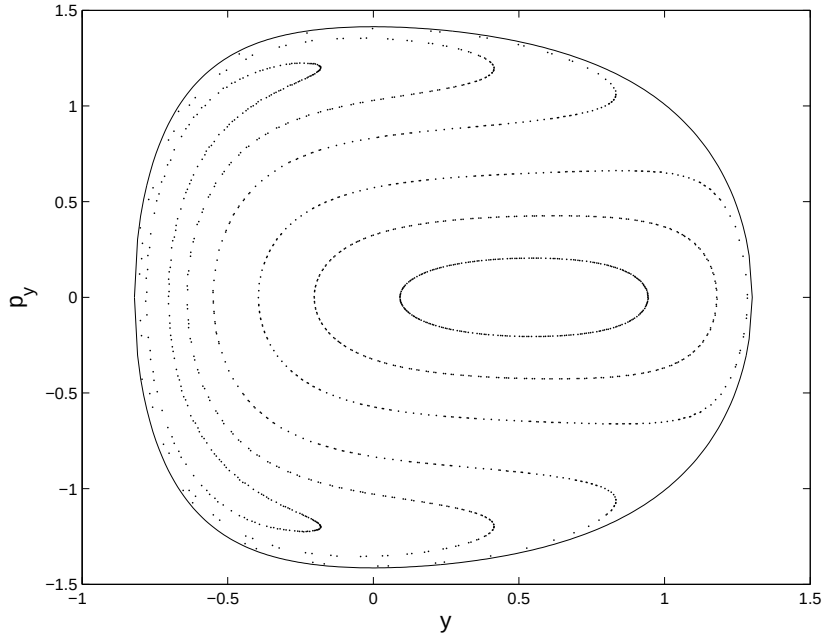
Classical Mechanics and Symplectic Integration

Nordkvist, Nikolaj; Hjorth, Poul G.

Publication date:
2005

Document Version
Publisher's PDF, also known as Version of record

Classical Mechanics and Symplectic Integration



Lecture Notes

Poul G. Hjorth Nikolaj Nordkvist

Contents

Chapter 1

Preliminaries

We start by recaling some usefull theory, that is some results from multivariable calculus and the concept of a Poincaré map.

Elements of Multivariable Calculus

In the theory of analytical mechanics two main theorems are frequently used, namely the chain rule and the inverse function theorem. Therefore we start be recalling these basic mathematical facts.

Throughout this text we will denote a vector $\boldsymbol{x} \in \mathbb{R}^n$ by boldface and it's elements by x_i, i.e. $\boldsymbol{x} = (x_1, \ldots, x_n)$. For a smooth function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x})$, that is $\boldsymbol{f} \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$, the Jacobian matrix $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$ is the $m \times n$ matrix given by[1]

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

With this the chain rule can be given as

Theorem 1 (The chain rule)**.** *For the smooth functions $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x})$, and $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^k$, $g = g(\boldsymbol{y})$, the Jacobian of $\boldsymbol{g} \circ \boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^k$ is given by*

$$\frac{\partial \boldsymbol{g} \circ \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}) = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}(\boldsymbol{f}(\boldsymbol{x}))\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x})$$

Example 1. Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$, $\boldsymbol{f} = \boldsymbol{f}(t)$, and $g : \mathbb{R}^n \to \mathbb{R}$, $g = g(\boldsymbol{x})$ then we

[1] Often, when there is no risk of confusion, we will for functions $f : \mathbb{R}^n \to \mathbb{R}$ also use the notation $\frac{\partial f}{\partial \boldsymbol{x}}$ for it's transpose, i.e. the gradient of f.

have by the chain rule that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} g \circ \boldsymbol{f}(t) &= \frac{\partial g}{\partial \boldsymbol{x}}(\boldsymbol{f}(t)) \frac{\mathrm{d}\boldsymbol{f}}{\mathrm{d}t}(t) \\ &= \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\boldsymbol{f}(t)) \dot{f}_i(t) \\ &= \nabla g(\boldsymbol{f}(t)) \cdot \dot{\boldsymbol{f}}(t)\end{aligned}$$

where the dot '$\cdot$' is the usual Euclidian scalar product.

If a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is bijective (one-to-one and onto) with inverse $\boldsymbol{f}^{-1}$, and if both $\boldsymbol{f}$ and $\boldsymbol{f}^{-1}$ are smooth, then $\boldsymbol{f}$ is said to be a diffeomorphism. A diffeomorphism is also sometimes refered to as a coordinate change, since it can be used to define new coordinates in which every function will be as many times differentiable as in the old coordinates. We will often specify a coordinate transformation $\psi : \mathbb{R}^n \to \mathbb{R}^n$ by $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$ to state explicitly that it takes the x-coordinates and transforms to the y-coordinates, and ψ^{-1} vice versa. The inverse function gives sufficient conditions on a map $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ to be a diffeomorphism.

Theorem 2 (Inverse function theorem)**.** *Let $U \subset \mathbb{R}^n$ be open, and let $\boldsymbol{f} : U \to \mathbb{R}^n$, $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x})$, be smooth. If $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$ is non-singular at $\boldsymbol{x}_0 \in U$, i.e.*

$$\det\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}_0)\right) \neq 0$$

then there exists an open set V with $\boldsymbol{x}_0 \in V \subset U$ such that $\boldsymbol{f}|_V$ is a diffeomorphism.

Example 2. If $g : \mathbb{R}^{2n} \to \mathbb{R}$, $g = g(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, is smooth and if

$$\det\left(\frac{\partial^2 g}{\partial \boldsymbol{y} \partial \boldsymbol{y}}\right) \neq 0$$

where $\frac{\partial^2 g}{\partial \boldsymbol{y} \partial \boldsymbol{y}}$ is the Jacobian matrix of $\frac{\partial g}{\partial \boldsymbol{y}}$, then we know by the inverse function theorem, that it is possible to define a coordinate transformation $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $(\boldsymbol{x}, \boldsymbol{y}) \leftrightarrow (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$, as

$$(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \left(\boldsymbol{x}, \frac{\partial g}{\partial \boldsymbol{y}}(\boldsymbol{x}, \boldsymbol{y})\right)$$

This transformation is important in the theory of analytical mechanics and is called the **Legendre transform**.

A more general result is the following

Theorem 3 (Implicit function theorem)**.** *Let $U \subset \mathbb{R}^{n-m} \times \mathbb{R}^m$ be open, and let $\boldsymbol{f} : U \to \mathbb{R}^m$ be smooth. Denote the Cartesian coordinate system on $\mathbb{R}^{n-m} \times \mathbb{R}^m$ by $(\boldsymbol{x}, \boldsymbol{y}) = (x_1, \ldots, x_{n-m}, y_1, \ldots, y_m)$. Suppose that at the point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in U$*

$$\boldsymbol{f}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0$$

and the matrix $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}$ is nonsingular, i.e.

$$\det\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}(\boldsymbol{x}_0, \boldsymbol{y}_0)\right) \neq 0$$

Then there exists an open neighborhood V of $\boldsymbol{x}_0$ in $\mathbb{R}^{n-m}$ and an open neighborhood W of $\boldsymbol{y}_0$ in $\mathbb{R}^m$ such that $V \times W \subset U$, and there exists a smooth map $\boldsymbol{g} : V \to W$ such that for each $(\boldsymbol{x}, \boldsymbol{y}) \in V \times W$

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = 0 \qquad \Leftrightarrow \qquad \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$$

The inverse function theorem can be deduced from the implicit function theorem as a corrolary.

Poincaré Maps

For a general autonomous differential equation $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$, a **surface of section** S is the image of a map $\boldsymbol{g} : U \to \mathbb{R}^n$, $U \subset \mathbb{R}^{n-1}$, e.i. $S = \boldsymbol{g}(U)$,[2] such that $\boldsymbol{f}$ is transversal to it, that is the flow $\boldsymbol{f}(\boldsymbol{x})$ intersects S only in points, not lines. Let φ_t be the flow of $\boldsymbol{f}$, $\Gamma = \varphi_{[0,\infty)}(\boldsymbol{p})$ the trajectory from $\boldsymbol{p} \in \mathbb{R}^n$ — assuming that $\varphi_t(\boldsymbol{p})$ is defined for all $t \in [0, \infty]$ — and $\boldsymbol{x}_k$ the k-th intersection, in a particular sence (e.g. from right to left) of Γ and S, $\boldsymbol{x}_{k+1} = \varphi_\tau(\boldsymbol{x}_k)$ for some $\tau \in (0, \infty)$, then the **Poincaré map** $\sigma : S \to S$ is defined as the mapping giving

$$\sigma(\boldsymbol{x}_k) = \boldsymbol{x}_{k+1}$$

Studying the Poincaré map of a differential equation can simplify some of the qualitative analysis. Consider a differential equation in $\mathbb{R}^3$ and the Poincaré map σ given by considering succesive intersections with a plane, from one specific side to the other. If σ has a fixed point we know that the differential equation has a closed orbit, and if $\boldsymbol{x}_1, \boldsymbol{x}_2 = \sigma(\boldsymbol{x}_1), \boldsymbol{x}_3 = \sigma^2(\boldsymbol{x}_1), \ldots$ lies on a closed curve we know that near the plane of section the motion takes place on something similar to a cylinder.

[2]This difinition slightly differs from the standerd one, where S is a $n-1$ dimensional submanifold of $\mathbb{R}^n$. So we see that our definition is contained in the "correct" one.

Chapter 2

Lagrangian Mechanics

2.1 Calculus of Variations

A functional is a mapping from a vector space to the real numbers. Let $L : \mathbb{R}^{n+1} \to \mathbb{R}$ be a smooth function, called the **Lagrangian**, and consider a smooth curve $\gamma : [t_0, t_1] \to \mathbb{R}^n$, then we define the functional I as

$$I(\gamma) = \int_{t_0}^{t_1} L(\gamma(t), \dot{\gamma}(t), t)\mathrm{d}t$$

The number n is refered to as the number of degrees of freedom, for reasons which will become clear later. The **variation** δI of I is for a smooth curve $\eta : [t_0, t_1] \to \mathbb{R}^n$ defined as

$$\delta I(\gamma, \eta) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} I(\gamma + s\eta) \tag{2.1}$$

We see that a necessary condition for γ to be an ekstremum of I, i.e. $I(\gamma) \leq \lim_{s\to 0} I(\gamma+s\eta)$ or $I(\gamma) \geq \lim_{s\to 0} I(\gamma+s\eta)$ for all curves η, is that $\delta I(\gamma, \cdot) = 0$. If we wish to find the curves starting at $\boldsymbol{q}^0$ and ending at $\boldsymbol{q}^1$ that are extremals of I we therefore calculate 2.1 with the condition on η that $\eta(t_0) = \eta(t_1) = 0$, which gives

$$\begin{aligned}
\delta I(\gamma, \eta) &= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) \cdot \eta(t) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \dot{\eta}(t) \right) \mathrm{d}t \\
&= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}}(\gamma, \dot{\gamma}(t), t) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \right) \cdot \eta(t)\mathrm{d}t + \left[\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \cdot \eta \right]_{t_0}^{t_1} \\
&= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \right) \cdot \eta(t)\mathrm{d}t
\end{aligned} \tag{2.2}$$

where $\boldsymbol{q} = (q_1, \ldots, q_n)$, $\dot{\boldsymbol{q}} = (\dot{q}_1, \ldots, \dot{q}_n)$ are Euclidian coordinates and $L = L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$. In order to proceed we need the so called fundamental lemma of the calculus of variations

Lemma 1. *Let* $\boldsymbol{g} \in C^0([t_0, t_1], \mathbb{R}^n)$*, then*

$$\int_{t_0}^{t_1} \boldsymbol{g}(t) \cdot \boldsymbol{h}(t) \mathrm{d}t = 0 \quad \forall \quad \boldsymbol{h} \in C^\infty([t_0, t_1], \mathbb{R}^n)$$

if and only if

$$\boldsymbol{g} = 0$$

Proof. The if part is obvious. The only if part we show by contradiction, so assume that $\int_{t_0}^{t_1} \boldsymbol{g}(t) \cdot \boldsymbol{h}(t) \mathrm{d}t = \int_{t_0}^{t_1} \sum_{i=1}^n g_i(t) h_i(t) \mathrm{d}t = 0$ for some $g_{i_0} \neq 0$. Then there exists $\tau \in (t_0, t_1)$ such that $g_{i_0}(\tau) \neq 0$. Since g_{i_0} is continous there exists $\delta > 0$ such that $g_{i_0}(t) > \frac{1}{2} g_{i_0}(\tau)$ for $t \in (\tau - \delta, \tau + \delta)$. Then choose h smooth such that

$$\begin{aligned} h_{i_0}(t) &= 0 \quad \text{for } t \in [t_0, \tau - \delta] \cup [\tau + \delta, t_1] \\ h_{i_0}(t) &> 0 \quad \text{for } t \in (\tau - \delta, \tau + \delta) \\ h_i(t) &= 0 \quad \text{for } i \neq i_0,\ t \in [t_0, t_1] \end{aligned}$$

But then we get

$$\int_{t_0}^{t_1} \sum_{i=1}^n g_i(t) h_i(t) \mathrm{d}t = \int_{\tau - \delta}^{\tau + \delta} g_{i_0}(t) h_{i_0}(t) \mathrm{d}t > \delta g_{i_0}(\tau)$$

which is a contradiction. □

Using this lemma we thus get the following

Theorem 4. *If a curve* $\gamma \in C^\infty([t_0, t_1], \mathbb{R}^n)$ *with* $\gamma(t_0) = \boldsymbol{q}^0$ *and* $\gamma(t_1) = \boldsymbol{q}^1$ *is an extremum of* I *among the curves satisfying these boundary conditions then it satisfies the equations*

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) - \frac{\partial L}{\partial \boldsymbol{q}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = 0$$

These equations are called the **Euler-Lagrange equations**. If we use the chain rule the Euler-Lagrange equations become

$$\sum_{j=1}^n \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_j + \sum_{j=1}^n \frac{\partial^2 L}{\partial \dot{q}_i \partial q_j} \dot{q}_j + \frac{\partial^2 L}{\partial \dot{q}_i \partial t} - \frac{\partial L}{\partial q_i} = 0 \qquad i = 1, \ldots, n$$

So we see that the Euler-Lagrange equations are a set of coupled second order differential equations. From the classical theorem on the existence and uniqueness of solutions to a differential equation we see that a sufficient condition to ensure the existence and uniqueness of a solution to the Euler-Lagrange equations is

$$\det \left(\frac{\partial^2 L}{\partial \dot{\boldsymbol{q}} \partial \dot{\boldsymbol{q}}} \right) \neq 0$$

Example 3. In 1696 John Bernoulli posed the following problem, which was later solved by John Bernoulli, James Bernoulli, Newton, and L'Hospital, and played an important part in the development of the theory of the calculus of variations.

