q_B,t_B) = \int_{t_A}^{t_B} \frac{1}{2}m\left(\frac{q_B-q_A}{t_B-t_A}\right)^2 \mathrm{d}t = \frac{1}{2}m\frac{(q_B-q_A)^2}{t_B-t_A}.$$

We now want to study the dependency of S^* on its arguments.

THEOREM 4.3. *We have*

$$\frac{\partial S^*}{\partial q_B^i} = p_{Bi}, \qquad \frac{\partial S^*}{\partial t_B} = -H_B,$$
$$\frac{\partial S^*}{\partial q_A^i} = -p_{Ai}, \qquad \frac{\partial S^*}{\partial t_A} = H_A,$$

where

$$p_{Bi}(q_A,t_A,q_B,t_B) := \frac{\partial L}{\partial v^i}(q_B, \dot q^*_{q_A,t_A,q_B,t_B}(t_B), t_B),$$
$$p_{Ai}(q_A,t_A,q_B,t_B) := \frac{\partial L}{\partial v^i}(q_A, \dot q^*_{q_A,t_A,q_B,t_B}(t_A), t_A),$$
$$H_B(q_A,t_A,q_B,t_B) = \sum_{i=1}^d p_{Bi}(q_A,t_A,q_B,t_B)\dot q^{*i}(t_B) - L(q_B,\dot q^*(t_B),t_B),$$
$$H_A(q_A,t_A,q_B,t_B) = \sum_{i=1}^d p_{Ai}(q_A,t_A,q_B,t_B)\dot q^{*i}(t_A) - L(q_A,\dot q^*(t_A),t_A).$$

One can also compactly write

$$\mathrm{d}S^* = -\sum_{i=1}^d p_{Ai}\mathrm{d}q_A^i + H_A\mathrm{d}t_A + \sum_{i=1}^d p_{Bi}\mathrm{d}q_B^i - H_B\mathrm{d}t_B.$$

EXAMPLE 4.4. Let us check the above formulae in the case of Example 4.2. We explicitly have

$$\frac{\partial S^*}{\partial q_B} = m\frac{q_B-q_A}{t_B-t_A} = m\dot q^*_{q_A,t_A,q_B,t_B}(t_B)$$

and

$$\frac{\partial S^*}{\partial t_B} = -\frac{1}{2}m\left(\frac{q_B-q_A}{t_B-t_A}\right)^2 = -\frac{1}{2}m\dot q^*_{q_A,t_A,q_B,t_B}(t_B)^2.$$

PROOF OF THEOREM 4.3. We begin with the derivatives with respect to q_B. Let δq_B be a vector in $\mathbb{R}^d$. We have

$$\lim_{\epsilon\to 0}\frac{S^*(q_A,t_A,q_B+\epsilon\delta q_B,t_B) - S^*(q_A,t_A,q_B,t_B)}{\epsilon} = \sum_{i=1}^d \frac{\partial S^*}{\partial q_B^i}\delta q_B^i.$$

Write $q^*_\epsilon := q^*_{q_A,t_A,q_B+\epsilon\delta q_B,t_B}$ and $q^* := q^*_\epsilon|_{\epsilon=0}$. We have

$$q^*_\epsilon = q^* + \epsilon\delta q + O(\epsilon^2)$$

for a uniquely defined path $\delta q\colon [t_A, t_B] \to \mathbb{R}^d$. Thus,

$$\lim_{\epsilon\to 0} \frac{S[q^*_\epsilon] - S[q^*]}{\epsilon} = \frac{\delta S}{\delta q}[q^*, \delta q],$$

which implies

$$\sum_{i=1}^d \frac{\partial S^*}{\partial q^i_B}\delta q^i_B = \frac{\delta S}{\delta q}[q^*, \delta q].$$

Now observe that, since q^*_ϵ is an extremal path, we have

$$\frac{\delta S}{\delta q}[q^*, \delta q] =$$
$$= \sum_{i=1}^d \left(\frac{\partial L}{\partial v^i}(q_B, \dot q^*(t_B), t_B)\delta q^i(t_B) - \frac{\partial L}{\partial v^i}(q_A, \dot q^*(t_A), t_A)\delta q^i(t_A) \right). \tag{4.1}$$

Since $q^*_\epsilon(t_A) = q_A$, we get $\delta q(t_A) = 0$. From $q^*_\epsilon(t_B) = q_B + \epsilon\delta q_B$, we conclude $\delta q(t_B) = \delta q_B$. Thus,

$$\frac{\delta S}{\delta q}[q^*, \delta q] = = \sum_{i=1}^d \frac{\partial L}{\partial v^i}(q_B, \dot q^*(t_B), t_B)\delta q^i_B,$$

which proves the first equation.

