ELAM XIV, Montevideo december 2005
Conservative dynamics and the calculus of variations

Alain Chenciner
Université Paris 7 & IMCCE
chenciner@imcce.fr

October 28, 2005

Abstract

Starting with elementary calculus of variations and Legendre transform, it is shown how the mathematical structures of conservative dynamics (Poincaré-Cartan integral invariant, symplectic structure, Hamiltonian form of the equations) arise from the simple computation of the variations of an action integral. The study of simple examples of integrable geodesic flows on the 2-torus then leads to the notion of Lagrangian submanifolds and to the Hamilton-Jacobi equation, whose relation to the Hamiltonian vector-field is the first step of the duality between particles and waves. The two last lectures are a brief introduction to KAM and weak KAM theories which describe what remains of complete integrability for more general hamiltonians, in particular for perturbations of integrable convex hamiltonians.[1] Due to the format of the course, only the easy results are proved while the harder (or longer) ones are admitted without remorse. This is, I hope, justified by the fact that it allows to give in a very short time some idea of the general architecture of the theory.

[1]The first three lectures are directly inspired by the course [C1] on weak KAM theory given by the author in Barcelona in July 2004. The last one owes much to the paper [Fe] and to discussions with Jacques Féjoz.

1 From the calculus of variations to the structures of conservative dynamics

1.1 Introduction

Classical mechanics (see [A]) deals in general with second order ordinary differential equations of the form

$$\ddot{q} = F(q, \dot{q}). \tag{E_1}$$

The terms depending on the time derivative $\dot{q}$ are termed "dissipative": they correspond to frictions (damping) or excitations. In their absence, one gets "conservative" equations $\ddot{q} = F(q)$ which are often of the form

$$\ddot{q} = \nabla U(q), \tag{E_2}$$

where U is a "potential function" and the gradient is relative to some Riemannian metric on the configuration space, which defines a "kinetic energy". The paradigmatic example is "the n-body problem", where the configuration $q = (\vec{r}_1, \vec{r}_2, , \vec{r}_n) \in E^n$ is the set of positions of n point masses in an euclidean space E and the equations are

$$\ddot{\vec{r}}_i = g \sum_{j \neq i} \frac{m_j(\vec{r}_j - \vec{r}_i)}{||\vec{r}_i - \vec{r}_j||_E^3}.$$

Here the m_i are positive masses, the potential function is

$$U(q) = \sum_{i<j} \frac{m_i m_j}{||\vec{r}_i - \vec{r}_j||_E}$$

and the Riemannian metric is defined by the (constant) scalar product

$$\langle (\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n), (\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n) \rangle = \sum_{i=1}^{n} m_i \langle \vec{r}_i, \vec{s}_i \rangle_E.$$

Such equations are known, since Lagrange, to be the so-called Euler-Lagrange equations of an "action functional", the Lagrangian action

$$L(q, \dot{q})) = \frac{1}{2} ||\dot{q}||^2 + U(q),$$

which is the difference between the kinetic energy $\frac{1}{2}||\dot{q}||^2$ and the potential energy $-U(q)$. This means that the solutions of (E_2) are exactly the set of "extremal" curves of the action functional. It is the mathematical formulation of the so-called "principle of least action". In the case when $U \equiv 0$, one gets the "geodesics" of the Riemannian metric. This origin makes natural the following "convexity" hypotheses:

General convexity hypotheses. The configuration space M will be either an open subset of $\mathbb{R}^n$ or the n-torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. The theory works with an arbitrary compact manifold but this hypothesis will allow us to work with global coordinates. The C^∞ (C^3 would be enough) Lagrangian $L(q,\dot q,t)$

$$L : TM \times \mathbb{R} = M \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$$

will be assumed to satisfy the "Tonelli" hypotheses which insure the existence of minimizers under natural hypotheses of coercivity:

1) L is strictly convex in $\dot q$, that is (in the sense of quadratic forms) :

$$\forall q, \dot q, t,\ \frac{\partial^2 L}{\partial \dot q^2}(q,\dot q,t) > 0;$$

2) L is superlinear in $\dot q$:

$$\forall C \in \mathbb{R}, \exists D \in \mathbb{R},\ \forall q, \dot q, t,\ L(q,\dot q,t) \geq C||\dot q|| - D,$$

that is $\lim_{||\dot q||\to\infty} \frac{L(q,\dot q,t)}{||\dot q||} = +\infty$ *uniformly in* (q,t).