Let A and B be fixed points in a vertical plane, and assume that B is lower than A. Let a particle slide without friction along a curve joining A and B, then the time it takes to reach B from A is a functional of the curve, and the curve which takes the least time is called the brachistochrone, and can be calculated using the Euler-Lagrange equations as follows:

We choose normal Euclidian x-y coordinates in the plane, where gravity acts in the positive direction of the y-axis, and assume for simplicity that the particle starts at rest and that A is the origin of coordinates. Letting the curve be a function of x we get

$$v = \sqrt{1 + (y')^2}\frac{\mathrm{d}x}{\mathrm{d}t}$$

and due to conservation of energy we have that

$$v = \sqrt{2gy}$$

where g is the gravitational accelleration. The transit time T of a curve y is therefore given by

$$T(y) = \int_{t_{start}}^{t_{end}} \mathrm{d}t = \int_0^{x_1} \frac{\sqrt{1 + (y')^2}}{\sqrt{2gy}}\mathrm{d}x$$

So the Lagrangian is given by

$$L(y, y') = \frac{\sqrt{1 + (y')^2}}{\sqrt{2gy}}$$

and the Euler-Lagrange equation for this problem becomes

$$2y''(x)y(x) + (y'(x))^2 + 1 = 0$$

We notice that this equation is independent of the values of g and the mass of the particle.

It turns out that the solution to the Euler-Lagrange equation is a family of cycloids

$$x = r(\theta - \sin\theta) \qquad\qquad y = r(1 - \cos\theta)$$

parametrized by θ.

Geodesic Curves

Consider a curve γ in $U \subset \mathbb{R}^n$. The length $\mathcal{L}$ of a curve in U is the functional given by

$$\mathcal{L}(\gamma) = \int_{t_0}^{t_1} \|\dot{\gamma}(t)\| \mathrm{d}t$$

The curves which minimize length are called geodesic curves.

If we define the functional $\mathcal{E}$ as

$$\mathcal{E}(\gamma) = \frac{1}{2} \int_{t_0}^{t_1} \|\dot{\gamma}(t)\|^2 \mathrm{d}t$$

we get using Schwartz' inequality

$$\begin{aligned} \mathcal{L}(\gamma) &= \int_{t_0}^{t_1} 1 \cdot \|\dot{\gamma}\| \mathrm{d}t \\ &\leq \left(\int_{t_0}^{t_1} \mathrm{d}t \right)^{1/2} \left(\int_{t_0}^{t_1} \|\dot{\gamma}(t)\|^2 \mathrm{d}t \right)^{1/2} \leq \sqrt{2(t_1 - t_0)} \sqrt{\mathcal{E}(\gamma)} \end{aligned}$$

where we have equality if and only if $\|\dot{\gamma}(t)\| = \textit{constant}$. So when minimizing $\mathcal{E}$ we find a geodesic curve with constant "speed".

Example 4. Consider $U = \mathbb{R}^n$ with coordinates $\boldsymbol{x} = (x_1, \ldots, x_n)$. We have that the Lagrangian for $\mathcal{E}$ is

$$L(\boldsymbol{x}, \dot{\boldsymbol{x}}) = \frac{1}{2} \|\dot{\boldsymbol{x}}\|^2 = \frac{1}{2} (\dot{x}_1^2 + \ldots + \dot{x}_n^2)$$

So according to the Euler-Lagrange equations we get that the geodesic curves satisfies

$$\ddot{\boldsymbol{x}} = 0$$

which means that the curves in $\mathbb{R}^n$ with minimal length are

$$\gamma(t) = \boldsymbol{a}t + \boldsymbol{b}$$

i.e. straight lines, a well known fact.

Another instructive example is the following

Example 5. Consider the sphere $S^2 \subset \mathbb{R}^3$. In $\mathbb{R}^3$ we can choose spherical coordinates (θ, ϕ, r) and in these coordinates S^2 will be given by

$$\begin{aligned} x &= \sin\theta \cos\phi \\ y &= \sin\theta \sin\phi \\ z &= \cos\theta \end{aligned}$$

Therefore on S^2 we have that the Lagrangian for the energy is

$$\begin{aligned} L(\theta,\phi,\dot\theta,\dot\phi) &= \frac{1}{2}\|\dot{\boldsymbol{x}}|_{S^2}\|^2 \\ &= \frac{1}{2}\left(\dot x|_{S^2}^2 + \dot y|_{S^2}^2 + \dot z|_{S^2}^2\right) \\ &= \frac{1}{2}\big((\dot\theta\cos\theta\cos\phi - \dot\phi\sin\theta\sin\phi)^2 \\ &\quad + (\dot\theta\cos\theta\sin\phi + \dot\phi\sin\theta\cos\phi)^2 + (-\dot\theta\sin\theta)^2\big) \\ &= \frac{1}{2}(\dot\theta^2 + \dot\phi^2\sin^2\theta) \end{aligned}$$

and this is thus the Lagrangian determining the geodesic curves on S^2 via the Euler-Lagrange equations.

If we have a smooth coordinate transformation $\psi : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{q} \mapsto \tilde{\boldsymbol{q}}(\boldsymbol{q})$, then a smooth curve in the q-coordinates must be a smooth curve in the $\tilde q$-coordinates and vice versa. Thus it seems that if a curve in the q-coordinates is an ekstremum for I then it must also be so in the $\tilde q$-coordinates. The following result would be a direct consequence of such a result.

Proposition 1. *The Euler-Lagrange equations are invariant under a smooth coordinate transformation $\psi : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{q} \leftrightarrow \tilde{\boldsymbol{q}}(\boldsymbol{q})$. That is if $\gamma : [t_0, t_1] \to \mathbb{R}^n$, $\tilde\gamma(t) = \psi(\gamma(t))$, and $L = \widetilde{L} \circ \psi$ then we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t),\dot\gamma(t),t) - \frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t),\dot\gamma(t),t) = 0$$

if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \widetilde{L}}{\partial \dot{\tilde{\boldsymbol{q}}}}(\tilde\gamma(t),\dot{\tilde\gamma}(t),t) - \frac{\partial \widetilde{L}}{\partial \tilde{\boldsymbol{q}}}(\tilde\gamma(t),\dot{\tilde\gamma}(t),t) = 0$$

Proof. From the chain rule we have

$$\dot{\tilde q}_j = \sum_{k=1}^{n}\frac{\partial \tilde q_j}{\partial q_k}\dot q_k \qquad\qquad \Rightarrow \qquad\qquad \frac{\partial \dot{\tilde q}_j}{\partial \dot q_i} = \frac{\partial \tilde q_j}{\partial q_i}$$

Thus we get using the chain rule on $L(\boldsymbol{q},\dot{\boldsymbol{q}},t) = \widetilde{L}(\tilde{\boldsymbol{q}}(\boldsymbol{q}),\dot{\tilde{\boldsymbol{q}}}(\boldsymbol{q},\dot{\boldsymbol{q}}),t)$

$$\frac{\partial L}{\partial \dot q_i} = \sum_{j=1}^{n}\frac{\partial \widetilde{L}}{\partial \dot{\tilde q}_j}\frac{\partial \dot{\tilde q}_j}{\partial \dot q_i} = \sum_{j=1}^{n}\frac{\partial \widetilde{L}}{\partial \dot{\tilde q}_j}\frac{\partial \tilde q_j}{\partial q_i}$$

giving

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot q_i}\right) = \sum_{j=1}^{n}\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \widetilde{L}}{\partial \dot{\tilde q}_j}\right)\frac{\partial \tilde q_j}{\partial q_i} + \sum_{j=1}^{n}\frac{\partial \widetilde{L}}{\partial \dot{\tilde q}_j}\sum_{k=1}^{n}\frac{\partial^2 \tilde q_j}{\partial q_i \partial q_k}\dot q_k$$

The second part of the Euler-Lagrange equations is similarly calculated using the chain rule

$$\begin{aligned}\frac{\partial L}{\partial q_i} &= \sum_{j=1}^{n} \frac{\partial \widetilde{L}}{\partial \tilde{q}_j}\frac{\partial \tilde{q}_j}{\partial q_i} + \sum_{j=1}^{n} \frac{\partial \widetilde{L}}{\partial \dot{\tilde{q}}_j}\frac{\partial \dot{\tilde{q}}_j}{\partial q_i} \\ &= \sum_{j=1}^{n} \frac{\partial \widetilde{L}}{\partial \tilde{q}_j}\frac{\partial \tilde{q}_j}{\partial q_i} + \sum_{j=1}^{n} \frac{\partial \widetilde{L}}{\partial \dot{\tilde{q}}_j}\sum_{k=1}^{n} \frac{\partial^2 \tilde{q}_j}{\partial q_i \partial q_k}\dot{q}_k\end{aligned}$$

Combining these we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \frac{\partial \tilde{q}_j}{\partial q_i}\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \widetilde{L}}{\partial \dot{\tilde{q}}_j} - \frac{\partial \widetilde{L}}{\partial \tilde{q}_j}\right)$$

We notice that $\frac{\partial \tilde{q}_j}{\partial q_i}$ are the elements of the Jacobian matrix $\frac{\partial \tilde{\boldsymbol{q}}}{\partial \boldsymbol{q}}$ for ψ and we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} - \frac{\partial L}{\partial \boldsymbol{q}} = \frac{\partial \tilde{\boldsymbol{q}}}{\partial \boldsymbol{q}}^T\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \widetilde{L}}{\partial \dot{\tilde{\boldsymbol{q}}}} - \frac{\partial \widetilde{L}}{\partial \tilde{\boldsymbol{q}}}\right) \tag{2.3}$$

Since ψ is a coordinate transformation $\frac{\partial \tilde{\boldsymbol{q}}}{\partial \boldsymbol{q}}$ is nonsingular and therefore we get the desired result. □

Problems that are invariant under smooth coordinate transformations can typically be stated in a differential geometric setting which makes it possible to work with the problem in a coordinate free way giving many powerfull tools for analysis.

2.2 Lagrange's Equations

Consider a mechanical system of k point masses m_i. If we by $\boldsymbol{x}_j = (x_j, y_j, z_j)$ denote the Cartesian coordinates of the i-th point mass, the entire system can be described by $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{R}^{3k}$, i.e. by $3k$ coordinates. Throughout this monograph we will only consider conservative systems which are mechanical systems for which the total force on the elements in the system can be derived from a potential $V : \mathbb{R}^{3k} \to \mathbb{R}$ according to

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{x}_i}$$

so newtons second law gives

$$m_i \ddot{\boldsymbol{x}}_i = -\frac{\partial V}{\partial \boldsymbol{x}_i}$$

If we for a mechanical system of point masses define the Lagrangian as

$$L = \sum_{i=1}^{k} \frac{1}{2} m_i \|\dot{\boldsymbol{x}}_i\|^2 - V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$$

that is the kinetic energy K minus the potential energy V, the Euler-Lagrange equations for this system becomes

$$m_i \ddot{\boldsymbol{x}}_i = -\frac{\partial V}{\partial \boldsymbol{x}_i}$$

so we see that Newtons laws and Lagrange's equations with $L = K - V$ for k point masses are equivalent.

Due to proposition 1 we know that we can choose arbitrary coordinates for the analysis of the system. Consider a mechanical system of point masses described locally in $V \subset \mathbb{R}^d$ by the coordinates $\boldsymbol{z} = (z_1, \ldots, z_d)$, not necessarily Cartesian coordinates, and assume that the potential is given by

$$V = V_1(z_1, \ldots, z_d) + \frac{1}{2}\frac{1}{\epsilon}(z_1 - z_d)^2$$

for some $\epsilon > 0$. Since the solution to a differential equation depending continously on a parameter will depend continously on that parameter the solution of Euler-Lagrange's equation for this potential will depend continuosly on ϵ. For ϵ very small we will have $z_d \approx z_1$, when $z_d(0) = z_1(0)$, so effectively in the equations the z_d coordinate can be considered a constant and the Euler-Lagrange equations will still describe the motion. So if we examine a mechanical system of point masses with a potential which to a good approximation gives the constraints

$$\dot{z}_{d-c+1} = 0 \quad \ldots \quad \dot{z}_d = 0$$

then we expect the motion of the system still to be expressed by the Euler-Lagrange equations, where $z_{d-c+1}, \ldots, z_d$ are just constants. The number $n = c - d$ is called the degrees of freedom of the system since this is the least number of coordinates which is needed to specify the configuration of the system. The coordinates $\boldsymbol{q} = (q_1, \ldots, q_n) = (z_1, \ldots, z_n) \in U \subset \mathbb{R}^n$ are called generalised coordinates, and with

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = L(q_1, \ldots, q_n, z_{d-c+1}, \ldots, z_d, \dot{q}_1, \ldots, \dot{q}_n, 0, \ldots, 0)$$

the motion of the system is thus determined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) - \frac{\partial L}{\partial \boldsymbol{q}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = 0$$

These equations are called **Lagrange's equations** — so basically Lagrange's equations are the Euler-Lagrange equations for a mechanical system with Lagrangian equal to the kinetic- minus the potential energy.

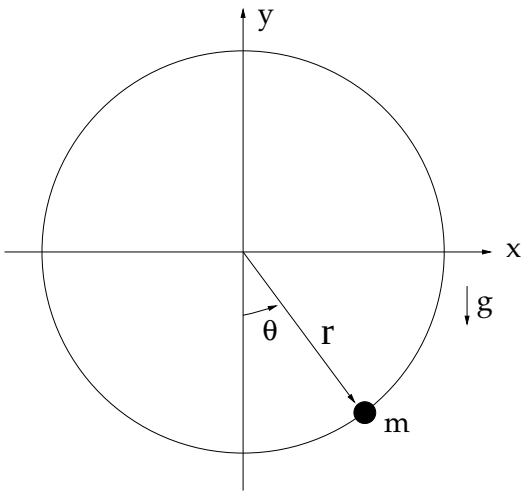


Figure 2.1: The pendulum.

Example 6. For a pendulum, see figure 2.1, we have, due to the fact that the length ℓ of the pendulum approximative is constant, that in cartesian coordinates (x, y)

$$x^2 + y^2 = \ell^2$$

If we instead choose polar coordinates (r, θ), where $\theta = 0$ in the vertical downward position, we have $\dot{r} = 0$ and the kinetic energy K and the potential energy V is given by

$$K = \frac{1}{2} m \ell^2 \dot{\theta}^2 \qquad\qquad V = -mg\ell\cos(\theta)$$

where m is the mass and g is the gravitational accelleration. From this the motion is determined by Lagrange's equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = 0$$

giving

$$\ddot{\theta} = -g\ell^{-1}\sin(\theta)$$

This equation would in the Newtonian framework have been deduced from angular momentum considerations and not Newton's second law directly, but in the Lagrangian framework everything is deduced from Lagrange's equations alone.

2.3 Symmetries and Conservation Laws

If the Lagrangian L is independent of q_i then q_i is called a cyclic coordinate, and from the Euler-Lagrange equations we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i}$$

so if q_i is a cyclic coordinate, then $\frac{\partial L}{\partial \dot{q}_i}$ is a constant of the motion.