We now come to the second equation, the derivative with respect to t_B. For $\delta t_B \in \mathbb{R}$, we have

$$\lim_{\epsilon\to 0} \frac{S^*(q_A, t_A, q_B, t_B + \epsilon\delta t_B) - S^*(q_A, t_A, q_B, t_B)}{\epsilon} = \frac{\partial S^*}{\partial t_B}\delta t_B.$$

Write $q^*_\epsilon := q^*_{q_A,t_A,q_B,t_B+\epsilon\delta t_B}$ and $q^* := q^*_\epsilon|_{\epsilon=0}$. Notice that q^*_ϵ is defined on the interval $[t_A, t_B + \epsilon\delta t_B]$. Assume $\epsilon\delta t_B \geq 0$ and denote by $\widetilde{q}^*_\epsilon$ the restriction of q^*_ϵ to $[t_A, t_B]$. We then have

$$S[q^*_\epsilon] = S[\widetilde{q}^*_\epsilon] + \int_{t_B}^{t_B+\epsilon\delta t_B} L(q^*_\epsilon(t), \dot q^*_\epsilon(t), t)\, \mathrm{d}t.$$

Notice that we have

$$\widetilde{q}^*_\epsilon = q^* + \epsilon\delta q + O(\epsilon^2)$$

for a uniquely defined path $\delta q\colon [t_A, t_B] \to \mathbb{R}^d$. Thus,[1]

$$\lim_{\epsilon\to 0} \frac{S[q^*_\epsilon] - S[q^*]}{\epsilon} = \frac{\delta S}{\delta q}[q^*, \delta q] + L(q^*_\epsilon(t_B), \dot q^*_\epsilon(t_B), t_B)\delta t_B,$$

[1]The limit is for $\epsilon \to 0^+$ if $\delta t_B \geq 0$ and for $\epsilon \to 0^-$ otherwise.

which implies

$$\frac{\partial S^*}{\partial t_B}\delta t_B = \frac{\delta S}{\delta q}[q^*, \delta q] + L(q^*_\epsilon(t_B), \dot{q}^*_\epsilon(t_B), t_B)\delta t_B.$$

We use again (4.1). Notice that $\widetilde{q}^*_\epsilon(t_A) = q^*_\epsilon(t_A) = q_A$ implies $\delta q(t_A) = 0$. On the other hand,

$$\begin{aligned} q_B = q^*_\epsilon(t_B + \epsilon\delta t_B) &= q^*_\epsilon(t_B) + \epsilon\dot{q}^*_\epsilon(t_B)\delta t_B + O(\epsilon^2) \\ &= q^*(t_B) + \epsilon(\delta q(t_B) + \dot{q}^*_\epsilon(t_B)\delta t_B) + O(\epsilon^2) \\ &= q_B + \epsilon(\delta q(t_B) + \dot{q}^*_\epsilon(t_B)\delta t_B) + O(\epsilon^2), \end{aligned}$$

so $\delta q(t_B) = -\dot{q}^*_\epsilon(t_B)\delta t_B$. We conclude that

$$\frac{\delta S}{\delta q}[q^*, \delta q] = -\sum_{i=1}^{d} \frac{\partial L}{\partial v^i}(q_B, \dot{q}^*(t_B), t_B)(\dot{q}^*_\epsilon)^i(t_B)\delta t_B,$$

which proves the second equation.

The third and the fourth equations are proved along the same lines. □

Now assume that L is hyperregular and denote by H its Legendre transform. We then have

$$H_B(q_A, t_A, q_B, t_B) = H(q_B, p_B(q_A, t_A, q_B, t_B), t_B).$$

The first two equations in Theorem 4.3 then imply

$$\frac{\partial S^*}{\partial t_B} + H\left(q_B, \frac{\partial S^*}{\partial q_B}, t_B\right) = 0;$$

i.e., S^* as a function of the end variables (q_B, t_B) satisfies the Hamilton–Jacobi equation.