1.2 The fundamental computation

The whole structure of classical conservative mechanics is the consequence of a single computation, the one giving the variation of the action

$$\mathcal{A}_L(\gamma) = \int_a^b L(\gamma(t), \dot\gamma(t), t)\, dt$$

of a path $\gamma : [a,b] \to M$ under an *arbitrary* variation of the path where neither the end-points nor the interval of variation of the parameter are fixed. Let us start with a regular (say at least C^2) path γ and consider a variation of γ, that is a family of paths $\gamma_u : [a(u), b(u)] \to M,\ u \in]-\epsilon, +\epsilon[$, regular with respect to both variables (u,t) and such that $a(0) = a, b(0) = b, \gamma_0 = \gamma$. The *infinitesimal variation* is by definition the vector-field on M along γ defined by

$$X(t) = \frac{\partial \Gamma}{\partial u}(0,t),$$

where we used the notation $\Gamma(u,t) = \gamma_u(t)$. It plays the role of a tangent vector at γ to the "manifold" of paths. More generally, we shall note $X_u(t) = \frac{\partial \Gamma}{\partial u}(u,t)$. The $X_u(t)$ are the velocity vectors of the "path of paths" $u \mapsto \gamma_u$. Thinking of $\gamma_u(t)$ as a point $q \in M$ depending on u and t, we shall also use the convenient notation $X_u = \frac{\partial q}{\partial u}$.

Computing the derivative of the function $u \mapsto \mathcal{A}_L(\gamma_u)$ via an integration by parts, one gets the following

Fundamental formula:

$$\frac{d}{du}\Big(\mathcal{A}_L\big(\gamma_u\big)\Big) = \frac{d}{du}\int_{a(u)}^{b(u)} L\left(\gamma_u(t), \dot\gamma_u(t), t\right)\, dt$$

$$= \int_{a(u)}^{b(u)} \left[\left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot q}\right)\big(\gamma_u(t), \dot\gamma_u(t), t\big)\right]\cdot X_u(t)\, dt$$

$$+\frac{\partial L}{\partial \dot q}\big(\gamma_u(t), \dot\gamma_u(t), t\big)\cdot X_u(t)\,\Big|_{t=b(u)} - \frac{\partial L}{\partial \dot q}\big(\gamma_u(t), \dot\gamma_u(t), t\big)\cdot X_u(t)\,\Big|_{t=a(u)}$$

$$+L\big(\gamma_u(t), \dot\gamma_u(t), t\big)\frac{db}{du}(u)\,\Big|_{t=b(u)} - L\big(\gamma_u(t), \dot\gamma_u(t), t\big)\frac{da}{du}(u)\,\Big|_{t=a(u)},$$

a formula that we shall abreviate in

$$\frac{d\mathcal{A}_L}{du} = \int_a^b \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot q}\right)\cdot\frac{\partial q}{\partial u}\, dt \;+\; \left[\frac{\partial L}{\partial \dot q}\cdot\frac{\partial q}{\partial u} + L\frac{dt}{du}\right]_a^b.$$

Restricting to variations parametrized by a fixed interval $[a,b]$ and such that the end-points are fixed, i.e. $\gamma_u(a) = \gamma(a)$ and $\gamma_u(b) = \gamma(b)$, one gets the classical *Euler-Lagrange* equations for the *extremals*, that is the paths γ such that $d\mathcal{A}_L(\gamma)X = 0$ for any infinitesimal variation X with fixed interval and fixed end points (i.e. $X(a) = 0$ and $X(b) = 0$):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot q_i}\big(\gamma(t), \dot\gamma(t), t\big)\right) = \frac{\partial L}{\partial q_i}\big(\gamma(t), \dot\gamma(t), t\big), \quad i = 1, \cdots, n. \qquad (E)$$