The energy $E : \mathbb{R}^{2n+1} \to \mathbb{R}$ is defined, using the Lagrangian, as

$$E(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) := \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \cdot \dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$$

Letting $\gamma(t)$ be a solution to the Euler-Lagrange equations we get using the chain rule and the Euler-Lagrange equations

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}E(\gamma(t), \dot{\gamma}(t), t) &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \dot{\gamma}(t) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \ddot{\gamma}(t) \\
&\quad - \frac{\mathrm{d}}{\mathrm{d}t}L(\gamma(t), \dot{\gamma}(t), t) \\
&= \frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) \cdot \dot{\gamma}(t) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \ddot{\gamma}(t) \\
&\quad - \left(\frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) \cdot \dot{\gamma}(t) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \ddot{\gamma}(t) \right. \\
&\qquad \left. + \frac{\partial L}{\partial t}(\gamma(t), \dot{\gamma}(t), t) \right) \\
&= -\frac{\partial L}{\partial t}(\gamma(t), \dot{\gamma}(t), t)
\end{aligned}$$

So if the Lagrangian is independent of t the energy is a constant of the motion.

The fact that conservation of momentum or angular momentum of a mechanical system follows from invariance of the problem with respect to translation or rotation respectively is a consequence of a more general result originally showed by E. Noether.

Theorem 5 (Noether's theorem)**.** *Let $L : \mathbb{R}^{2n+1} \to \mathbb{R}$ be smooth and assume there exists a one-parameter family of smooth maps $\boldsymbol{h}_s : \mathbb{R}^n \to \mathbb{R}^n$ smooth in $s \in (-\epsilon, \epsilon)$, $\epsilon > 0$, and with $\boldsymbol{h}_0 = \mathrm{id}$. If*

$$L\left(\boldsymbol{h}_s(\gamma(t)), \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{h}_s(\gamma(t)), t\right) = L(\gamma(t), \dot{\gamma}(t), t)$$

for all $s \in (-\epsilon, \epsilon)$ and all smooth curves $\gamma : [t_0, t_1] \to \mathbb{R}^n$. Then for any solution of the Euler-Lagrange equations the function

$$F(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \cdot \left.\frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\right|_{s=0}(\boldsymbol{q})$$

is a constant of the motion.

Proof. By assumption we have that

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} L\left(\boldsymbol{h}_s(\gamma(t)), \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{h}_s(\gamma(t)), t\right) \\ &= \frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) \cdot \frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\bigg|_{s=0}(\gamma(t)) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}\,\frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\bigg|_{s=0}(\gamma(t)) \end{aligned}$$

Let $\gamma(t)$ be a solution to the Euler-Lagrange equations, i.e. we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) - \frac{\partial L}{\partial \boldsymbol{q}}(\gamma(t), \dot{\gamma}(t), t) = 0$$

Inserting this we get

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\bigg|_{s=0}(\gamma(t)) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}\,\frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\bigg|_{s=0}(\gamma(t)) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\gamma(t), \dot{\gamma}(t), t) \cdot \frac{\mathrm{d}\boldsymbol{h}_s}{\mathrm{d}s}\bigg|_{s=0}(\gamma(t))\right) \end{aligned}$$

which shows that F is a constant of the motion. □

Example 7. Consider a mechanical system of n point masses m_j and coordinates $\boldsymbol{q}_j = (x_j, y_j, z_j)$

$$L = \sum_{j=1}^{n} \frac{1}{2} m_j \|\dot{\boldsymbol{q}}_j\|^2 - V(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$$

Assume that the system is invariant under a translation along the x-axis, i.e.

$$\boldsymbol{h}_s(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) = (\boldsymbol{q}_1 + se_x, \ldots, \boldsymbol{q}_n + se_x)$$

where e_x is a unit vector in the x-direction. Then we get that

$$\sum_{j=1}^{n} m_j \dot{x}_j$$

which is the total momentum along the x-axis, is conserved by the flow. From proposition 1 we thus know that a mechanical system of point masses which is invariant under translation in some direction has total momentum along this axis as a constant of the motion.

A system consisting of two point masses with a potential given by

$$V(\boldsymbol{q}_1, \boldsymbol{q}_2) = V(\|\boldsymbol{q}_1 - \boldsymbol{q}_2\|)$$

is invariant under a translation along any axis and therefore the total momentum in any direction is conserved. Examples of such systems include two masses interacting gravitationally, e.g. the sun and a planet, and two charges interacting due to Coulomb's law.

Chapter 3

Hamiltonian Mechanics

3.1 Hamilton's Equations

Given a smooth Lagrangian $L : \mathbb{R}^{2n+1} \to \mathbb{R}$ the smooth map $(\boldsymbol{q}, \dot{\boldsymbol{q}}) \mapsto \mathbb{R}^n$ given by

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$$

gives the Legendre transformation $(\boldsymbol{q}, \dot{\boldsymbol{q}}) \leftrightarrow (\boldsymbol{q}, \boldsymbol{p})$ if

$$\det\left(\frac{\partial^2 L}{\partial \dot{\boldsymbol{q}} \partial \dot{\boldsymbol{q}}}\right) \neq 0$$

which insures according to the inverse function theorem that $(\boldsymbol{q}, \dot{\boldsymbol{q}}) \leftrightarrow (\boldsymbol{q}, \boldsymbol{p})$ indeed locally defines a smooth coordinate transformation.[1]

Remark 1. Often for mechanical systems the potential energy is a function of the configuration only and the kinetic energy K is given by a quadratic form, i.e.

$$K = \frac{1}{2}\dot{\boldsymbol{q}}^T \mathrm{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(\boldsymbol{q})\dot{q}_i\dot{q}_j$$

where G is a positive definite and symmetric $n \times n$ matrix, possibly depending on $\boldsymbol{q}$, with elements g_{ij}. Since G is positive definite it's determinant is strictly positive and we get

$$\det\left(\frac{\partial^2 L}{\partial \dot{\boldsymbol{q}} \partial \dot{\boldsymbol{q}}}\right) = \det\left(\frac{\partial^2 K}{\partial \dot{\boldsymbol{q}} \partial \dot{\boldsymbol{q}}}\right) = \det(\mathrm{G}) \neq 0$$

So for these systems the Legendre transformation does exist.

[1] If L depends on t we know from the implicit function theorem that the Legendre transformation is smooth in t.

The Legendre transformation makes it possible to define the **Hamiltonian** which is the smooth function $H : \mathbb{R}^{2n+1} \to \mathbb{R}$ given by

$$H(\boldsymbol{p}, \boldsymbol{q}, t) = \sum_{j=1}^{n} p_j \dot{q}_j - L(\boldsymbol{q}, \dot{\boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{p}, t), t)$$

Remark 2. Consider a mechanical system with potential energy a function of the configuration only and the kinetic energy K given by a quadratic form G

$$K = \frac{1}{2}\dot{\boldsymbol{q}}^T \mathrm{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(\boldsymbol{q})\dot{q}_i\dot{q}_j$$

The Legendre transformation gives

$$\begin{aligned} \boldsymbol{p} &= \mathrm{G}(\boldsymbol{q})\dot{\boldsymbol{q}} & \qquad \dot{\boldsymbol{q}} &= \mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} \\ p_i &= \sum_{j=1}^{n} g_{ij}(\boldsymbol{q})\dot{q}_j & \qquad \dot{q}_i &= \sum_{j=1}^{n} g^{ij}(\boldsymbol{q})p_j \end{aligned}$$

where g^{ij} are the elements of the matrix G^{-1}, which is also positive definite and symmetric. Then,

$$\begin{aligned} K &= \frac{1}{2}\boldsymbol{p}^T \mathrm{G}^{-T}(\boldsymbol{q})\mathrm{G}(\boldsymbol{q})\mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} \\ &= \frac{1}{2}\boldsymbol{p}^T \mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} \end{aligned}$$

and

$$\begin{aligned} H &= \boldsymbol{p}^T \mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} - \frac{1}{2}\boldsymbol{p}^T \mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} + V(\boldsymbol{q}) \\ &= \frac{1}{2}\boldsymbol{p}^T \mathrm{G}^{-1}(\boldsymbol{q})\boldsymbol{p} + V(\boldsymbol{q}) \\ &= K(\boldsymbol{p}, \boldsymbol{q}) + V(\boldsymbol{q}) \end{aligned}$$

Thus we see that for such a system the Hamiltonian is equal to the total energy of the system.

How the dynamics of the Euler-Lagrange equations is expressed in the coordianates $\boldsymbol{q}$ and $\boldsymbol{p}$ is given by the following

Theorem 6. *The Euler-Lagrange equations are equivalent to the following equations*

$$\begin{aligned} \dot{\boldsymbol{p}} &= -\frac{\partial H}{\partial \boldsymbol{q}} \\ \dot{\boldsymbol{q}} &= \frac{\partial H}{\partial \boldsymbol{p}} \end{aligned}$$

These equations are called ***Hamilton's equations****.*

Proof. The differential of H is given by

$$\mathrm{d}H = \sum_{j=1}^{n} \left(\frac{\partial H}{\partial p_j} \mathrm{d}p_j + \frac{\partial H}{\partial q_j} \mathrm{d}q_j \right) + \frac{\partial H}{\partial t} \mathrm{d}t \tag{3.1}$$

If instead H is considered as a function in $\boldsymbol{q}$ and $\dot{\boldsymbol{q}}$ we get the following differential

$$\begin{aligned} \mathrm{d}H &= \sum_{j=1}^{n} \left(p_j \mathrm{d}\dot{q}_j + \dot{q}_j \mathrm{d}p_j - \frac{\partial L}{\partial q_i} \mathrm{d}q_i - \frac{\partial L}{\partial \dot{q}_j} \mathrm{d}\dot{q}_j \right) + \frac{\partial H}{\partial t} \mathrm{d}t \\ &= \sum_{j=1}^{n} \left(\dot{q}_j \mathrm{d}p_j - \frac{\partial L}{\partial q_i} \mathrm{d}q_i \right) + \frac{\partial H}{\partial t} \mathrm{d}t \end{aligned} \tag{3.2}$$

The Euler-Lagrange equations is in the coordinates $\boldsymbol{q}$ and $\boldsymbol{p}$ given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{q}}$$

Inserting this in 3.2 gives

$$\mathrm{d}H = \sum_{j=1}^{n} \left(\dot{q}_j \mathrm{d}p_j - \dot{p}_j \mathrm{d}q_j \right) + \frac{\partial H}{\partial t} \mathrm{d}t \tag{3.3}$$

Comparing equation 3.3 and 3.1 gives Hamilton's equations. □

Example 8. Consider the pendulum. We have that the kinetic energy K and the potential energy V are given by

$$K = \frac{1}{2} m \ell^2 \dot{\theta}^2 \qquad\qquad V = -mg\ell \cos(\theta)$$

According to remark 2 we thus have

$$p = m\ell^2 \dot{\theta}$$

and

$$H = \frac{1}{2} \frac{p^2}{m\ell^2} - mg\ell \cos(\theta)$$

Therefore Hamilton's equations for the pendulum are

$$\begin{aligned} \dot{p} &= -mg\ell \sin(\theta) \\ \dot{\theta} &= \frac{p}{m\ell^2} \end{aligned}$$

We define the $2n \times 2n$ matrix J as

$$\mathrm{J} = \begin{bmatrix} 0 & \mathrm{I} \\ -\mathrm{I} & 0 \end{bmatrix} \qquad\qquad \mathrm{J}^{-1} = \mathrm{J}^T = -\mathrm{J}$$

where I is the $n \times n$ identity matrix. This Matrix is called the symplectic identity. Hamiltons equations can then be written as

$$\begin{bmatrix} \dot{\boldsymbol{p}} \\ \dot{\boldsymbol{q}} \end{bmatrix} = -\begin{bmatrix} 0 & \mathrm{I} \\ -\mathrm{I} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial \boldsymbol{p}} \\ \frac{\partial H}{\partial \boldsymbol{q}} \end{bmatrix}$$

We define $\boldsymbol{y} = (\boldsymbol{p}, \boldsymbol{q})$ which gives the equations in the compact form

$$\dot{\boldsymbol{y}} = \mathrm{J}^{-1} \nabla H(\boldsymbol{y}, t)$$

We will denote by φ_t^H the flow of Hamilton's equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H(\boldsymbol{y}, t)$.

Proposition 2. *We have*

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\varphi_t^H(\boldsymbol{y}_0, t_0), t) = \frac{\partial H}{\partial t}(\varphi_t^H(\boldsymbol{y}_0, t_0), t)$$

So if H is independent of t, H is a constant of the motion for Hamilton's equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H(\boldsymbol{y}, t)$.

Proof. This follows from the direct calculation using the chain rule and Hamiltons equations. Take $(\boldsymbol{p}(t), \boldsymbol{q}(t)) = \varphi_t^H(\boldsymbol{y}_0, t_0)$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} H(\boldsymbol{p}(t), \boldsymbol{q}(t), t) &= \frac{\partial H}{\partial \boldsymbol{p}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \cdot \dot{\boldsymbol{p}}(t) + \frac{\partial H}{\partial \boldsymbol{q}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \cdot \dot{\boldsymbol{q}}(t) \\
&\quad + \frac{\partial H}{\partial t}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \\
&= \frac{\partial H}{\partial \boldsymbol{p}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \cdot \left(-\frac{\partial H}{\partial \boldsymbol{q}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \right) \\
&\quad + \frac{\partial H}{\partial \boldsymbol{q}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \cdot \frac{\partial H}{\partial \boldsymbol{p}}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) + \frac{\partial H}{\partial t}(\boldsymbol{p}(t), \boldsymbol{q}(t), t) \\
&= \frac{\partial H}{\partial t}(\boldsymbol{p}(t), \boldsymbol{q}(t), t)
\end{aligned}$$

□

If a system does indeed depend explicitly on t, that is $H = H(\boldsymbol{p}, \boldsymbol{q}, t)$, then we can define a new Hamiltonian $\mathcal{H} : \mathbb{R}^{2n+2} \to \mathbb{R}$ as

$$\mathcal{H}(\boldsymbol{p}, p_0, \boldsymbol{q}, q_0) = p_0 + H(\boldsymbol{p}, \boldsymbol{q}, q_0)$$

Hamilton's equation for this system then are

$$\begin{aligned}
\dot{\boldsymbol{p}} &= -\frac{\partial H}{\partial \boldsymbol{q}} & \dot{p}_0 &= -\frac{\partial H}{\partial q_0} \\
\dot{\boldsymbol{q}} &= \frac{\partial H}{\partial \boldsymbol{p}} & \dot{q}_0 &= 1
\end{aligned}$$

Thus every time dependent system with n degrees of freedom can be written as a system independent of time with $n + 1$ degrees of freedom. For this reason from now on only time independent systems will be considered.

Poincaré Maps for Hamiltonian Systems with 2 Degrees of Freedom

When you have a time independent Hamiltonian system with 2 degrees of freedom with Hamiltonian H, we know from proposition 2 that the Hamiltonian will be a constant of the motion. Thus we know that a trajectory will be confined to move on a set $W_E = \{(p_1, p_2, q_1, q_2) \in \mathbb{R}^4 \mid H(p_1, p_2, q_1, q_2) = E\}$. On W_E we can locally write $p_2 = p_2(p_1, q_1, q_2)$ due to the implicit function theorem, so we can regard W_E as a subset of $\mathbb{R}^3$. Therefore instead of investigating a Poincaré map for a specific surface of section in $\mathbb{R}^4$, that is a volume, we can simplify matters and only consider the trajectory in the p_1, q_1, and q_2 coordinates, i.e. in $W_E \subset \mathbb{R}^3$, and in these coordinates examine the Poincaré map for a surface of section, now indeed a surface, in $\mathbb{R}^3$, e.g. the plane given by $q_2 = 0$.