4.3. Solving the Cauchy problem for the Hamilton–Jacobi equation

The Cauchy problem for the Hamilton–Jacobi equation is the system

$$\begin{aligned} \frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right) &= 0, \\ S(q, t_0) &= \sigma(q), \end{aligned}$$

where σ is a given function on U. For simplicity, and actually without loss of generality, we assume that H is time independent and take $t_0 = 0$.

We want to show that we can solve the Cauchy problem for the Hamilton–Jacobi equation by integrating the Hamilton equations. First define

$$\mathcal{L} := \left\{(q,p) \in U \times \mathbb{R}^d : p_i = \frac{\partial \sigma}{\partial q^i}(q)\ \forall i\right\}$$

and $\mathcal{L}_t := \phi_t(\mathcal{L})$, where ϕ_t is the flow of the Hamiltonian vector field of H.

Notice that $\mathcal{L}$ is defined as the graph of a map. This will be the case also for $\mathcal{L}_\tau$ for τ small. Let t_1 be the largest number such that $\mathcal{L}_\tau$ is a graph for all $\tau \in (0, t_1)$. Then for each $q \in U$ and for each $\tau \in (0, t_1)$ there is a unique $p(q,\tau)$ such that $(q, p(q,\tau)) \in \mathcal{L}_\tau$. We then have a unique orbit (q^*, p^*) on the interval $[0,\tau]$ such that $q^*(\tau) = q$ and $p^*(\tau) = p(q,\tau)$: namely, $(q^*, p^*)(t) = \phi_t(\phi_\tau^{-1}(q, p(q,\tau)))$.

Equivalently, (q^*, p^*) is the unique orbit with $q^*(\tau) = q$ and $p_i^*(0) = \frac{\partial \sigma}{\partial q^i}(q^*(0))\ \forall i$.[2] These orbits are called characteristics of the system.

Let $\phi\colon U \times (0, t_1) \to U$ be the map that assigns $q^*(0)$ to a pair (q,τ). Notice that $\lim_{\tau\to 0} \phi(q,\tau) = q\ \forall q \in U$, so we can extend ϕ to $U \times [0, t_1)$ Then we have the

Theorem 4.5. *If H is the Legendre transform of the Lagrangian L, then the function*

$$\boxed{S(q,\tau) := \sigma(\phi(q,\tau)) + \int_0^\tau L(q^*(t), \dot q^*(t))\,\mathrm{d}t}$$

solves the Cauchy problem for $\tau \in [0, t_1)$.

Proof. We clearly have $S(q,0) = \sigma(q)$. Then let

$$S_{\mathrm{char}}(q,\tau) := \int_0^\tau L(q^*(t), \dot q^*(t))\,\mathrm{d}t = S^*(\phi(q,\tau), 0, q, \tau).$$

By Theorem 4.3 we have

$$\frac{\partial S_{\mathrm{char}}}{\partial q^j}(q,\tau) = p_j(q,\tau) - \sum_{i=1}^d \frac{\partial \sigma}{\partial q^i}(\phi(q,\tau))\frac{\partial \phi^i}{\partial q^j}(q,\tau),$$

$$\frac{\partial S_{\mathrm{char}}}{\partial \tau}(q,\tau) = -H(q, p(q,\tau)) - \sum_{i=1}^d \frac{\partial \sigma}{\partial q^i}(\phi(q,\tau))\frac{\partial \phi^i}{\partial \tau}(q,\tau).$$

[2] In practice, one solves the backward Cauchy problem with final conditions $q^*(\tau) = q$ and $p^*(\tau) = p$ for some p and then uses the conditions $p_i^*(0) = \frac{\partial \sigma}{\partial q^i}(q^*(0))\ \forall i$ to determine p as a function of q and τ.

Hence

$$\frac{\partial S}{\partial q^j}(q,\tau) = p_j(q,\tau),$$
$$\frac{\partial S}{\partial \tau}(q,\tau) = -H(q,p(q,\tau)),$$

so S solves the Hamilton–Jacobi equation. □

For a general Hamiltonian H, we have the

THEOREM 4.6. *The function*

$$\boxed{S(q,\tau) := \sigma(\phi(q,\tau)) + \int_0^\tau \left(\sum_{i=1}^d p_i^*(t)\dot{q}^{*i}(t) - H(q^*(t),p^*(t))\right) \mathrm{d}t}$$

solves the Cauchy problem for $\tau \in [0,t_1)$.