In order to put these equations into a nice "explicit" form, we notice that the "general hypotheses" we made on L imply that the *Legendre mapping*

$$\Lambda : TM\times\mathbb{R} = M\times\mathbb{R}^n\times\mathbb{R} \to (\mathbb{R}^n)^*\times M\times\mathbb{R} = T^*M\times\mathbb{R}$$

defined by

$$\Lambda(q,\dot q,t) = (p,q,t), \quad p = \frac{\partial L}{\partial \dot q}(q,\dot q,t),$$

is a global diffeomorphism (strict convexity for all p of $\dot q \mapsto L(q,\dot q,t) - p\cdot\dot q$ implies the injectivity of Λ and surlinearity implies that it is proper, hence surjective). One says that L is *globally regular*. From this, two important results follow:

1) Regularity of extremals: any extremal is as regular as L. This means that if we had assumed paths to be only C^0 and piecewise C^1, and its variations accordingly, the extremals would still be as regular as the Lagrangian. This justifies our working only with regular paths. In the case of minimizers, one could even work with absolutely continuous paths. In fact, small enough extremals are minimizers and their regularity amounts, as in the case of straight lines, to the remark that a broken curve can always be shortened by smoothing the angle.

2) Existence of the Euler-Lagrange flows: it follows from the fact that Λ is a diffeomorphism that equations (E) define a (time-dependant if L is) vector-field X_L in TM and X_H^* in T^*M (the notation X_H^* will be explained below). For example, in T^*M:

$$\frac{dp_i}{dt} = \frac{\partial L}{dq_i} \circ \Lambda^{-1}(p,q,t), \ \frac{dq_i}{dt} = \dot{q}_i \circ \Lambda^{-1}(p,q,t) \qquad (X_H^*)$$

These vector-fields are intrinsically defined (i.e. they do not depend on the choice of local or global coordinates on M). Their flows will both be called the *Euler-Lagrange flow.*

Indeed, their variational origin implies that the Euler-Lagrange equations (E) take exactly the same form in any local or global coordinate system. In other words, the mapping $[L]_\gamma : [a,b] \to T^*M$ defined by

$$[L]_\gamma(t) = \frac{\partial L}{\partial q}\big(\gamma(t),\dot{\gamma}(t),t\big) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\big(\gamma(t),\dot{\gamma}(t),t\big)\right) \in T^*_{\gamma(t)}M$$

is an intrinsically defined field of covectors tangent to M "along γ" and the derivative of the action (for variations with fixed interval and fixed end-points) can be written

$$d\mathcal{A}_L(\gamma)\cdot X = \int_a^b [L]_\gamma(t)\cdot X(t)\,dt.$$

1.3 The Poincaré-Cartan integral invariant

From the fundamental formula, it follows that, if γ_u is a family of extremals of the action $\mathcal{A}_L = \int L\,dt$, we get

$$\frac{d\mathcal{A}_L}{du} = \left[\frac{\partial L}{\partial \dot{q}}\cdot\frac{\partial q}{\partial u} + L\,\frac{dt}{du}\right]_a^b.$$

We replace now the partial derivative $\frac{\partial q}{\partial u}$ (that is $\frac{\partial \Gamma}{\partial u}$), deprived of geometric meaning, by the "effective variation"

$$\frac{d}{du}\Big(\Gamma\big(u,t(u)\big)\Big) = \frac{dq}{du} = \frac{\partial q}{\partial u} + \frac{\partial q}{\partial t}\frac{dt}{du} = \frac{\partial q}{\partial u} + \dot{q}\,\frac{dt}{du}, \quad t(u) = a(u) \text{ or } b(u),$$

of the extremities of the path γ_u as a fonction of u (figure 1).

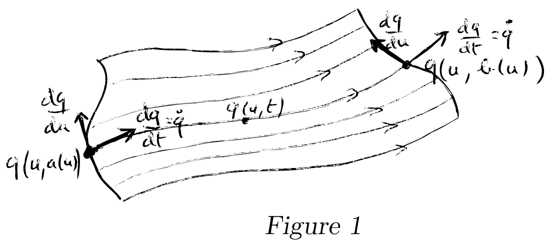

Figure 1

This transforms the expression of $\frac{d}{du}\left(\mathcal{A}_L(\gamma_u)\right)$ for a family of extremals into an identity between *differential 1-forms* on the interval $\mathcal{U}$ of definition of the parameter u :