3.2 Canonical Transformations

The concept of a canonical coordinate transformation is very important since this is the class of coordinate transformations which leave invariant the form of Hamilton's equations.

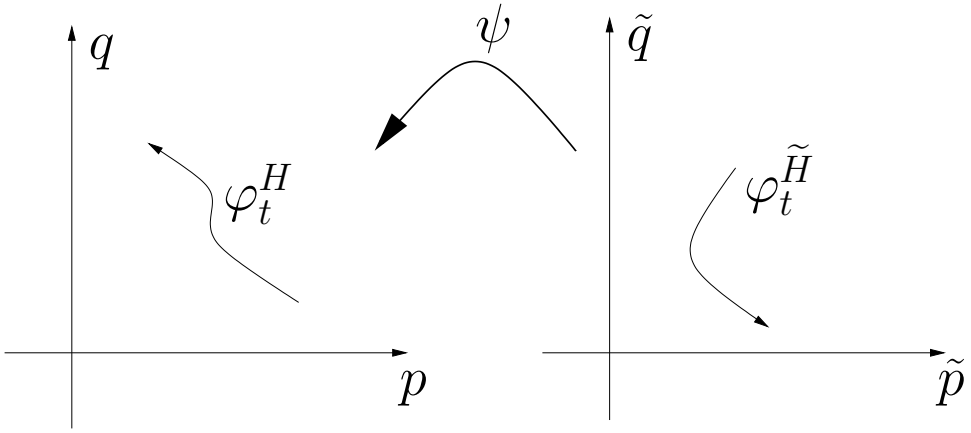


Figure 3.1: Coordinate transformation. It is canonical if $H = \widetilde{H} \circ \psi$.

Definition 1. *The smooth coordinate transformation* $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}) \leftrightarrow (\boldsymbol{p}, \boldsymbol{q})$ *is said to be canonical if for any Hamiltonian* H *Hamiltons's equations are equivalent to*

$$\dot{\tilde{\boldsymbol{p}}} = -\frac{\partial \widetilde{H}}{\partial \tilde{\boldsymbol{q}}} \qquad\qquad \dot{\tilde{\boldsymbol{q}}} = \frac{\partial \widetilde{H}}{\partial \tilde{\boldsymbol{p}}}$$

where $\widetilde{H} = H \circ \psi$.

We then have that ψ is a canonical transformation if and only if

$$\varphi_t^H \circ \psi = \psi \circ \varphi_t^{\widetilde{H}}$$

i.e. solutions are mapped into solutions. Denoting by $\frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}}$ the Jacobian of ψ we have the following result giving an easy characterisation of canonical transformations.

Proposition 3. *The transformation $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $\tilde{\boldsymbol{y}} \leftrightarrow \boldsymbol{y}$, is canonical if and only if its Jacobian $\frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}}$ satisfies the relation*

$$\frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}}^T \mathrm{J} \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} = \mathrm{J}$$

Proof. Since with $\boldsymbol{y} = \boldsymbol{y}(\tilde{\boldsymbol{y}})$ we have

$$\dot{\boldsymbol{y}} = \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \dot{\tilde{\boldsymbol{y}}}$$

we get that two general differential equations

$$\dot{\boldsymbol{y}} = Y(\boldsymbol{y}) \qquad\qquad \dot{\tilde{\boldsymbol{y}}} = \widetilde{Y}(\tilde{\boldsymbol{y}})$$

are equal if and only if

$$Y(\boldsymbol{y}(\tilde{\boldsymbol{y}})) = \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \widetilde{Y}(\tilde{\boldsymbol{y}})$$

using this we get that the transformation is canonical if and only if

$$\begin{aligned} \mathrm{J}^{-1} \nabla H(\boldsymbol{y}(\tilde{\boldsymbol{y}})) &= \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \mathrm{J}^{-1} \nabla \left(\widetilde{H}(\tilde{\boldsymbol{y}}) \right) \\ &= \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \mathrm{J}^{-1} \nabla \left(H(\boldsymbol{y}(\tilde{\boldsymbol{y}}) \right) \\ &= \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \mathrm{J}^{-1} \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}}^T \nabla H(\boldsymbol{y}(\tilde{\boldsymbol{y}})) \end{aligned}$$

giving the relation

$$\frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}} \mathrm{J}^{-1} \frac{\partial \boldsymbol{y}}{\partial \tilde{\boldsymbol{y}}}^T = \mathrm{J}^{-1}$$

Transposing this expression and using $\mathrm{J}^{-1} = \mathrm{J}^T = -\mathrm{J}$ gives the desired relation. □

This immediately gives that a composition of canonical transformations is a canonical transformation.

A $2n \times 2n$ matrix A is said to be symplectic if it satisfies the relation

$$\mathrm{A}^T \mathrm{J} \mathrm{A} = \mathrm{J}$$

Therefore a smooth map, which is a canonical transformation, is also said to be **symplectic**.

Proposition 3 can be used to show the following theorem, which is interesting in it's own right, but which also encourage us to examine symplectic integrators, i.e. integrators that are symplectic mappings.

Proposition 4. *For fixed $t \in \mathbb{R}$ the flow $\varphi_t^H : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ of Hamilton's equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H(\boldsymbol{y})$ is a canonical transformation.*

Proof. Since H is assumed smooth the flow $\varphi_t^H(\boldsymbol{y})$ is a smooth function of t and $\boldsymbol{y}$ and since $\left(\varphi_t^H\right)^{-1} = \varphi_{-t}^H$ it defines a smooth coordinate transformation.

Denoting by $\nabla^2 H$ the Hessian matrix of H we get from

$$\nabla\left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^H(y)\right) = \nabla\left(\mathrm{J}^{-1}\nabla H(\varphi_t^H(\boldsymbol{y}))\right)$$

that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\varphi_t}{\partial\boldsymbol{y}}(\boldsymbol{y}) = \mathrm{J}^{-1}\nabla^2 H(\varphi_t^H(\boldsymbol{y}))\frac{\partial\varphi_t}{\partial\boldsymbol{y}}(\boldsymbol{y})$$

We use this, the fact that $\nabla^2 H$ is symmetric, and the relations $\mathrm{JJ} = -\mathrm{I}$, $\mathrm{J}^{-1} = -\mathrm{J}$, and $\mathrm{J}^T = -\mathrm{J}$, to calculate the following

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\mathrm{J}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\right)^T\mathrm{J}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}} + \frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\mathrm{J}\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\right)\\
&= \frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\left(\nabla^2 H(\varphi_t^H(\boldsymbol{y}))\right)^T\mathrm{J}^{-T}\mathrm{J}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\\
&\quad + \frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\mathrm{J}\mathrm{J}^{-1}\nabla^2 H(\varphi_t^H(\boldsymbol{y}))\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\\
&= -\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\nabla^2 H(\varphi_t^H(\boldsymbol{y}))\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}} + \frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\nabla^2 H(\varphi_t^H(\boldsymbol{y}))\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\\
&= 0
\end{aligned}$$

Since $\varphi_0 = \mathrm{id}$ we have

$$\frac{\partial\varphi_0^H}{\partial\boldsymbol{y}}^T\mathrm{J}\frac{\partial\varphi_0^H}{\partial\boldsymbol{y}} = \mathrm{J}$$

so φ_0^H must be a canonical transformation and because of the above φ_t^H must be a canonical transformation for all t. □

This proposition can also be used to search for a Hamiltonian χ such that the canonical transformation φ_t^χ, t fixed, gives the Hamiltonian system under consideration a Hamiltonian which is simpler in some sense.

From the fact that the flow φ_t^H of Hamilton's equations is a canonical transformation we get for fixed t

$$\begin{aligned}
\det\left(\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}^T\mathrm{J}\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\right) &= \left(\det\left(\frac{\partial\varphi_t^H}{\partial\boldsymbol{y}}\right)\right)^2\det(\mathrm{J})\\
&= \det(\mathrm{J})
\end{aligned}$$

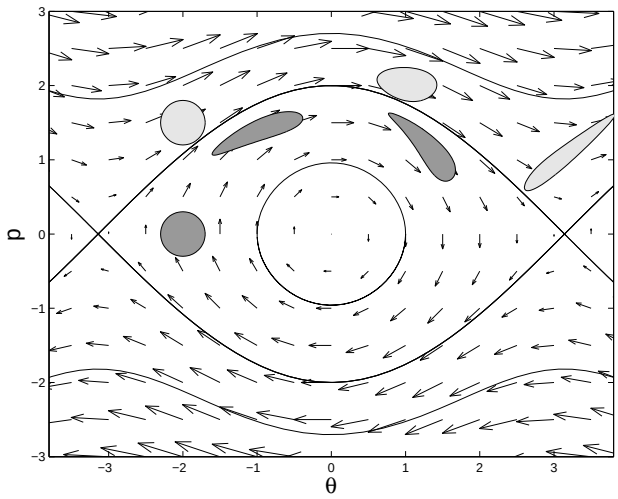


Figure 3.2: The pendulum. The vector field and some solution curves. Two disks of initial conditions and their appearence after time $t = 1.5$ and $t = 3$. The volume of the disks are unchanged.

Since $\det(\mathrm{J}) \neq 0$, φ_t^H is a smooth function of $\boldsymbol{y}$, and $\varphi_0^H = \mathrm{id}$ we thus get

$$\det\left(\frac{\partial \varphi_t^H}{\partial \boldsymbol{y}}\right) = 1$$

The volume of a subset U of $\mathbb{R}^{2n}$, $\mathrm{Vol}(U) = \int_U \mathrm{d}y_1 \dots \mathrm{d}y_{2n}$, when mapped by the flow φ_t^H for fixed t, can be calculated using the above and the change of variables formula for integrals as

$$\begin{aligned}
\mathrm{Vol}(\varphi_t^H(U)) &= \int_{\varphi_t^H(U)} \mathrm{d}y_1 \dots \mathrm{d}y_{2n} \\
&= \int_U \left|\det\left(\frac{\partial \varphi_t^H}{\partial \boldsymbol{y}}\right)\right| \mathrm{d}y_1 \dots \mathrm{d}y_{2n} \\
&= \int_U \mathrm{d}y_1 \dots \mathrm{d}y_{2n} \\
&= \mathrm{Vol}(U)
\end{aligned}$$

Thus we see that the flow of Hamilton's equations preserves the volume. This leads to several important facts about the flow of Hamilton's equations, e.g. it is not possible for the flow to have asymptotically stable equilibrium positions or asymptotically stable limit cycles.

A canonical transformation $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $\tilde{\boldsymbol{y}} \mapsto \boldsymbol{y}$, can very well mix the p and q coordinates. This is very different from the Lagrangian setting where only point transformations $\tilde{\boldsymbol{q}} \mapsto \boldsymbol{q}$ are considered.

The following theorem gives a very usefull way of constructing canonical transformations.

Proposition 5. *Given a smooth function* $S_1 : \mathbb{R}^{2n} \to \mathbb{R}$, $S_1 = S_1(\tilde{\boldsymbol{q}}, \boldsymbol{q})$, $\tilde{\boldsymbol{q}}, \boldsymbol{q} \in \mathbb{R}^n$, *such that*

$$\det\left(\frac{\partial^2 S_1}{\partial \tilde{\boldsymbol{q}} \partial \boldsymbol{q}}\right) \neq 0$$

the equations

$$\tilde{\boldsymbol{p}} = -\frac{\partial S_1}{\partial \tilde{\boldsymbol{q}}} \qquad\qquad \boldsymbol{p} = \frac{\partial S_1}{\partial \boldsymbol{q}}$$

implicitly define a local canonical coordinate transformation $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$.

Proof. If we define $\tilde{\boldsymbol{p}} = \tilde{\boldsymbol{p}}(\tilde{\boldsymbol{q}}, \boldsymbol{q})$ and $\boldsymbol{p} = \boldsymbol{p}(\tilde{\boldsymbol{q}}, \boldsymbol{q})$ by

$$\tilde{\boldsymbol{p}} = -\frac{\partial S_1}{\partial \tilde{\boldsymbol{q}}} \qquad\qquad \boldsymbol{p} = \frac{\partial S_1}{\partial \boldsymbol{q}} \tag{3.4}$$

then the condition

$$\det\left(\frac{\partial^2 S_1}{\partial \tilde{\boldsymbol{q}} \partial \boldsymbol{q}}\right) \neq 0$$

gives according to the implicit function theorem that $\boldsymbol{q} = \boldsymbol{q}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$ and thus $\boldsymbol{p} = \boldsymbol{p}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$. We can therefore calculate

$$\begin{aligned}\frac{\partial(\boldsymbol{p}, \boldsymbol{q})}{\partial(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})} &= \begin{bmatrix} \frac{\partial \boldsymbol{p}}{\partial \tilde{\boldsymbol{p}}} & \frac{\partial \boldsymbol{p}}{\partial \tilde{\boldsymbol{q}}} \\ \frac{\partial \boldsymbol{q}}{\partial \tilde{\boldsymbol{p}}} & \frac{\partial \boldsymbol{q}}{\partial \tilde{\boldsymbol{q}}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \boldsymbol{p}}{\partial \tilde{\boldsymbol{p}}} & \frac{\partial^2 S_1}{\partial \tilde{\boldsymbol{q}} \partial \boldsymbol{q}} \\ \left(\frac{\partial^2 S_1}{\partial \boldsymbol{q} \partial \tilde{\boldsymbol{q}}}\right)^{-1} & \frac{\partial \boldsymbol{q}}{\partial \tilde{\boldsymbol{q}}} \end{bmatrix}\end{aligned}$$

which is nonsingular so the inverse function theorem insures that 3.4 implicitly defines a coordinate transformation $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$.