Notice, by the way, that for H the Legendre transform of L one has

$$\int_0^\tau \left(\sum_{i=1}^d p_i^*(t)\dot{q}^{*i}(t) - H(q^*(t),p^*(t))\right) \mathrm{d}t = \int_0^\tau L(q^*(t),\dot{q}^*(t))\,\mathrm{d}t.$$

PROOF. We clearly have $S(q,0) = \sigma(q)$. Then recall that the Hamilton equations for H are also the EL equations for the the Lagrangian function

$$\widetilde{L}(q,p,v_q,v_p,t) := \sum_{i=1}^d p_i v_q^i - H(q,p,t)$$

defined on $(U \times \mathbb{R}^d) \times \mathbb{R}^{2d} \times I$. Denote by $\widetilde{S}$ the action functional corresponding to $\widetilde{L}$ and observe that

$$S_{\mathrm{Ham}}(q,\tau) := \int_0^\tau \left(\sum_{i=1}^d p_i^*(t)\dot{q}^{*i}(t) - H(q^*(t),p^*(t))\right) \mathrm{d}t = \widetilde{S}[(q^*,p^*)].$$

We now want to compute derivatives of S_{Ham} with respect to its arguments. We essentially proceed as in the proof of Theorem 4.3. The first remark is that, for any solution (Q,P) of the Hamilton equations on an interval $[0,\tau]$, we have

$$\begin{aligned}\delta\widetilde{S}[(Q,P),(\delta Q,\delta P)] &:= \lim_{\epsilon\to 0}\frac{\widetilde{S}[(Q+\epsilon\delta Q,P+\epsilon\delta P)] - \widetilde{S}[(Q,P)]}{\epsilon}\\ &= \sum_{i=1}^d \left(P_i(\tau)\delta Q^i(\tau) - P_i(0)\delta Q^i(0)\right).\end{aligned}$$

In particular, for the characteristic (q^*, p^*) we have

$$\delta\widetilde{S}[(q^*,p^*),(\delta Q,\delta P)] = \sum_{i=1}^d p_i(q,\tau)\delta Q^i(\tau) - \sum_{i=1}^d \frac{\partial\sigma}{\partial q^i}(\phi(q,\tau))\delta Q^i(0),$$

where again we have written $p^*(\tau) = p(q,\tau)$.

Now consider the characteristic $(q^*_\epsilon, p^*_\epsilon)$ corresponding to $(q+\epsilon\delta q, \tau)$. We write $q^*_\epsilon = q^* + \epsilon\delta q^* + O(\epsilon^2)$ and $p^*_\epsilon = p^* + \epsilon\delta p^* + O(\epsilon^2)$. Since

$$q + \epsilon\delta q = q^*_\epsilon(\tau) = q + \epsilon\delta q^*(\tau) + O(\epsilon^2),$$

we get $\delta q^*(\tau) = \delta q$. Since

$$\phi(q+\epsilon\delta q,\tau) = q^*_\epsilon(0) = \phi(q,\tau) + \epsilon\delta q^*(0) + O(\epsilon^2),$$

we get $\delta q^{*i}(0) = \sum_{j=1}^d \frac{\partial\phi^i}{\partial q^j}(q,\tau)\delta q^j$. Hence

$$\delta\widetilde{S}[(q^*,p^*),(\delta q^*,\delta p^*)] = \sum_{i=1}^d p_i(q,\tau)\delta q^i - \sum_{i,j=1}^d \frac{\partial\sigma}{\partial q^i}(\phi(q,\tau))\frac{\partial\phi^i}{\partial q^j}(q,\tau)\delta q^j.$$

Since

$$\lim_{\epsilon\to 0}\frac{S_{\text{Ham}}(q+\epsilon\delta q,\tau) - S_{\text{Ham}}(q,\tau)}{\epsilon} = \sum_{j=1}^d \frac{\partial S_{\text{Ham}}}{\partial q^j}(q,\tau)\delta q^j,$$

we finally get

$$\frac{\partial S_{\text{Ham}}}{\partial q^j}(q,\tau) = p_j(q,\tau) - \sum_{i=1}^d \frac{\partial\sigma}{\partial q^i}(\phi(q,\tau))\frac{\partial\phi^i}{\partial q^j}(q,\tau)$$