$$d\mathcal{A}_L = \delta_b^*\varpi_L - \delta_a^*\varpi_L,$$

where $\delta_a, \delta_b : \mathcal{U} \to T^*\mathbb{T}^n \times \mathbb{R}$ denote the mappings

$$\delta_t(u) = \left(\Gamma\big(u, t(u)\big), \frac{\partial \Gamma}{\partial t}\big(u, t(u)\big), t(u)\right), \quad t(u) = a(u) \text{ or } b(u),$$

and ϖ_L is the differential 1-form on $T\mathbb{T}^n \times \mathbb{R}$ defined by

$$\varpi_L = \frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t) \cdot dq - \left(\frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t) \cdot \dot{q} - L(q, \dot{q}, t)\right) dt.$$

Finally, we can simplify the formulas by transporting everything on the cotangent side with the Legendre diffeomorphism Λ. The function on $T^*\mathbb{T}^n \times \mathbb{R}$ defined by

$$H(p, q, t) = p \cdot \dot{q} - L(q, \dot{q}, t),$$

where $\dot{q}$ is expressed in terms of p, q, t via Λ is called the *Legendre transform* of L, or the *Hamiltonian* associated to the Lagrangian L. If ϖ_H denotes the 1-form on $T^*\mathbb{T}^n \times \mathbb{R}$ defined by

$$\varpi_H = p \cdot dq - H(p, q, t)dt,$$

the formula for the unconstrained variations of extremals becomes

$$d\mathcal{A}_L = (\Lambda \circ \delta_b)^*\varpi_H - (\Lambda \circ \delta_a)^*\varpi_H.$$

The 1-form ϖ_H is the *Poincaré-Cartan integral invariant* (*tenseur impulsion-énergie* in Cartan's terminology).

Rewriting the action. As $L = p \cdot \dot{q} - H$, the action istself can now be written as the integral of $\varpi_H = p \cdot dq - H dt$ on the *lift* $\Gamma^*(t) = \left(\frac{\partial L}{\partial \dot{q}}(\gamma(t), \dot{\gamma}(t), t), \gamma(t), t\right)$ to $T^*\mathbb{T}^n \times \mathbb{R}$ of the path $\gamma(t)$ in $\mathbb{T}^n$:

$$\mathcal{A}_L(\gamma) = \int_{\Gamma^*} \varpi_H.$$

This expression is the basis of *Hamilton's least action principle.*

In order to understand where it comes from, it is enough to apply the general formula for the variation of the action to a family of paths which differ only by the domain of definition (that is, all the γ_u are the restrictions to some time interval of one and the same path).

1.4 The symplectic structure and Hamilton's equations

A paraphrase of equations (E) is that a path $t \mapsto \gamma(t)$ in $\mathbb{T}^n$ is an extremal if and only if the parametrized curve in $T^*\mathbb{T}^n \times \mathbb{R}$

$$t \mapsto \left(\frac{\partial L}{\partial \dot{q}}\big(\gamma(t), \dot{\gamma}(t), t\big), \gamma(t), t \right) = \Lambda\big(\gamma(t), \dot{\gamma}(t), t\big)$$

is an integral curve of the (time-dependant) vector-field

$$\Xi_H^* = (X_H^*, 1) = \Lambda_*(X_L, 1)$$

on $T^*\mathbb{T}^n \times \mathbb{R}$. The last formula of the preceding section then implies that, if C_a and C_b are two oriented *loops* in $T^*\mathbb{T}^n \times \mathbb{R}$, such that $C_b - C_a$ is the oriented boundary of a cylinder C generated by pieces of of integral curves of $\Xi_H^* = (X_H^*, 1)$, one has

$$\int_{C_a} p \cdot dq - H(p,q)dt = \int_{C_b} p \cdot dq - H(p,q)dt\,.$$

In Cartan's terminology, $\varpi_H = p \cdot dq - Hdt$ is a *relative* and *complete* integral invariant : relative because its invariance holds only if the integral is taken on loops C_i, complete because C_a and C_b are not supposed to be contained in slices where t is constant (i.e. C_b is not supposed to be the image of C_a under the element φ_a^b of the flow of Ξ_H^*.