Hamiltons equations with Hamiltonian $H : \mathbb{R}^{2n} \to \mathbb{R}$, $H = H(\boldsymbol{p}, \boldsymbol{q})$, are the Euler-Lagrange equations with Lagrangian $L : \mathbb{R}^{4n} \to \mathbb{R}$ given by

$$L(\boldsymbol{q}, \boldsymbol{p}, \dot{\boldsymbol{q}}, \dot{\boldsymbol{p}}) = \sum_{i=1}^{n} p_i \dot{q}_i - H(\boldsymbol{p}, \boldsymbol{q})$$

For $L : \mathbb{R}^{2m} \to \mathbb{R}$ and $S : \mathbb{R}^m \to \mathbb{R}$ smooth, the functionals

$$\begin{aligned} I_1(\gamma) &= \int_{t_0}^{t_1} L(\gamma(t), \dot{\gamma}(t)) \mathrm{d}t \\ I_2(\gamma) &= \int_{t_0}^{t_1} \left(L(\gamma(t), \dot{\gamma}(t)) + \frac{\mathrm{d}}{\mathrm{d}t} S(\gamma(t)) \right) \mathrm{d}t \end{aligned}$$

where γ is a smooth curve $\gamma : [t_0, t_1] \to \mathbb{R}^m$, leads to the same Euler-Lagrange equations. Combining these two facts and proposition 1 we see that if the canonical coordinates $(\boldsymbol{p}, \boldsymbol{q})$ and the general coordinates $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$ with $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$ satisfies

$$\sum_{i=1}^{n} p_i \dot{q}_i - H(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} \tilde{p}_i \dot{\tilde{q}}_i - \widetilde{H}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}) + \frac{\mathrm{d}}{\mathrm{d}t} S(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$$

where $\widetilde{H} = H \circ \psi$, then ψ must be a canonical transformation. This sufficient condition is equivalent to

$$\begin{aligned} \mathrm{d}S &= -\sum_{i=1}^{n} \tilde{p}_i \mathrm{d}\tilde{q}_i + \sum_{i=1}^{n} p_i \mathrm{d}q_i + (\widetilde{H} - H)\mathrm{d}t \\ &= -\sum_{i=1}^{n} \tilde{p}_i \mathrm{d}\tilde{q}_i + \sum_{i=1}^{n} p_i \mathrm{d}q_i \end{aligned} \tag{3.5}$$

Thus when comparing

$$\mathrm{d}S_1 = \sum_{i=1}^{n} \frac{\partial S_1}{\partial \tilde{q}_i} \mathrm{d}\tilde{q}_i + \sum_{i=1}^{n} \frac{\partial S_1}{\partial q_i} \mathrm{d}q_i$$

with the sufficient condition on canonicity, that is equation 3.5, we see that the coordinate transformation $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$ given by 3.4 is canonical. □

A similar constructive proposition for finding canonical transformations is the following regarding a function of $\tilde{\boldsymbol{p}}$ and $\boldsymbol{q}$

Proposition 6. *Given a smooth function $S_2 : \mathbb{R}^{2n} \to \mathbb{R}$, $S_2 = S_2(\tilde{\boldsymbol{p}}, \boldsymbol{q})$, $\tilde{\boldsymbol{p}}, \boldsymbol{q} \in \mathbb{R}^n$, such that*

$$\det\left(\frac{\partial^2 S_2}{\partial \tilde{\boldsymbol{p}} \partial \boldsymbol{q}}\right) \neq 0$$

the equations

$$\boldsymbol{p} = \frac{\partial S_2}{\partial \boldsymbol{q}} \qquad\qquad \tilde{\boldsymbol{q}} = \frac{\partial S_2}{\partial \tilde{\boldsymbol{p}}}$$

implicitly define a local canonical coordinate transformation $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})$.

Proof. The argumentation that this indeed does implicitly define a coordinate transformation is completely similar to that in proposition 5.

If we in proposition 5 use

$$S(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}) = \bar{S}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}) - \sum_{i=1}^{n} \tilde{p}_i \tilde{q}_i$$

The sufficient condition on canonicity, that is equation 3.5, is seen to be equivalent to

$$\mathrm{d}\bar{S} = \sum_{i=1}^{n} p_i \mathrm{d}q_i + \sum_{i=1}^{n} \tilde{q}_i \mathrm{d}\tilde{p}_i$$

Since

$$\begin{aligned}\mathrm{d}S_2 &= \sum_{i=1}^{n} \frac{\partial S_2}{\partial \tilde{p}_i} \mathrm{d}\tilde{p}_i + \sum_{i=1}^{n} \frac{\partial S_2}{\partial q_i} \mathrm{d}q_i \\ &= \sum_{i=1}^{n} \tilde{q}_i \mathrm{d}\tilde{p}_i + \sum_{i=1}^{n} p_i \mathrm{d}q_i\end{aligned}$$

we therefore see that this transformation is canonical. □

Similar theorems can be made for functions $S_3(\boldsymbol{p}, \tilde{\boldsymbol{q}})$ and $S_4(\boldsymbol{p}, \tilde{\boldsymbol{p}})$. The functions S_1, S_2, S_3, and S_4 are called **generating functions**. The sign when differentiating on of the generating functions can be read from the mnemonic figure 3.3.

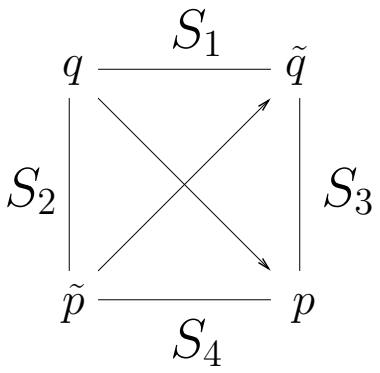


Figure 3.3: Canonical transformations given by a generating function. Going along an arrow, in any direction, gives the new variable as the end point. Going in the arrows direction corresponds to differentiating, w.r.t. the variable at the starting point, the generating function at the starting point and nothing else whereas going opposite the arrows direction corresponds to differentiating, and changing the sign.

Example 9. Consider the generating function S_2 given by

$$S_2(\tilde{\boldsymbol{p}}, \boldsymbol{q}) = \tilde{\boldsymbol{p}} \cdot \boldsymbol{g}(\boldsymbol{q})$$

which gives the canonical transformation

$$\boldsymbol{p} = \left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{q}}(\boldsymbol{q}) \right)^T \tilde{\boldsymbol{p}} \qquad\qquad \tilde{\boldsymbol{q}} = \boldsymbol{g}(\boldsymbol{q})$$

This is seen to give the way in which a transformation of the q's extends canonically to the p's. Such a transformation is called a contact transformation. For $\boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{q}$ this is the identity transformation.

These propositions give a constructive way of searching for canonical transformations which makes the Hamiltonian in the new coordinates much simpler, e.g. $\widetilde{H}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}) = \widetilde{H}(\tilde{\boldsymbol{q}})$. This is accomplished if for example

$$H\left(\left.\frac{\partial S_1}{\partial \boldsymbol{q}}\right|_{\boldsymbol{q}=\boldsymbol{q}(\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}})}, \boldsymbol{q}(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}})\right) = \widetilde{H}(\tilde{\boldsymbol{q}})$$

so we get a partial differential equation which, if it has a solution, can be used to find such a generating function. The above partial differential equation, along with similar equations for the other generating functions, is called the **Hamilton-Jacobi equation**.

3.3 Integrable Systems

For two smooth functions $f, g : \mathbb{R}^{2n} \to \mathbb{R}$ their **Poisson bracket** $\{f, g\}$ is the smooth function given by

$$\begin{aligned}\{f, g\} &= \big(\nabla f(\boldsymbol{y})\big)^T \mathrm{J}^{-1} \nabla g(\boldsymbol{y}) \\ &= \sum_{j=1}^{n} \left(\frac{\partial f}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial p_j}\frac{\partial g}{\partial q_j}\right)\end{aligned}$$

The following proposition is an immediate consequense of the definition.

Proposition 7. *The Poisson bracket* $\{,\} : C^\infty(\mathbb{R}^{2n}, \mathbb{R}) \times C^\infty(\mathbb{R}^{2n}, \mathbb{R}) \to C^\infty(\mathbb{R}^{2n}, \mathbb{R})$*, given by* $\{f, g\} = \big(\nabla f(\boldsymbol{y})\big)^T \mathrm{J}^{-1} \nabla g(\boldsymbol{y})$*,* $f, g, h \in C^\infty(\mathbb{R}^{2n}, \mathbb{R})$ *, satisfies:*

1. $\{f, g\}$ *is bilinear in* f *and* g*.*
2. $\{f, g\} = -\{g, f\}$*, skew-symmetry.*
3. $\{fg, h\} = f\{g, h\} + g\{f, h\}$*, derivation in each argument.*
4. $\{\{f, g\}, h\} + \{\{h, f\}, g\} + \{\{g, h\}, f\} = 0$*, Jacobi's identity.*

We then have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} F(\varphi_t^H(\boldsymbol{y})) &= \sum_{i=1}^{n} \frac{\partial F}{\partial y_i}(\varphi_t^H(\boldsymbol{y})) \left(\dot{\varphi}_t^H(\boldsymbol{y})\right)_i \\ &= \big(\nabla F(\varphi_t^H(\boldsymbol{y})\big)^T \mathrm{J}^{-1} \nabla H(\varphi_t^H(\boldsymbol{y})) \\ &= \{F, H\}(\varphi_t^H(\boldsymbol{y}))\end{aligned}$$

Thus we see that F is a constant of the motion of Hamilton's equations with Hamiltonian H if and only if $\{F, H\} = 0$, and therefore we also immediately get that H is a constant of the motion since $\{H, H\} = 0$ due to the skew-symmetry of the bracket. Since we get from Jacobi's identity

$$\{F_1, H\} = 0 \ , \ \{F_2, H\} = 0 \qquad \Rightarrow \qquad \{\{F_1, F_2\}, H\} = 0$$

we have that if F_1 and F_2 are constants of the motion of Hamiltons equations then so is $\{F_1, F_2\}$.

If $F_1, \ldots, F_n$ are smooth and linearly independent, i.e.

$$\text{Rank}\left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{y}}\right) = n$$

where $\boldsymbol{F} : \mathbb{R}^{2n} \to \mathbb{R}^n$, $\boldsymbol{y} \mapsto [F_1(\boldsymbol{y}) \ \ldots \ F_n(\boldsymbol{y})]^T$, then we know from the implicit function theorem that $\boldsymbol{F} = c = constant$ defines an n dimensional subset of $\mathbb{R}^{2n}$ for which it is possible to assign coordinates. If furthermore $\{F_i, F_j\} = 0$, for all i, j, we have since

$$\{F_i, F_j\} = 0 \qquad \Rightarrow \qquad \varphi_t^{F_i} \circ \varphi_s^{F_j} = \varphi_s^{F_j} \circ \varphi_t^{F_i}$$

that we can take $s_1, \ldots, s_n$ as local coordinates on $\Sigma_c = \{\boldsymbol{y} \in \mathbb{R}^{2n} | \boldsymbol{F}(\boldsymbol{y}) = c\}$ defined by

$$\varphi_{s_1}^{F_1} \circ \ldots \circ \varphi_{s_n}^{F_n}(\boldsymbol{y}_0) : \mathbb{R}^n \to \Sigma_c$$

where $\boldsymbol{y}_0 \in \Sigma_c$. These considerations are of importance in the proof of the following theorem

Theorem 7 (Liouville-Arnol'd theorem)**.** *Assume there exist n smooth functions $F_1, \ldots, F_n : \mathbb{R}^{2n} \to \mathbb{R}$, such that $\{F_i, F_j\} = 0$ for all i, j. Assume $F_1, \ldots, F_n$ are linearly independent on a level set Σ_c. Furthermore assume that Σ_c is compact, then:*

1. *Σ_c is diffeomorphic to the n-dimensional torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$*
2. *There exists a neighborhood $U \subset \mathbb{R}^n$ of c such that the set*

$$D_U = \bigcup_{c' \in U} \Sigma_{c'}$$

is diffeomorphic to $U \times \Sigma_c$ and in D_U there exists a canonical transformation $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\boldsymbol{I}, \boldsymbol{\varphi})$, defined for $\boldsymbol{I} \in B \subset \mathbb{R}^n$ and $\boldsymbol{\varphi} \in \mathbb{T}^n$, $D_U = \psi(B \times \mathbb{T}^n)$, such that $\boldsymbol{I}$ is constant on tori and, conversely, $\widetilde{\boldsymbol{F}} = \boldsymbol{F} \circ \psi$ depend only on $\boldsymbol{I}$, i.e. $\widetilde{\boldsymbol{F}}(\boldsymbol{I}, \boldsymbol{\varphi}) = \widetilde{\boldsymbol{F}}(\boldsymbol{I})$.

Proof. See [1] □

From this we immediately get

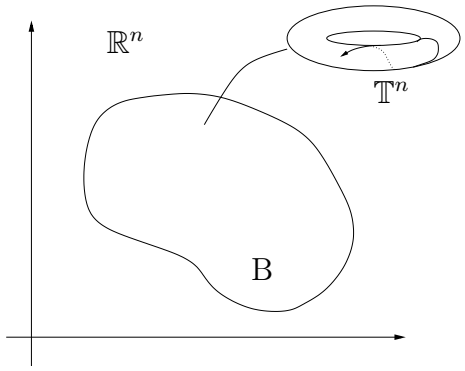


Figure 3.4: Integrable system. The foliation of $B \subset \mathbb{R}^n$ with tori with quasi periodic motion.

Corollary 1. *Consider Hamilton's equations* $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H(\boldsymbol{y})$. *Let* $F_1 = H$ *and assume that there exists* $n-1$ *smooth functions* $F_j : \mathbb{R}^{2n} \to \mathbb{R}$, $j = 2, \ldots, n$, *such that* $\{F_i, F_j\} = 0$ *for all* $i, j = 1, \ldots, n$, *i.e.* F_j, $j = 1, ..., n$, *are constants of the motion of the flow of* $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla F_1(\boldsymbol{y})$. *If* $F_1, \ldots, F_n$ *are linearly independent on the level sets* Σ_c *of the* n *constants of the motion and if these level sets are compact, then each of them is diffeomorphic to the* n*-dimensional torus* $\mathbb{T}^n$. *In a neighborhood of* Σ_c *there exist a canonical coordinate transformation* $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $(\boldsymbol{p}, \boldsymbol{q}) = \psi(\boldsymbol{I}, \boldsymbol{\varphi})$, *such that the new hamiltonian* $\widetilde{H} = H \circ \psi$ *is given by*

$$\widetilde{H}(\boldsymbol{I}, \boldsymbol{\varphi}) = h(\boldsymbol{I})$$

The I-coordinates are called action variables and the φ coordinates are called the angle variables, and together $(\boldsymbol{I}, \boldsymbol{\varphi})$ are called **action-angle variables**. Hamiltonian systems which satisfy the assumptions of the Liouville-Arnol'd theorem are said to be **integrable**. Thus for integrable systems the dynamics is given by

$$\begin{aligned} \dot{\boldsymbol{I}} &= 0 \\ \dot{\boldsymbol{\varphi}} &= \boldsymbol{\omega}(\boldsymbol{I}) \end{aligned}$$

where $\boldsymbol{\omega}(\boldsymbol{I}) = \frac{\partial h}{\partial \boldsymbol{I}}(\boldsymbol{I})$. So the motion takes place on a torus $\mathbb{T}^n$, one for each value of $\boldsymbol{I}$, with constant angular velocity. Such a motion is said to be quasi periodic with frequency $\boldsymbol{\omega}(\boldsymbol{I})$.

Poincaré Maps for integrable Systems with 2 Degrees of Freedom

If we follow the procedure explained earlier about Poincaré maps for Hamiltonian systems with 2 degrees of freedom, we know according to the Liouville-Arnol'd theorem that for every initial condition the trajectory will lie on a torus $\mathbb{T}^2$ imbedded in $W_E \subset \mathbb{R}^3$. Thus when making a Poincaré map for

such a system each point will lie on on a closed curve, and if we change the definition of direction in the map the points of this map will lie on another closed curve. Therefore a Poincaré map can immediately show if a system with 2 degrees of freedom is not integrable.

3.4 Nearly Integrable Systems

Two cornerstones in pertubation theory for integrable Hamiltonian systems are the KAM theorem and the Nekhoroshev theorem. The KAM theorem gives sufficent conditions for a perturbed integrable system to behave qualitatively as the integrable system. The Nekhoroshev theorem instead gives conditions on the system for the action variables to be almost conserved quantities of the motion. Here we will only state the theorems without proofs since these are very long and technical.