and so

$$\frac{\partial S}{\partial q^j}(q,\tau) = p_j(q,\tau).$$

Similarly, we now denote by $(q^*_\epsilon, p^*_\epsilon)$ the characteristic corresponding to $(q,\tau+\epsilon\delta\tau)$. We assume $\epsilon\delta\tau \geq 0$ and denote by $(\widetilde{q}^*_\epsilon, \widetilde{p}^*_\epsilon)$ the restriction of $(q^*_\epsilon, p^*_\epsilon)$ to $[0,\tau]$. We then have

$$S_{\text{Ham}}(q,\tau+\epsilon\delta\tau) = \\ = \widetilde{S}[(\widetilde{q}^*_\epsilon,\widetilde{p}^*_\epsilon)] + \left(\sum_{i=1}^d p_i(q,\tau)\dot{q}^{*i}(\tau) - H(q,p(q,\tau))\right)\epsilon\delta\tau + O(\epsilon^2).$$

We write $\widetilde{q}^*_\epsilon = q^* + \epsilon\delta q^* + O(\epsilon^2)$ and $\widetilde{p}^*_\epsilon = p^* + \epsilon\delta p^* + O(\epsilon^2)$. Reasoning as above we get $\delta q^*(\tau) = -\dot{q}^*(\tau)\delta\tau$ and $\delta q^{*i}(0) = \frac{\partial\phi^i}{\partial\tau}(q,\tau)\delta\tau$. Therefore, putting everything together, we get

$$\frac{\partial S_{\text{Ham}}}{\partial\tau}(q,\tau) = -H(q,p(q,\tau)) - \sum_{i=1}^d \frac{\partial\sigma}{\partial q^i}(\phi(q,\tau))\frac{\partial\phi^i}{\partial\tau}(q,\tau)$$

and so

$$\frac{\partial S}{\partial \tau}(q,\tau) = -H(q, p(q,\tau)).$$

Hence S solves the Hamilton–Jacobi equation. □

REMARK 4.7. In Remark 4.1 we have seen that the Hamilton–Jacobi equation is related to the asymptotics of the Schrödinger equation. With the results of this section, we now also see that $\psi(q,t) := \mathrm{e}^{\frac{\mathrm{i}}{\hbar}S(q,t)}$ solves the Schrödinger equation up to $O(\hbar)$. This shows the role of the exponential of the action functional in quantum mechanics (its full-fledged role appears in Feynman's path integral).

4.4. Generating functions

Let $\omega = \sum_{i=1}^d \mathrm{d}p_i \mathrm{d}q^i$ by the symplectic form on $W \subseteq \mathbb{R}^{2d}$ with coordinates (q,p) and $\Omega = \sum_{i=1}^d \mathrm{d}P_i \mathrm{d}Q^i$ be the symplectic form on $Z \subseteq \mathbb{R}^{2d}$ with coordinates (Q,P). Recall that a symplectomorphism (in this context a.k.a. canonical transformation) is a diffeomorphism $\phi\colon W \to Z$ such that $\phi^*\Omega = \omega$. Also recall that the orbits of the Hamiltonian system with Hamiltonian function H on W are bijectively mapped by the symplectomorphsm ϕ to the orbits of the Hamiltonian system with Hamiltonian function $\widetilde{H} := H \circ \phi^{-1}$ on Z. The idea is to look for a symplectomorphism that makes $\widetilde{H}$ very simple, so that its Hamilton equations can be solved explixitly.

We actually look for a symplectomorphism that makes $\widetilde{H}$ depend only on the P variables: $\widetilde{H}(Q,P) = K(P)$ for some function K. In this case, the Hamilton equations are just $\dot{P} = 0$ and $\dot{Q}^i = \frac{\partial K}{\partial P_i}(P)$, $\forall i$. The solution of the Cauchy problem with initial condition, say at time $t = 0$, given by (Q_0, P_0) is then $P(t) = P_0$ and $Q^i(t) = Q^i_0 + \frac{\partial K}{\partial P_i}(P_0)t$, $\forall i$.

Notice that $\omega = \mathrm{d}\alpha$ with $\alpha = \sum_{i=1}^d p_i \mathrm{d}q^i$ and that $\Omega = \mathrm{d}\beta$ with $\beta = \sum_{i=1}^d P_i \mathrm{d}Q^i$. Notice that a diffeomorphism $\phi\colon W \to Z$ such that $\alpha - \phi^*\beta$ is the differential of a function F is in particular a symplectomorphism. We will only consider symplectomorphisms of this form.[3] More explicitly, we denote by $Q^i(q,p)$ and $P_i(q,p)$ the components of ϕ. So we have

$$\sum_{i=1}^d p_i \mathrm{d}q^i - \sum_{i=1}^d P_i(q,p)\mathrm{d}Q^i(q,p) = \mathrm{d}F(q,p). \tag{4.2}$$

[3]In general, ϕ is a symplectomorphism if and only if $\alpha - \phi^*\beta$ is closed. If W and Z are contractible, in particular star shaped, then every closed 1-form is automatically exact.