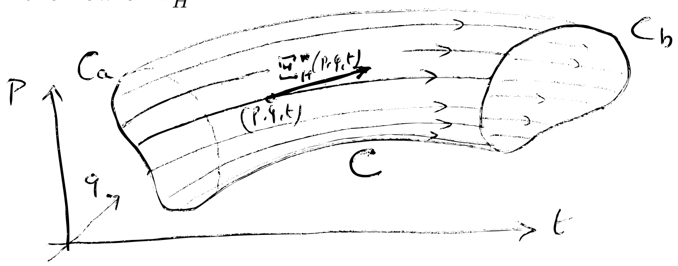

Figure 2

Applying Stokes formula to small disks D_a et D_b contained respectively in the time slices $T^*\mathbb{T}^n \times \{a\}$ and $T^*\mathbb{T}^n \times \{b\}$ and such that $D_b = \varphi_a^b(D_a)$ is the image of D_a under the flow of Ξ_H^*, one gets the

Theorem 1 *The time-dependant vector-field X_H^* defined on $T^*\mathbb{T}^n$, preserves the* standard symplectic 2-form $\omega = \sum_{i=1}^n dp_i \wedge dq_i$.

A corollary of the preservation of the symplectic structure is

Theorem 2 (Liouville's theorem) *The flow of the time-dependant vector-field X_H^* preserves the $2n$-form ω^n, hence the Lebesgue measure (volume).*

Hamilton's equations. We now deduce the structure of the vector-field X_H^* (i.e. the structure of the Euler-Lagrange equations (E) seen from the cotangent side) from the following characterization of integral invariants:

(K) The 1-form ϖ_H *is an integral invariant of the vector-field* Ξ_H^* *if and only if, at each point* $(p,q,t) \in T^*\mathbb{T}^n \times \mathbb{R}$, *the vector* $\Xi_H^*(p,q,t)$ *belongs to the kernel of the bilinear form* $d\varpi_H(p,q,t)$, *i.e. if* $i_{\Xi_H^*} d\varpi_H = 0$.

The proof is a consequence of Stokes formula applied to oriented cylinders.

This determines the direction of Ξ_H^*, hence X_H^*, because the kernel of

$$d(p \cdot dq - H dt) = \sum_{i=1}^{n} \left[\left(dp_i + \frac{\partial H}{\partial q_i} dt \right) \wedge \left(dq_i - \frac{\partial H}{\partial p_i} dt \right) \right],$$

is easily seen to be 1-dimensional and generated at each point (p,q,t) by the vector $\left(-\frac{\partial H}{\partial q}(p,q,t), \frac{\partial H}{\partial p}(p,q,t), 1 \right)$.

Finally, we get

$$X_H^* = \left(-\frac{\partial H}{\partial q_1}, \cdots - \frac{\partial H}{\partial q_n}, \frac{\partial H}{\partial p_1}, \cdots \frac{\partial H}{\partial p_n} \right).$$

Hence, when transported in $T^*\mathbb{T}^n$ by the Legendre diffeomorphism, the Euler-Lagrange equations (E) take the particularly symmetric form of *Hamilton's equations* (or *canonical equations*) :

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, i = 1 \cdots n, \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, i = 1 \cdots n\,.$$

As the equations depend only on H, this justifies the notation X_H^*. It is fair to remember that this particularly symmetric form of the equations of classical mechanics already appear in Lagrange's works.

Symplectic changes of coordinates. If $\Phi(p,q) = (a,b)$ is symplectic, that is if $dp \wedge dq = da \wedge db$ (or more correctly $\Phi^*(\omega) = \omega$), the direct image of the Hamiltonian vector-field X_H^* is the Hamiltonian vector-field $X_{H\circ\Phi^{-1}}^*$.

The Legendre transform in the convex case. It follows from Hamilton's equations that the Legendre transform $L \mapsto H$ is involutive :

$$H(p,q,t) = p \cdot \dot{q} - L(q,\dot{q},t), \quad p = \frac{\partial L}{\partial \dot{q}}(q,\dot{q},t),$$

$$L(q,\dot{q},t) = p \cdot \dot{q} - H(p,q,t), \quad \dot{q} = \frac{\partial H}{\partial p}(p,q,t).$$

This symmetry makes it natural to write the correspondance $L \leftrightarrow H$ in the following form, where the variables (q,t) play the role of mere parameters :

$$p \cdot \dot{q} = L(q,\dot{q},t) + H(p,q,t).$$

The convexity of $\dot{q} \mapsto L(q,\dot{q},t)$ is equivalent to that of $p \mapsto H(p,q,t)$ and if a function satisfies the general convexity hypotheses, so does its transform.