A **nearly integrable Hamiltonian system** is a system with Hamiltonian $H : B \times \mathbb{T}^n \to \mathbb{R}$, $B \subset \mathbb{R}^n$, of the form

$$H(\boldsymbol{I}, \boldsymbol{\varphi}) = h(\boldsymbol{I}) + \epsilon f(\boldsymbol{I}, \boldsymbol{\varphi}) \qquad \epsilon > 0 \tag{3.6}$$

Suppose we can find a canonical transformation $(\boldsymbol{I}, \boldsymbol{\varphi}) = \psi(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}})$ which simplifies the Hamiltonian as

$$\widetilde{H} = \tilde{h}(\tilde{\boldsymbol{I}}) + \epsilon g(\tilde{\boldsymbol{I}}) + \epsilon^2 \tilde{f}(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}})$$

Such a generation function must depend on ϵ and for $\epsilon = 0$ it should be the identity. A near the identity canonical transformation is a canonical transformation $\psi = \psi_\epsilon$, ϵ small, with

$$\lim_{\epsilon \to 0} \psi_\epsilon = \mathrm{id}$$

Therefore we need a near the identity canonical transformation to simplify the Hamiltonian. It can be proven that any near the identity canonical transformation can be constructed using a generating function of the form

$$S_2^\epsilon(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi}) = \tilde{\boldsymbol{I}} \cdot \boldsymbol{\varphi} - \epsilon \chi^\epsilon(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi}, \epsilon)$$

Since we will need only the first order part in ϵ we will use

$$S_2(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi}) = \tilde{\boldsymbol{I}} \cdot \boldsymbol{\varphi} - \epsilon \chi(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi})$$

According to proposition 6 we have

$$\boldsymbol{I} = \tilde{\boldsymbol{I}} - \epsilon \frac{\partial \chi}{\partial \boldsymbol{\varphi}}(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi})$$

$$\tilde{\boldsymbol{\varphi}} = \boldsymbol{\varphi} - \epsilon \frac{\partial \chi}{\partial \tilde{\boldsymbol{I}}}(\tilde{\boldsymbol{I}}, \boldsymbol{\varphi})$$

Inverting these we get

$$\begin{aligned}\boldsymbol{I} &= \tilde{\boldsymbol{I}} - \epsilon \frac{\partial \chi}{\partial \boldsymbol{\varphi}}(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}}) + \mathcal{O}(\epsilon^2) \\ \boldsymbol{\varphi} &= \tilde{\boldsymbol{\varphi}} + \mathcal{O}(\epsilon)\end{aligned}$$

and the hamiltonian transforms to

$$\widetilde{H}(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}}) = h(\tilde{\boldsymbol{I}}) + \epsilon \left(-\boldsymbol{\omega}(\tilde{\boldsymbol{I}}) \cdot \frac{\partial \chi}{\partial \tilde{\boldsymbol{\varphi}}}(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}}) + f(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{\varphi}}) \right) + \mathcal{O}(\epsilon^2)$$

where $\boldsymbol{\omega} = \frac{\partial h}{\partial \boldsymbol{I}}$. Thus, when dropping the tilde, we wish to find a function χ such that

$$g(\boldsymbol{I}) = -\boldsymbol{\omega}(\boldsymbol{I}) \cdot \frac{\partial \chi}{\partial \boldsymbol{\varphi}}(\boldsymbol{I}, \boldsymbol{\varphi}) + f(\boldsymbol{I}, \boldsymbol{\varphi}) \tag{3.7}$$

for some function g. Since χ is a function on $B \times \mathbb{T}^n$ it can be written

$$\chi(\boldsymbol{I}, \boldsymbol{\varphi}) = \sum_{\nu \in \mathbb{Z}^n} \chi_\nu(\boldsymbol{I}) e^{i\nu \cdot \boldsymbol{\varphi}}$$

and therefore

$$\left\langle \frac{\partial \chi}{\partial \boldsymbol{\varphi}} \right\rangle = 0$$

where $\langle \rho \rangle = \frac{1}{(2\pi)^n} \int_0^{2\pi} \rho(\boldsymbol{I}, \boldsymbol{\varphi}) \mathrm{d}\varphi_1 \dots \mathrm{d}\varphi_n$ denotes the average in the angle variables. Equation 3.7 therefore gives $g(\boldsymbol{I}) = \langle f \rangle$ and simplifies to

$$\boldsymbol{\omega}(\boldsymbol{I}) \cdot \frac{\partial \chi}{\partial \boldsymbol{\varphi}}(\boldsymbol{I}, \boldsymbol{\varphi}) = f(\boldsymbol{I}, \boldsymbol{\varphi}) - \langle f \rangle \tag{3.8}$$

This equation however can generally not be solved, a fact which is known as the Poincaré difficulty, and can be stated precisely as follows

Proposition 8. *Let the Hamiltonian be given by*

$$H(\boldsymbol{I}, \boldsymbol{\varphi}) = h(\boldsymbol{I}) + \epsilon f(\boldsymbol{I}, \boldsymbol{\varphi}) \qquad (\boldsymbol{I}, \boldsymbol{\varphi}) \in B \times \mathbb{T}^n$$

and assume:

1. *h is non degenerate, i.e.*

$$\det \left(\frac{\partial^2 h}{\partial \boldsymbol{I} \partial \boldsymbol{I}} \right) \neq 0$$

in an open subset $B_0 \subset B$.

2. f has essentially full Fourier series, more precisely, denoting

$$f(\boldsymbol{I},\boldsymbol{\varphi}) = \sum_{\nu\in\mathbb{Z}^n} f_\nu(\boldsymbol{I})e^{i\nu\cdot\boldsymbol{\varphi}}$$

for any $\nu \in \mathbb{Z}^n$ there exits ν' parallel to ν such that $f_{\nu'}(\boldsymbol{I}) \neq 0$ in B_0. Then in B_0 there doesn't exist a function χ solving equation 3.8

Proof. Since

$$\boldsymbol{\omega}(\boldsymbol{I})\cdot\frac{\partial\chi}{\partial\boldsymbol{\varphi}} = \sum_{\nu\in\mathbb{Z}^n} i\nu\cdot\boldsymbol{\omega}(\boldsymbol{I})\chi_\nu(\boldsymbol{I})e^{i\nu\cdot\boldsymbol{\varphi}}$$

we see that in order for 3.8 to be satisfied in B_0 it would be necessary that

$$i\nu\cdot\boldsymbol{\omega}(\boldsymbol{I})\chi_\nu(\boldsymbol{I}) = f_\nu(\boldsymbol{I}) \qquad \forall\nu\in\mathbb{Z}^n\setminus\{0\}$$

in B_0. This can formally be solved as

$$\chi_\nu(\boldsymbol{I}) = \frac{f_\nu(\boldsymbol{I})}{i\nu\cdot\boldsymbol{\omega}(\boldsymbol{I})}$$

According to assumption 1 $\boldsymbol{\omega}$ is a diffeomorphism and since $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$ we will have $\nu\cdot\boldsymbol{\omega}(\boldsymbol{I}) = 0$ in a dense subset $B_0' = \boldsymbol{\omega}^{-1}(\mathbb{Q}^n\cap\boldsymbol{\omega}(B_0)) \subset B_0$, and therefore for all ν' parallel to ν, but $f_{\nu'} \neq 0$ for at least one ν' due to assumption 2, which proves the proposition. □

From the proof of this we see that in order to proceed we need to define sufficient conditions for

$$\chi_\nu(\boldsymbol{I},\boldsymbol{\varphi}) = \sum_{\nu\in\mathbb{Z}^n}\chi_\nu(\boldsymbol{I})e^{i\nu\cdot\boldsymbol{\varphi}} \qquad \chi_\nu(\boldsymbol{I}) = \frac{f_\nu(\boldsymbol{I})}{i\nu\cdot\boldsymbol{\omega}(\boldsymbol{I})} \quad \nu\in\mathbb{Z}^n\setminus\{0\}$$

to be solvable. We see that it is necesarry that $\boldsymbol{\omega}$ is nonresonant, i.e.

$$\nu\cdot\boldsymbol{\omega}(\boldsymbol{I}) \neq 0 \qquad \forall\nu\in\mathbb{Z}^n$$

It turns out however that it is not sufficient that $\boldsymbol{\omega}$ is nonresonant.

The proper sufficiency condition on a frequency $\boldsymbol{\omega}\in\mathbb{R}^n$ is that there must exist real constants $\gamma > 0$ and $\tau > n-1$ such that

$$|\nu\cdot\boldsymbol{\omega}| \geq \frac{\gamma}{|\nu|^\tau} \qquad \forall\nu\in\mathbb{Z}^n\setminus\{0\}$$

where $|\nu| = |\nu_1| + \ldots|\nu_n|$. Then $\boldsymbol{\omega}$ is said to be **Diophantine**, or more precisely (γ,τ)-Diophantine. The set $\Omega_{\gamma,\tau}$ consisting of all (γ,τ)-Diophantine frequencies in a ball K in $\mathbb{R}^n$, has the counter intuitive property that it has large measure, but its complement $K\setminus\Omega_{\gamma,\tau}$ is open and dense in K.

Example 10. Consider the interval $[0,1]$. Then the set $[0,1] \cap \mathbb{Q}$ of rational numbers in $[0,1]$ is countable and denote by a_i the i-th element. Denote by W and V the sets

$$W = \bigcup_{i \in \mathbb{N}} \left(a_i - \frac{1}{2}\left(\frac{1}{4}\right)^i, a_i + \frac{1}{2}\left(\frac{1}{4}\right)^i \right) \cap [0,1] \qquad V = [0,1] \setminus W$$

Then we have that W is open and dense in $[0,1]$, since the rational numbers are dense in the real numbers. Furthermore we have

$$1 \geq \text{Measure}(V) = 1 - \text{Measure}(W) \geq 1 - \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^i = \frac{2}{3}$$

So the set V has the same peculiar proporties as $\Omega_{\gamma,\tau}$.

In the different theorems on nearly integrable systems some definitions will be needed:

1. $\rho = (\rho_I, \rho_\varphi) \in \mathbb{R}^+ \times \mathbb{R}^+$.
2. $\| \cdot \|$ denotes the Euclidian norm in $\mathbb{R}^n$.
3. The strip $\mathcal{S}_\rho = \{\boldsymbol{\varphi} \in \mathbb{C}^n \ : \ |\text{Im}\varphi_j| < \rho_\varphi, \ j = 1, \ldots, n\}$.
4. The supremum norm $\| \cdot \|_\rho^\infty$ of a function $F : \mathcal{S}_\rho \to \mathbb{C}$
$$\|F\|_\rho^\infty = \sup_{\boldsymbol{\varphi} \in \mathcal{S}_\rho} \|F(\boldsymbol{\varphi})\|$$
5. The Fourier norm $\| \cdot \|_\rho$ of a function $F : \mathcal{S}_\rho \to \mathbb{C}$, periodic of real period 2π in each argument
$$\|F\|_\rho = \sum_{\nu \in \mathbb{Z}^n} |F_\nu| e^{\rho_\varphi |\nu|}$$
6. For a ball $B \subset \mathbb{R}^n$
$$\mathcal{B}_\rho = \bigcup_{\boldsymbol{I} \in B} \Delta_\rho(\boldsymbol{I}) \ , \quad \Delta_\rho(\boldsymbol{I}) = \{\tilde{\boldsymbol{I}} \in \mathbb{C}^n : |\tilde{I}_j - I_j| < \rho_I, \ j = 1, \ldots, n\}$$
7. $\mathcal{D}_\rho = \mathcal{B}_\rho \times \mathcal{S}_\rho$
8. The Fourier norm $\| \cdot \|_\rho$ of a function $F : \mathcal{D}_\rho \to \mathbb{C}$, $f = F(\boldsymbol{I}, \boldsymbol{\varphi})$, periodic of real period 2π in φ_i
$$\|F\|_\rho = \sup_{\boldsymbol{I} \in \mathcal{B}_\rho} \|F(\boldsymbol{I}, \cdot)\|_\rho$$

A complex function $F : U \to \mathbb{C}$, $U \subset \mathbb{C}^n$ open, is said to be analytic if it is differentiable in every point in U or equivalently if it's Taylor series in around every point in U converges in a neighbourhood of that point (see e.g. [14]).

3.4.1 The KAM Theorem

This theorem is named after A. N. Kolmogorov who first proved it (in 1954) and V. I. Arnold, and J. K. Moser who shortly after proved some slightly different theorems about esentially the same subject. Here we give the theorem in the form originally given by Kolmogorov.

Consider the Nearly integrable system with Hamiltonian

$$H(\boldsymbol{I}, \boldsymbol{\varphi}) = h(\boldsymbol{I}) + \epsilon f(\boldsymbol{I}, \boldsymbol{\varphi}) \qquad (\boldsymbol{I}, \boldsymbol{\varphi}) \in B \times \mathbb{T}^n$$

Consider a fixed $\boldsymbol{I}^* \in B$ and define

$$H^*(\boldsymbol{J}, \boldsymbol{\varphi}) = H(\boldsymbol{I}^* + \boldsymbol{J}, \boldsymbol{\varphi})$$

Then a Taylor expansion gives

$$\begin{aligned} H^*(\boldsymbol{J}, \boldsymbol{\varphi}) = h(\boldsymbol{I}^*) + \boldsymbol{\omega}^* \cdot \boldsymbol{J} + \frac{1}{2}\Gamma \boldsymbol{J} \cdot \boldsymbol{J} \\ + \epsilon \left(A(\boldsymbol{\varphi}) + B(\boldsymbol{\varphi}) \cdot \boldsymbol{J} + \frac{1}{2} C(\boldsymbol{\varphi}) \boldsymbol{J} \cdot \boldsymbol{J} \right) + \mathcal{O}(\|\boldsymbol{J}\|^3) \end{aligned}$$

where

$$\boldsymbol{\omega}^* = \boldsymbol{\omega}(\boldsymbol{I}^*) \qquad \Gamma = \frac{\partial^2 h}{\partial \boldsymbol{I} \partial \boldsymbol{I}}(\boldsymbol{I}^*)$$

$$A(\boldsymbol{\varphi}) = f(\boldsymbol{I}^*, \boldsymbol{\varphi}) \qquad B(\boldsymbol{\varphi}) = \frac{\partial f}{\partial \boldsymbol{I}}(\boldsymbol{I}^*, \boldsymbol{\varphi}) \qquad C(\boldsymbol{\varphi}) = \frac{\partial^2 f}{\partial \boldsymbol{I} \partial \boldsymbol{I}}(\boldsymbol{I}^*, \boldsymbol{\varphi})$$

Kolmogorov's theorem then asserts that under certain conditions it is possible to construct a canonical transformation such that A and B vanish — the resulting Hamiltonian is then said to be in Kolmogorov weak normal form.