Now we assume that the graph of ϕ in $Z \times W$ may be parametrized by the (q, P) variables (instead of the (q, p) variables). Namely, we want to solve the equations $P_i = P_i(q, p)$ with respect to the p variables getting them as smooth functions of the Ps and the qs. By the implicit function theorem, this is possible if the following condition is satisfied:

Assumption 4.8. We assume that the matrix $\left(\frac{\partial P_i}{\partial p^j}\right)_{i,j=1,\dots,d}$ is nondegenerate for all $(q, p) \in W$.

Under this assumption, we then get functions $\widetilde{p}(q, P)$ and define $\widetilde{Q}(q, P) := Q(q, \widetilde{p}(q, P))$. Equation (4.2) now becomes

$$\sum_{i=1}^{d} \widetilde{p}_i(q, P)\mathrm{d}q^i - \sum_{i=1}^{d} P_i \mathrm{d}\widetilde{Q}^i(q, P) = \mathrm{d}\widetilde{F}(q, P)$$

with $\widetilde{F}(q, P) = F(q, \widetilde{p}(q, P))$. Setting

$$S(q, P) := \widetilde{F}(q, P) + \sum_{i=1}^{d} P_i \widetilde{Q}^i(q, P),$$

we finally get

$$\sum_{i=1}^{d} p_i(q, P)\mathrm{d}q^i + \sum_{i=1}^{d} Q^i(q, P)\mathrm{d}P_i = \mathrm{d}S(q, P),$$

where we have removed the tildes for simplicity of notation. Notice that this equation is equivalent to the system

$$p_i = \frac{\partial S}{\partial q^i}, \tag{4.3}$$

$$Q^i = \frac{\partial S}{\partial P_i}, \tag{4.4}$$

for $i = 1, \dots, d$. Notice that Assumption 4.8 is satisfied if the following holds:

Assumption 4.9. The matrix $\left(\frac{\partial^2 S}{\partial q^j \partial P_i}\right)_{i,j=1,\dots,d}$ is nondegenerate for all (q, P).

As the map ϕ can then be reconstructed by these equations, an S satisfying this condition is called a generating function for ϕ. Next we want $\widetilde{H}(Q, P) = K(P)$. Since $\widetilde{H}(Q(q, P), P) = H(q, p(q, P))$, we get by (4.3) that

$$H\left(q, \frac{\partial S}{\partial q}\right) = K(P).$$

Hence S, as a function of q parametrized by P, solves the reduced Hamilton–Jacobi equation at energy $K(P)$. A solution satisfying Assumption 4.9 is called a complete integral.

What we have shown is Jacobi's theorem that a complete integral for the Hamilton–Jacobi equation for H allows one to solve its Hamilton equations.

Notice that the P variables are contants of motions for the $\widetilde{H}$ system. Also notice that their differentials are clearly linear independent and that their pairwise Poisson brackets vanish. Regarded as functions of the (q, p) variables, they are then independent constants of motions for the H system in involution. A d-dimensional Hamiltonian system with d independent constants of motions in involution is called integrable. We then see that the above method can only work for integrable systems.

APPENDIX A

Differential Forms

A.1. Notations

In these notes U will denote an open subset of $\mathbb{R}^n$, α a k-form, β an l-form, f a function and X a vector field on U. We will denote by $(x^1, \dots, x^n)$ the coordinates on U and accordingly write

$$\alpha(x) = \sum_{i_1,\dots,i_k=1}^{n} \alpha_{i_1,\dots,i_k}(x)\, \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k},$$

$$\beta(x) = \sum_{j_1,\dots,j_l=1}^{n} \beta_{j_1,\dots,j_l}(x)\, \mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_l},$$

$$X(x) = \sum_{i=1}^{n} X^i(x)\, \frac{\partial}{\partial x^i}.$$

For simplicity we will assume throughout that α, β, f and X are smooth, i.e., that all components $\alpha_{i_1,\dots,i_k}$, all components $\beta_{j_1,\dots,j_l}$, all components X^i and f are arbitrarily often continuously differentiable. Recall that functions and zero-forms are one and the same thing.