Young-Fenchel inequality. For all $q, t, \dot{q}, p$, the following holds :

$$p \cdot \dot{q} \le L(q,\dot{q},t) + H(p,q,t),$$

with equality if and only if $p = \dfrac{\partial L}{\partial \dot{q}}(q,\dot{q},t)$.

Figure 3 illustrates in dimension 1 this variational definition of the Legendre transform. One also reads on this figure the interpretation of the transform as the passage from a punctual to a tangential equation.

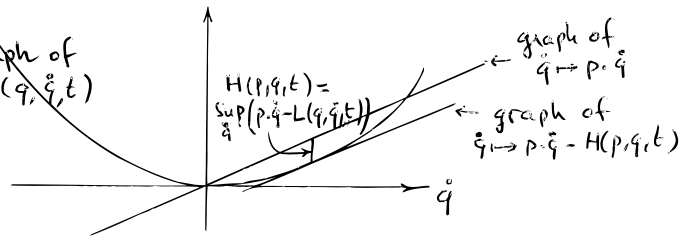

Figure 3

Autonomous Lagrangians. These are the Lagrangians $L(q, \dot{q})$ which do not depend explicitely on time. It follows from Hamilton's equations that the (autonomous) vector-field X_H^* preserves the Hamiltonian H. In mechanics, this is the preservation of the *total energy.*

An elegant way of proving this property is to notice that the property (K), that is $i_{\Xi_H^*} d(p \cdot dq - Hdt) = 0$, is equivalent to

$$i_{X_H^*}\omega \;=\; -dH,$$

where ω is the symplectic form (in the time-dependant case, one must replace dH by $\partial H = dH - \frac{\partial H}{\partial t}$). Hence X_H^* is characterized by the property that, for any vector field Y on $T^*\mathbb{T}^n$, one has

$$\omega(X_H^*, Y) = -dH \cdot Y.$$

By analogy with the gradient of a Riemannian metric, one calls X_H^* the *symplectic gradient* of H. The conservation of energy amounts now to the identity

$$L_{X_H^*} H = dH(X_H^*) = -\omega(X_H^*, X_H^*) = 0.$$

An important feature of autonomous Hamiltonian systems is that *up to the parametrization, integral curves of the flow of X_H^* are completely determined by the sole geometry of the level hypersurfaces of H* : this is clear on figure 4 : the direction of $\mathrm{grad}_\omega H$ depends only on the direction of $\mathrm{grad}\, H$ and not on its length or orientation.

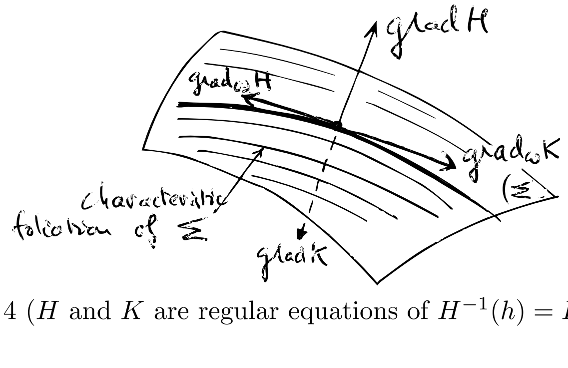

Figure 4 (H and K are regular equations of $H^{-1}(h) = K^{-1}(k)$ at x)

From time-dependant to time-independant. A time-dependant system can always be embedded into a time-independant one at the expense of adding dimensions and loosing track of time origin : indeed, the vector-field X_K^* on $T^*(\mathbb{T}^n \times \mathbb{R})$ corresponding to the *extended* Hamiltonian

$$K(p, E, q, \tau) = E + H(p, q, \tau)$$

restricts to $\Xi_H^* = (X_H^*, 1)$ when one identifies the energy hypersurface $K^{-1}(0)$ with $T^*\mathbb{T}^n \times \mathbb{R}$.
This extension may be useful even if H does not depend on time : because of the last component equal to 1, the geometry of the energy hypersurface $K \equiv E + H(p, q, t) = 0$ determines completely the vector-field Ξ_H^*, hence X_H^*.