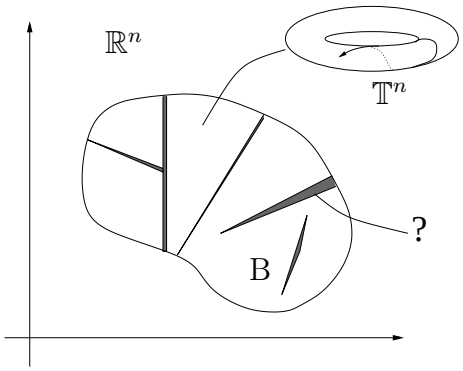


Figure 3.5: Nearly integrable system. The foliation of a large subset, with respect to measure, of $B \subset \mathbb{R}^n$ with tori with quasi periodic motion.

Theorem 8 (Kolmogorov's theorem)**.** *Consider the Hamiltonian system with Hamiltonian*

$$H(\boldsymbol{I}, \boldsymbol{\varphi}) = h(\boldsymbol{I}) + \epsilon f(\boldsymbol{I}, \boldsymbol{\varphi}) \qquad (\boldsymbol{I}, \boldsymbol{\varphi}) \in B \times \mathbb{T}^n$$

and suppose it is analytic in a complex neighborhood $\mathcal{D}_\rho$ of $B \times \mathbb{T}^n$ with some $\rho = (\rho_I, \rho_\varphi)$. Let $\boldsymbol{I}^* \in B$ be such that $\boldsymbol{\omega}^* = \boldsymbol{\omega}(\boldsymbol{I}^*)$ is (γ, τ)-Diophantine and $\Gamma = \frac{\partial^2 h}{\partial \boldsymbol{I} \partial \boldsymbol{I}}(\boldsymbol{I}^*)$ is invertible, and define $H^*(\boldsymbol{J}, \boldsymbol{\varphi}) = H(\boldsymbol{I}^* + \boldsymbol{J}, \boldsymbol{\varphi})$.

Then there exist a positive constant $\mathcal{E}$, depending on n, ρ, γ, $\|\boldsymbol{\omega}^*\|$, and $\|\Gamma^{-1}\|$, such that if $\epsilon \|f\|_\rho^\infty < \mathcal{E}$, then there exists a neighborhood B_0 of $\boldsymbol{J} = 0$, and a canonical transformation $(\boldsymbol{J}, \boldsymbol{\varphi}) = \psi(\boldsymbol{J}', \boldsymbol{\varphi}')$, defined in $B_0 \times \mathbb{T}^n$, such that the new Hamiltonian $H' = H^* \circ \psi$ is in Kolmogorov weak normal form

$$H'(\boldsymbol{J}', \boldsymbol{\varphi}') = \boldsymbol{\omega}^* \cdot \boldsymbol{J}' + \frac{1}{2} \Gamma' \boldsymbol{J}' \cdot \boldsymbol{J}' + \mathcal{O}(\|\boldsymbol{J}'\|^2)$$

Γ' being a constant matrix close to Γ.

Proof. See [4] □

From this it follows directly

Corollary 2. *Under the conditions as in Kolmogorov's theorem. For each Diophantine $\boldsymbol{\omega}^*$, if the norm of the pertubation is sufficiently small, the original Hamiltonian system has an invariant torus, close to the unpertubed torus $\boldsymbol{I}^* \times \mathbb{T}^n$, and on this torus the motion is quasi periodic with frequency $\boldsymbol{\omega}^*$.*

It can be proved that the set of invariant tori has relative measure differing from 1 by at most quantities of order $\mathcal{O}(\epsilon^{1/4})$.

Poincaré Maps for Nearly Integrable Systems with 2 Degrees of Freedom

If we investigate a Poincaré map for a nearly integrable system with 2 degrees of freedom we know that, if the conditions of Kolmogorov's theorem are satisfied, then the trajectory is most likely lying on a torus $\mathbb{T}^2$ imbedded in $W_E \subset \mathbb{R}^3$. If we for such a system plot several Poincaré maps in W_E, E being the same for all of them, we then have that most of what we will see are closed curves lying inside closed curves, and since two tori $\mathbb{T}^2$ in $\mathbb{R}^3$ will confine trajectories in between, a motion, which is no longer lying on a torus, will still be confined to lie in a bounded region.

3.4.2 Nekhoroshev's Theorem

The KAM theorem was concerned with which tori persist under a pertubation of the system. The Nekhoroshev theory presents a rather different perspective to nearly integrable systems, since it is concerned with sufficient conditions on the system to insure that the action variables change only a

little along the motion.

The Nekhoroshev theorem can be stated as follows

Theorem 9 (Nekhoroshev)**.** *Consider a nearly integrable Hamiltonian system*

$$H(\boldsymbol{I},\boldsymbol{\varphi}) = h(\boldsymbol{I}) + \epsilon f(\boldsymbol{I},\boldsymbol{\varphi}) \qquad\qquad (\boldsymbol{I},\boldsymbol{\varphi}) \in B \times \mathbb{T}^n$$

and assume:

1. *H is analytic in the complex neighborhood $\mathcal{D}_\rho$ of $B \times \mathbb{T}^n$.*
2. *h is (l,m)-quasi-convex in B, namely the system*

$$|\boldsymbol{\omega}(\boldsymbol{I}) \cdot v| < l\|v\| \qquad\qquad |\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{I}}(\boldsymbol{I})v \cdot v| < m\|v\|^2$$

   has no solution for $\boldsymbol{I} \in B$.

Denote $\boldsymbol{\omega}_0 = \inf_{\boldsymbol{I}\in B} \|\boldsymbol{\omega}(\boldsymbol{I})\|$. Then there exist positive constants ϵ^*, $\mathcal{E}_1$, $\mathcal{E}_2$ depending on ρ, l, m, $\boldsymbol{\omega}_0$, $\|h\|_\rho$, $\|f\|_\rho$, and positive constants a and b depending only on n, such that any motion $(\boldsymbol{I}(t),\boldsymbol{\varphi}(t)$ with initial values $(\boldsymbol{I}^0,\boldsymbol{\varphi}^0) \in B \times \mathbb{T}^n$ satisfies

$$\|\boldsymbol{I}(t) - \boldsymbol{I}^0\| < \mathcal{E}_1 \left(\frac{\epsilon}{\epsilon^*}\right)^b \qquad \textit{for} \qquad |t| < \mathcal{E}_2 e^{(\epsilon^*/\epsilon)^a}$$

Proof. See [6] □

The class of Hamiltonian systems with Hamiltonian $H(\boldsymbol{I},\boldsymbol{\varphi}) = \boldsymbol{\omega} \cdot \boldsymbol{I}$, $\boldsymbol{\omega} \in \mathbb{R}^n$, which is clearly seen to be integrable, is called isochronous systems. A theorem regarding the stability of the actions for a pertubed isochronous system is the following, which is considerably easier to prove than Nekhoroshev's theorem

Theorem 10. *Consider a pertubed isochronous system*

$$H(\boldsymbol{I},\boldsymbol{\varphi}) = \boldsymbol{\omega} \cdot \boldsymbol{I} + \epsilon f(\boldsymbol{I},\boldsymbol{\varphi}) \qquad\qquad (\boldsymbol{I},\boldsymbol{\varphi}) \in B \times \mathbb{T}^n$$

Assume:

1. *f is analytic in the complex neighborhood $\mathcal{D}_\rho$ of $B \times \mathbb{T}^n$.*
2. *$\boldsymbol{\omega}$ is (γ,n)-Diophantine.*
3. *$\epsilon < \epsilon^* = c\mathcal{E}_3/\|f\|_\rho$, where $\mathcal{E}_3 = \gamma\rho_I\rho_\varphi^n$ and c is a specific positive constant depending on n.*

Then for initial values $(\boldsymbol{I}(0),\boldsymbol{\varphi}(0)) = (\boldsymbol{I}^0,\boldsymbol{\varphi}^0) \in B\times\mathbb{T}^n$ we have, with $\mathcal{E}_4 = \frac{c}{\gamma}$ and $a = \frac{1}{n+1}$, that

$$\|\boldsymbol{I}(t) - \boldsymbol{I}^0\| \le 3\frac{\epsilon}{\epsilon^*}\rho_I \qquad \textit{for} \qquad |t| < \mathcal{E}_4 e^{(\epsilon^*/\epsilon)^a}$$

Proof. See [6] □

These two theorems dealing with the stability of the actions of nearly integrable systems are some of the strongest theorems in this area.

3.5 Symplectic Integrators

As described in [10] a symplectic integrator $\Psi_h : \mathbb{R}^n \to \mathbb{R}^n$ is an integrator which is also a symplectic map. Examples include the symplectic Euler method, the Störmer/Verlet method and the splitting method

$$\Phi_h = \varphi_{h/2}^{H_1} \circ \varphi_h^{H_2} \circ \varphi_{h/2}^{H_1}$$

The motivation for studying symplectic integrators comes from the fact that the exact flow of Hamilton's equations is a symplectic map, and therefore when studying Hamilton's equations numerically a better performance could be anticipated if using a symplectic integrator.

Before going in to details with the important theorems on symplectic integrators, we will first describe a very usefull way of writting the Taylor series for the flow of a differential equation, and give some necessary definitions .

A function $\boldsymbol{h} : U \to U, U \subset \mathbb{R}^d$ or $U \subset \mathbb{C}^d$, is said to be analytic in $z_0 \in C^d$ if the Taylor series for $\boldsymbol{h}$ in a neighborhood of $\boldsymbol{z}_0$ converges. If we consider a differential equation given by the vector field $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$$

then if $\boldsymbol{f}$ is analytic the solution φ_t must be analytic (see e.g. [7]), therefore we have in a neighborhood of $t = 0$

$$\varphi_t(\boldsymbol{x}_0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\right|_{t=0} \varphi_t(\boldsymbol{x}_0)$$

Since for a differentiable function $g : \mathbb{R}^d \to \mathbb{R}$ we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} g(\varphi_t(\boldsymbol{x}_0)) &= \frac{\partial g}{\partial \boldsymbol{x}}(\varphi_t(\boldsymbol{x}_0)) \cdot \dot{\varphi}_t(\boldsymbol{x}_0) \\ &= \frac{\partial g}{\partial \boldsymbol{x}}(\varphi_t(\boldsymbol{x}_0)) \cdot \boldsymbol{f}(\varphi_t(\boldsymbol{x}_0)) \end{aligned}$$

The Lie derivative with respect to the vector field $\boldsymbol{f}$, $L_{\boldsymbol{f}}$, is for a differentiable function $g : R^d \to \mathbb{R}$ defined as

$$L_{\boldsymbol{f}} g = \frac{\partial g}{\partial \boldsymbol{x}}(\boldsymbol{x}_0) \cdot \boldsymbol{f}(\boldsymbol{x}_0)$$

and the Lie derivative with respect to the vector field f, for a differentiable function $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^c$, $\boldsymbol{g} = (g_1, \ldots, g_c)$, is defined as $L_{\boldsymbol{f}} \boldsymbol{g} = (L_{\boldsymbol{f}} g_1, \ldots, L_{\boldsymbol{f}} g_c)$. Using these expressions we have by induction

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\right|_{t=0} \boldsymbol{g}(\varphi_t(\boldsymbol{x}_0)) = \left(L_{\boldsymbol{f}}^k \boldsymbol{g}\right)(\boldsymbol{x}_0)$$

With this we get for $\boldsymbol{g}$ and $\boldsymbol{f}$ analytic in a neighborhood of $t = 0$

$$\begin{aligned}\boldsymbol{g}(\varphi_t(\boldsymbol{x}_0)) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{\mathrm{d}^k}{\mathrm{d}t^k}\bigg|_{t=0} \boldsymbol{g}(\varphi_t(\boldsymbol{x}_0)) \\ &= \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} L_{\boldsymbol{f}}^k \boldsymbol{g}\right)(\boldsymbol{x}_0) \\ &= \left(\exp(tL_{\boldsymbol{f}})\boldsymbol{g}\right)(\boldsymbol{x}_0)\end{aligned}$$

From this we see that for $\boldsymbol{f}$ analytic we have in a neighborhood of $t = 0$ that

$$\varphi_t(\boldsymbol{x}_0) = \left(\exp(tL_{\boldsymbol{f}})\boldsymbol{x}\right)(\boldsymbol{x}_0)$$

Two more definitions — which are quite similar to some of the definitions necessary for the theory on nearly integrable systems — are needed. Let $U \subset \mathbb{R}^d$ and $\sigma \in \mathbb{R}^d$, $\sigma_i > 0$, then define

$$\mathcal{U}_\sigma = \bigcup_{\boldsymbol{x} \in U} \{\boldsymbol{z} \in \mathbb{C}^d : |z_i - x_i| \leq \sigma_i,\ i = 1, ..., d\}$$

For a function $\boldsymbol{w} : \mathcal{U}_\sigma \to \mathbb{C}^d$ we define the norm $\|\cdot\|_\sigma$ as

$$\|\boldsymbol{w}\|_\sigma = \max_{i \in \{1,\ldots,d\}} \frac{\sup_{\boldsymbol{x} \in \mathcal{U}_\sigma} |w_i(\boldsymbol{x})|}{\sigma_i}$$

With this in hand we can state the main theorem regarding symplectic integrators as follows.

Theorem 11. *Consider the mapping $\Psi_\epsilon : U \to U$, $U \subset \mathbb{R}^d$*

$$\Psi_\epsilon(\boldsymbol{x}) = \boldsymbol{x} + \sum_{k=1}^{\infty} \epsilon^k \psi_k(\boldsymbol{x}) \qquad\qquad \epsilon \geq 0$$

Assume the functions ψ_k, as extensions of functions to $\mathcal{U}_\sigma$, are real analytic in $\mathcal{U}_\sigma$ and satisfy there the estimates

$$\|\psi_k\|_\sigma \leq \mu^{k-1}\Gamma$$

for some positive constants μ and Γ. Denote

$$\beta = 4\max(\mu, \Gamma)$$

Then there exists a formal series of vector fields

$$\boldsymbol{f}_\epsilon^\infty = \boldsymbol{f}_1 + \epsilon \boldsymbol{f}_2 + \epsilon^2 \boldsymbol{f}_3 + \ldots$$

analytic in $\mathcal{U}_\sigma$ such that:

1. *One has formally*

$$\Psi_\epsilon = \exp\left(\epsilon L_{\boldsymbol{f}_\epsilon^\infty}\right)\boldsymbol{x}$$

2. *The vector fields* $\boldsymbol{f}_k$ *satisfies the estimates*

$$\begin{aligned}\|\boldsymbol{f}_1\|_\sigma &\leq \Gamma \\ \|\boldsymbol{f}_k\|_{\sigma/2} &< \frac{1}{2}k^{k-1}\beta^{k-1}\mu \quad , \quad k \geq 2\end{aligned}$$

3. *Let* $\epsilon^* = \frac{1}{2e\beta}$*, then the flow* φ_t *of the vector field*

$$\boldsymbol{f}_\epsilon^r = \boldsymbol{f}_1 + \epsilon\boldsymbol{f}_2 + \ldots + \epsilon^{[\epsilon^*/\epsilon]-1}\boldsymbol{f}_{[\epsilon^*/\epsilon]}$$

satisfies

$$\|\varphi_\epsilon - \Psi_\epsilon\|_{\sigma/4} < 3\epsilon\Gamma e^{-[\epsilon^*/\epsilon]}$$

where $[\cdot]$ *gives the integer part.*

4. *If* Ψ_ϵ *is a symplectic map, then all the vector fields* $\boldsymbol{f}_1$*,* $\boldsymbol{f}_2$*, ... are Hamiltonian, i.e.* $\boldsymbol{f}_1 = \mathrm{J}^{-1}\nabla\widehat{H}_1$*,* $\boldsymbol{f}_2 = \mathrm{J}^{-1}\nabla\widehat{H}_2$*, ..., for some Hamiltonians* $\widehat{H}_1$*,* $\widehat{H}_2$*, ...*

Proof. See [5] □

The meaning of "formal" in this theorem simply means that $\boldsymbol{f}_\epsilon^\infty$ can be written in the stated way, but this series doesn't necessarily converge. Since Ψ_ϵ is analytical in $\epsilon = 0$ it's Taylor series converge, and item 1 in the theorem gives the way to determine the Taylor coefficients of Ψ_ϵ — even though the series $\boldsymbol{f}_\epsilon^\infty$ may not converge.