Remark A.1. The symbols $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ denote the basis of $\mathbb{R}^n$ corresponding to our choice of coordinates. The symbols $\mathrm{d}x^1, \dots, \mathrm{d}x^n$ denote the dual basis of $(\mathbb{R}^n)^*$; i.e., the canonical pairing of $\mathrm{d}x^i$ with $\frac{\partial}{\partial x^j}$ is 1 if $i = j$ and 0 otherwise. The induced basis of $\wedge^k(\mathbb{R}^n)^*$ is given by the symbols $(\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k})_{1\le i_1<\cdots<i_k\le n}$. The wedge product of the symbols $\mathrm{d}x^i$ is defined by the identity

$$\mathrm{d}x^i \wedge \mathrm{d}x^j = -\mathrm{d}x^j \wedge \mathrm{d}x^i.$$

Using this identity one can rewrite all the terms in the above expansion of α into a linear combination of the basis elements $(\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k})_{1\le i_1<\cdots<i_k\le n}$ of $\wedge^k(\mathbb{R}^n)^*$. Notice that the coefficients of each basis element is given by the complete antisymmetrization of the components with respect to their indices. This means that it is enough to consider completely antisymmetric components, but it is quite convenient (see, e.g., the formulae for the wedge product and for the exterior derivative below) to allow for more general (though redundant) components.

REMARK A.2. If $k = 0$, then α is a function; if $k > n$ or $k < 0$, then α is 0.

REMARK A.3. The attentive reader might have noticed that we consistently use upper and lower indices (to denote components of vectors and forms, respectively). This helps in bookkeeping but is not essential. It also allows using Einstein's convention on repeated indices that a sum over an index is understood when the index is repeated, once as an upper index and once as a lower index. With this convention, which we will not use in this note, the last formula would, e.g., read $X(x) = X^i(x)\,\frac{\partial}{\partial x^i}$.

A.2. Definitions

The wedge product of α and β is the $(k+l)$-form

$$\alpha\wedge\beta(x) = \sum_{i_1,\dots,i_k,j_1,\dots,j_l=1}^{n} \alpha_{i_1,\dots,i_k}(x)\beta_{j_1,\dots,j_l}(x)\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_l}.$$

Notice that if $k + l > n$, then $\alpha\wedge\beta$ is automatically zero.

The differential or exterior derivative of α is the $(k+1)$-form

$$\mathrm{d}\alpha(x) = \sum_{j=1}^{n}\sum_{i_1,\dots,i_k=1}^{n} \frac{\partial}{\partial x^j}\alpha_{i_1,\dots,i_k}(x)\,\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}.$$

Notice that $\mathrm{d}x^i$ denotes at the same time the i-th basis vector of $(\mathbb{R}^n)^*$ and the differential of the coordinate function x^i. Also notice that if α is a top form, i.e., $k = n$, then automatically $\mathrm{d}\alpha = 0$.

If V is an open subset of $\mathbb{R}^m$ and ϕ a smooth map $V \to U$, the pullback of α is the k-form on V defined by

$$\phi^*\alpha(y) := \wedge^k\mathrm{d}\phi(y)^*\alpha(\phi(y)), \qquad y\in V,$$

where $\mathrm{d}\phi(y)\colon \mathbb{R}^m \to \mathbb{R}^n$ denotes the differential of ϕ at y, $\mathrm{d}\phi(y)^*$ its transpose and $\wedge^k\mathrm{d}\phi(y)^*$ the k-th exterior power of the latter. If $(y^1,\dots,y^m)$ are coordinates on V, we have

$$\phi^*\alpha(y) = \sum_{j_1,\dots,j_k=1}^{m}\sum_{i_1,\dots,i_k=1}^{n} \alpha_{i_1,\dots,i_k}(\phi(y))\,\frac{\partial\phi^{i_1}}{\partial y^{j_1}}(y)\cdots\frac{\partial\phi^{i_k}}{\partial y^{j_k}}(y)\,\mathrm{d}y^{j_1}\wedge\cdots\wedge\mathrm{d}y^{j_k}.$$