Thus according to 1 of this theorem we see that formally the map Ψ_ϵ is the flow for time ϵ of the formal vector field $\boldsymbol{f}_\epsilon^\infty$.

If we consider an analytic symplectic integrator Φ_h of order m approximating a solution to Hamiltons equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H$, with analytic Hamiltonian, we then get

$$\begin{aligned}\varphi_h^H - \Phi_h &= \exp\left(hL_{\mathrm{J}^{-1}\nabla H}\right)\boldsymbol{y} - \exp\left(hL_{(\mathrm{J}^{-1}\nabla\widehat{H}_1 + h\mathrm{J}^{-1}\nabla\widehat{H}_2 + \ldots)}\right)\boldsymbol{y} \\ &= \left(\mathrm{id} + hL_{\mathrm{J}^{-1}\nabla H} + \ldots + \frac{h^m}{m!}L^m_{\mathrm{J}^{-1}\nabla H} + \ldots\right)\boldsymbol{y} \\ &\quad - \left(\mathrm{id} + hL_{(\mathrm{J}^{-1}\nabla\widehat{H}_1 + h\mathrm{J}^{-1}\nabla\widehat{H}_2 + \ldots)} + \ldots + \frac{h^m}{m!}L^m_{(\mathrm{J}^{-1}\nabla\widehat{H}_1 + h\mathrm{J}^{-1}\nabla\widehat{H}_2 + \ldots)} + \ldots\right)\boldsymbol{y}\end{aligned}$$

but due to the order

$$\varphi_h^H - \Phi_h = \mathcal{O}(h^{m+1})$$

we get when collecting terms, using $L_{h\boldsymbol{f}} = hL_{\boldsymbol{f}}$, that

$$\begin{aligned}\widehat{H}_1 &= H \\ \widehat{H}_k &= 0 \quad k = 2, \ldots, m\end{aligned}$$

Combining this and theorem 11 we get

Corollary 3. *Let $\Phi_h : U \to U$, $U \subset \mathbb{R}^{2n}$, be a symplectic integrator of order m approximating a solution to Hamiltons equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H$, with analytic Hamiltonian, and assume that Φ_h satisfies the assumptions of theorem 11 and define σ, h^* and Γ according to this theorem. Then there exist a Hamiltonian*

$$\widehat{H} = H + h^m \widehat{H}_{m+1} + \ldots + h^{[h^*/h]-1}\widehat{H}_{[h^*/h]}$$

such that

$$\|\varphi_h^{\widehat{H}} - \Phi_h\|_{\sigma/4} < 3h\Gamma e^{-[h^*/h]}$$

The Hamiltonian $\widehat{H}$ is called the **modified Hamiltonian** of the symplectic integrator.

This corollary is very important in connection with numerical calculations using symplectic integrators, since a computer only has a finite precision. Therefore effectively a symplectic integrator can be constructed, by choosing the timestep h small enough, such that the numerical algorithm exactly, that is up to machine accuracy, solves a Hamiltonian system which is a pertubed version of the Hamiltonian system under consideration. This corollary thus gives a very usefull connection between symplectic integrators and pertubation theory for Hamiltonian systems.

Theorem 11 and corollary 3 is used to show the following proposition.

Proposition 9. *Let Φ_h, m, H, $\widehat{H}$, σ, h^*, and Γ be as in corollary 3, and suppose $\Phi_h(\boldsymbol{y}_0)$ stays within a compacs set K, where $\boldsymbol{y}_0 \in U$. Then there exist a positive constant λ such that*

$$\begin{aligned}\left|\widehat{H}(\Phi_h^i(\boldsymbol{y}_0)) - \widehat{H}(\boldsymbol{y}_0)\right| &\leq 3\lambda\Gamma e^{-\frac{[h^*/h]}{2}} \\ \left|H(\Phi_h^i(\boldsymbol{y}_0)) - H(\boldsymbol{y}_0)\right| &= \mathcal{O}(h^m)\end{aligned}$$

for $ih \leq e^{\frac{[h^*/h]}{2}}$.

Proof. Due to theorem 11 item 2 we know that there exist a Lipshitz constant λ such that for $\boldsymbol{y}_1, \boldsymbol{y}_2 \in U$ we have

$$|\widehat{H}(\boldsymbol{y}_2) - \widehat{H}(\boldsymbol{y}_1)| \leq \lambda\|\boldsymbol{y}_2 - \boldsymbol{y}_1\|_{\sigma/4}$$

and thus

$$\begin{aligned}|\widehat{H}(\Phi_h^i(\boldsymbol{y}_0)) - \widehat{H}(\boldsymbol{y}_0)| &\leq \sum_{j=1}^{i} |\widehat{H}\big(\Phi_h^j(\boldsymbol{y}_0)\big) - \widehat{H}\big(\varphi_h^{\widehat{H}}(\Phi^{j-1}(\boldsymbol{y}_0))\big)| \\ &\leq i\lambda \|\Phi_h - \varphi_h^{\widehat{H}}\|_{\sigma/4} \\ &\leq 3\lambda\Gamma e^{-\frac{[h^*/h]}{2}}.\end{aligned}$$

Since a continous function on a compact set attains a maximum we get using corollary 3

$$\begin{aligned}\big|H(\varphi_h^i(\boldsymbol{y}_0)) - H(\boldsymbol{y}_0)\big| &= \Big|\widehat{H}(\Phi_h^i(\boldsymbol{y}_0)) - \widehat{H}(\boldsymbol{y}_0) + h^m\widehat{H}_{m+1}(\boldsymbol{y}_0) - h^m\widehat{H}_{m+1}(\Phi_h^i(\boldsymbol{y}_0)) \\ &\quad + \ldots + h^{[h^*/h]-1}\widehat{H}_{[h^*/h]}(\boldsymbol{y}_0) - h^{[h^*/h]-1}\widehat{H}_{[h^*/h]}(\Phi_h^i(\boldsymbol{y}_0))\Big| \\ &\leq \Big|\widehat{H}(\Phi_h^i(\boldsymbol{y}_0)) - \widehat{H}(\boldsymbol{y}_0)\Big| + 2h^m \max_{\boldsymbol{y}\in K}\Big|\widehat{H}_{m+1}(\boldsymbol{y})\Big| \\ &\quad + \ldots + 2h^{[h^*/h]-1}\max_{\boldsymbol{y}\in K}\Big|\widehat{H}_{[h^*/h]}(\boldsymbol{y})\Big| \\ &\leq 2h^m M_{m+1} + \ldots + 2h^{[h^*/h]-1}M_{[h^*/h]} + 3\lambda\Gamma e^{-\frac{[h^*/h]}{2}}\end{aligned}$$

So the dominating term in this expression is seen to be of order $\mathcal{O}(h^m)$. □

Remark 3. Corollary 3 and proposition 9 shows why standard step size control cannot be immediately applied to a symplectic integrator without affecting it in a negative way. This comes from the fact that by regulating the step size the modified Hamiltonian will become time dependent and a time dependent Hamiltonian has not conservation of the Hamiltonian. Therefore a symplectic integrator with standard step size control can not be expected to have good properties when it comes to conservation of the Hamiltonian.

If we instead consider a completely general method Ψ_h of order m we get when performing the same estimates as in this proposition and again using theorem 11

$$\begin{aligned}\big|H(\Psi_h^i(\boldsymbol{y}_0)) - H(\boldsymbol{y}_0)\big| &\leq \sum_{j=1}^{i}\big|H(\Psi_h^j(\boldsymbol{y}_0)) - H(\Psi_h^{j-1}(\boldsymbol{y}_0))\big| \\ &\leq \lambda\sum_{j=1}^{i}\|\Psi_h^j(\boldsymbol{y}_0) - \Psi_h^{j-1}(\boldsymbol{y}_0)\|_{\sigma/4}\end{aligned}$$

Since $\|\Psi_h^j(\boldsymbol{y}_0)) - \Psi_h^{j-1}(\boldsymbol{y}_0)\|_{\sigma/4} = \mathcal{O}(h^{m+1})$ we get

$$\big|H(\Psi_h^i(\boldsymbol{y}_0)) - H(\boldsymbol{y}_0)\big| = \mathcal{O}(th^m)$$

where $t = ih$. Thus for a completely general nonsymplectic method we see that the error in conservation of energy grows up to linearly in time, so when it comes to energy properties symplectic integrators therefore behave superior to nonsymplectic methods.

Modified Hamiltonians for Symplectic Splitting Methods

The easiest methods to calculate modified Hamiltonians for are the splitting methods, for which the BCH theorem (see e.g. [10]) gives an elegant way of expressing them using the Poisson bracket. This also gives that for a symmetric splitting method the modified Hamiltonian only consists of even terms in h. Consider the Hamiltonian system with Hamiltonian

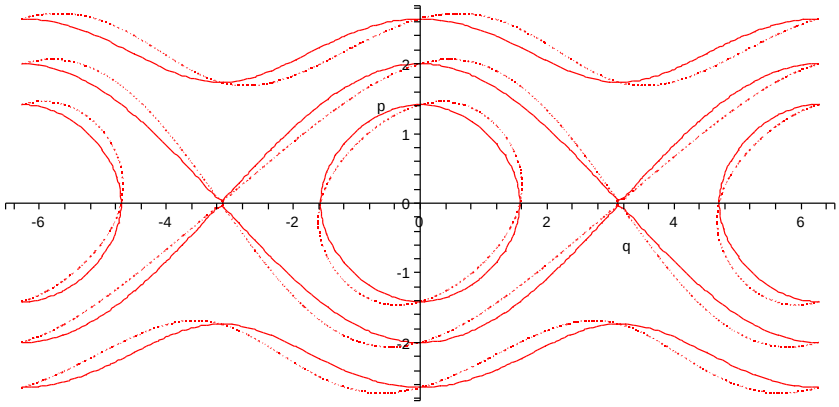


Figure 3.6: The pendulum. The solid curves are level curves for the pendulum Hamiltonian and the dashed curves are level curves of it's modified Hamiltonian for a 1st order splitting method.

$$H = H_1 + H_2$$

where H_1 and H_2 are analytic, and the two symplectic methods for solving Hamiltons equations $\dot{\boldsymbol{y}} = \mathrm{J}^{-1}\nabla H$

$$\Phi_h = \varphi_h^{H_1} \circ \varphi_h^{H_2} \qquad\qquad \Phi_h^S = \varphi_{h/2}^{H_1} \circ \varphi_h^{H_2} \circ \varphi_{h/2}^{H_1}$$

of order 1 and 2 respectively. Then the modified Hamiltonian $\widehat{H} = H + h\widehat{H}_2 + h^2\widehat{H}_3 + \mathcal{O}(h^3)$ for Φ_h is given by

$$\widehat{H}_2 = \frac{1}{2}\{H_1, H_2\}$$
$$\widehat{H}_3 = \frac{1}{12}\left(\{\{H_1, H_2\}, H_2\} + \{\{H_2, H_1\}, H_1\}\right)$$

whereas the modified Hamiltonian $\widehat{H}^S = H + h^2\widehat{H}_3^S + \mathcal{O}(h^4)$ for Φ_h^S is given by

$$\widehat{H}_3^S = -\frac{1}{24}\{\{H_2, H_1\}, H_1\} + \frac{1}{12}\{\{H_1, H_2\}, H_2\}$$

Higher order terms are likewise composed of Poisson brackets of Poisson brackets and so forth of H_1 and H_2.

Example 11. Consider the pendulum which has Hamiltonian

$$H = \frac{1}{2}p^2 - \cos(q)$$

If the Hamiltonian is split into it's kinetic energy $H_1 = \frac{1}{2}p^2$ and potential energy $H_2 = -\cos(q)$ then we get that the modified Hamiltonian for the first order symplectic splitting method $\Phi_h = \varphi_h^{H_1} \circ \varphi_h^{H_2}$ has first order element

$$\begin{aligned}
\widehat{H}_2 &= \frac{1}{2}\{H_1, H_2\} \\
&= \frac{1}{2}[p \;\; 0] \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \sin(q) \end{bmatrix} \\
&= -\frac{1}{2}p\sin(q)
\end{aligned}$$

Level curves for H and $H + h\widehat{H}_2$ are shown in figure 3.6 with $h = \frac{1}{2}$.

Bibliography

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics,* Springer (1989).

[2] G. Benettin, *The Elements of Hamiltonian Pertubation Theory,* Unpublished lecture notes (2001). Can be downloaded from www.math.unipd.it/∼benettin/

[3] G. Benettin, A. M. Cherubini, and F. Fassò, *A Changing-chart Symplectic Algorithm for Rigid Bodies and other Hamiltonian Systems on Manifolds,* SIAM J. Sci. Comput. **23**, 1189-1203 (2001).

[4] G. Benettin, L. Galgani, A. Giorgilli, and J. -M. Strelcyn *A Prooff of Kolmogorov's Theorem on invariant Tori Using Canonical Transformations Defined by the Lie Method,* Nuovo Cimento B **79**, 201-223 (1984).

[5] G. Benettin, and A. Giorgilli, *On the Hamiltonian Interpolation of Near-to-the-Identity Symlectic Mappings with Application to Symplectic Integration Algorithms,* Journal of Statistical Physics **74**, 1117-1143 (1994).

[6] A. M. Bloch, J. Baillieul, P. Crouch and J. E. Marsden, *Nonholonomic Mechanics and Control,* Springer (2003).

[7] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations,* McGraw-Hill (1955).

[8] I. M. Gelfand and S. V. Fomin, *Calculus of Variations,* Dover (2000).

[9] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations,* Springer (2002).

[10] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration Illustrated by the Störmer/Verlet Method,* Acta Numerica **12**, 399-450 (2003).

[11] A. J. Lichtenberg, and M. A. Lieberman, *Regular and Stochastic Motion,* Springer (1983).

[12] J. Jost, and X. Li-Jost, *Calculus of Variations,* Cambridge University Press (1999).

[13] J.E. Marsden, and T. S. Ratiu, *Introduction to Mechanics and Symmetry,* Springer (1994).

[14] E. B. Saff, and A. D. Snider, *Fundamentals of Complex Analysis for Mathematics, Science, and Engineering,* Prentice Hall (1993).

[15] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups,* Springer (1983).