Observe that, if W is an open subset of $\mathbb{R}^s$ and ψ a smooth map $W \to V$, we have

$$(\phi\circ\psi)^* = \psi^*\phi^*.$$

The Lie derivative with respect to X of α is the k-form

$$\mathsf{L}_X\alpha = \lim_{t\to 0}\frac{\phi^*_{X,t}\alpha-\alpha}{t},$$

where $\phi_{X,t}$ is the flow of X at time t. Explicitly we have

$$\mathsf{L}_X\alpha(x) = \sum_{i_1,\dots,i_k=1}^{n} X(\alpha_{i_1,\dots,i_k})(x)\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k} +$$
$$+\sum_{i_1,\dots,i_k=1}^{n}\sum_{r,s=1}^{n}(-1)^{r-1}\alpha_{i_1,\dots,i_k}(x)\frac{\partial X^{i_r}}{\partial x^s}(x)\,\mathrm{d}x^s\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\widehat{\mathrm{d}x^{i_r}}\wedge\cdots\wedge\mathrm{d}x^{i_k},$$

where $X(\alpha_{i_1,\dots,i_k}) = \sum_{j=1}^{n} X^j \frac{\partial}{\partial x^j}\alpha_{i_1,\dots,i_k}$ denotes the directional derivative of the function $\alpha_{i_1,\dots,i_k}$ in the direction of X and the caret $\hat{}$ denotes that the factor $\mathrm{d}x^{i_r}$ is omitted.

The contraction of X with α is the $(k-1)$-form

$$\iota_X\alpha(x) = \sum_{i_1,\dots,i_k=1}^{n}\sum_{r=1}^{n}(-1)^{r-1}\alpha_{i_1,\dots,i_k}(x)X^{i_r}(x)\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\widehat{\mathrm{d}x^{i_r}}\wedge\cdots\wedge\mathrm{d}x^{i_k}.$$

If α is a function, i.e., $k=0$, then automatically $\iota_X\alpha = 0$.

A.3. Properties

The wedge product is bilinear over $\mathbb{R}^n$, whereas the differential, the pullback, the Lie derivative and the contraction are linear over $\mathbb{R}^n$. Moreover, we have the following properties:

$$\alpha\wedge\beta = (-1)^{k+l}\beta\wedge\alpha, \tag{A.1}$$
$$\mathrm{d}^2\alpha = 0, \tag{A.2}$$
$$\mathrm{d}(\alpha\wedge\beta) = \mathrm{d}\alpha\wedge\beta + (-1)^k\alpha\wedge\mathrm{d}\beta, \tag{A.3}$$
$$\phi^* f = f\circ\phi, \tag{A.4}$$
$$\phi^*(\alpha\wedge\beta) = \phi^*\alpha\wedge\phi^*\beta, \tag{A.5}$$
$$\mathrm{d}\phi^*\alpha = \phi^*\mathrm{d}\alpha, \tag{A.6}$$
$$\mathsf{L}_X f = X(f), \tag{A.7}$$
$$\mathsf{L}_X(\alpha\wedge\beta) = \mathsf{L}_X\alpha\wedge\beta + \alpha\wedge\mathsf{L}_X\beta, \tag{A.8}$$
$$\mathsf{L}_X\mathrm{d}\alpha = \mathrm{d}\mathsf{L}_X\alpha, \tag{A.9}$$
$$\iota_X(\alpha\wedge\beta) = \iota_X\alpha\wedge\beta + (-1)^k\alpha\wedge\iota_X\beta, \tag{A.10}$$
$$L_X\alpha = \iota_X\mathrm{d}\alpha + \mathrm{d}\iota_X\alpha \qquad \text{(Cartan's formula)}. \tag{A.11}$$

Observe that the above properties characterize d, ϕ^*, L_X and ι_X completely. If Y is a second vector field, we also have

$$\iota_X \iota_Y \alpha = -\iota_Y \iota_X \alpha, \tag{A.12}$$

$$\mathsf{L}_X \mathsf{L}_Y \alpha - \mathsf{L}_Y \mathsf{L}_X \alpha = \mathsf{L}_{[X,Y]} \alpha, \tag{A.13}$$

$$\iota_X \mathsf{L}_Y \alpha - \mathsf{L}_Y \iota_X \alpha = \iota_{[X,Y]} \alpha, \tag{A.14}$$

where $[X, Y]$ is the Lie bracket of X and Y defined by

$$[X, Y](f) = X(Y(f)) - Y(X(f)).$$