The example of classical mechanics. The Lagrangian is the difference

$$L(q, \dot{q}) = \frac{1}{2}\, \dot{q} \cdot G(q)\dot{q} - V(q) = \frac{1}{2} g(q)(\dot{q}, \dot{q}) - V(q)$$

between *kinetic* and *potential* energy. The kinetic energy is defined by a Riemannian metric g on $M = \mathbb{T}^n$, that is for each q a positive definite quadratic form $g(q)$, represented by a symmetric matrix $G(q)$. When there is no potential V, the extremals are the *geodesics* of the metric. The Legendre transform $p = G(q)\dot{q}$ defines the conjugate momenta (the *impulsions*) p_i of the configuration variables q_i, the $\dfrac{\partial L}{\partial q_i}$ are the *forces* and the Hamiltonian is total energy, i.e. the sum of kinetic and potential energies

$$H(p, q) = \frac{1}{2}\, \dot{q} \cdot G(q)\dot{q} + V(q) = \frac{1}{2} p \cdot G(q)^{-1} p + V(q).$$

2 Complete integrability and the Hamilton-Jacobi equation

2.1 The simplest example of a completely integrable system : the geodesic flow of a flat torus

The Lagangian $L : T^*\mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2 \times \mathbb{R}^2 \to \mathbb{R}$ is $L(q, \dot{q}) = \frac{1}{2}||\dot{q}||^2$. We shall write the coordinates $q = (\varphi, \psi)$ and $\dot{q} = (\dot{\varphi}, \dot{\psi})$ (figure 5).

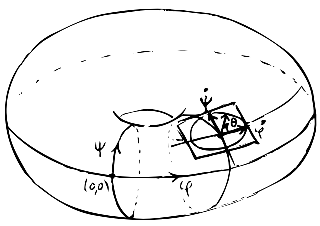

Figure 5

The Euler-Lagrange equation (E) is $\ddot{q} = 0$ and the extremals, the geodesics of $\mathbb{T}^2$, are the images by the canonical projection of the straight lines of $\mathbb{R}^2$ with an affine parametrization. The Legendre diffeomorphism is defined by $p = q$ and fixing the energy $H(p, q) = \frac{1}{2}||p||^2$ amounts to fixing the norm of the velocity. If the energy is different from 0, the energy hypersurface is diffeomorphic to $\mathbb{T}^3 = (\mathbb{R}/2\pi\mathbb{Z})^3$. The flow is depicted on figure 6.

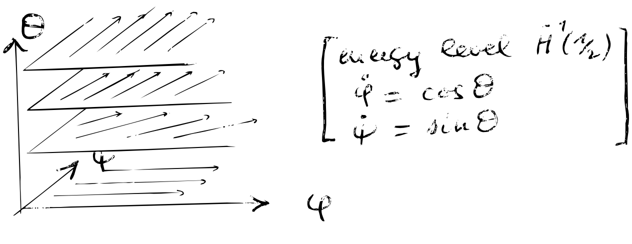

Figure 6

The whole *phase space* $T\mathbb{T}^2$ (or $T^*\mathbb{T}^2$) is foliated by the 2-dimensional tori $\dot{q} =$ constant (or $p =$ constant) which are invariant under the flow of X_L (or X_H^*). On these tori, the vector-field is constant (the flow is a flow of tranlations) and, depending on the rationality or irrationality of $\dot{\psi}/\dot{\varphi}$, the integral curves on the torus are all periodic or all dense.

Notice that the tori on which the integral curves are dense have a dynamical definition, as the closure of any of the integral curves they contain. This is not the case of the "periodic" tori which are a mere union of closed integral curves.

2.2 Opening of a resonance : the geodesic flow of a torus of revolution

We embed the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2$ in $\mathbb{R}^3$ by the mapping $(r < 1)$

$$(\varphi, \psi) \mapsto \big((1 + r\cos\psi)\cos\varphi, (1 + r\cos\psi)\sin\varphi, r\sin\psi\big).$$

The image is invariant under rotation around the z-axis. The extremals of the Lagrangian

$$L(\varphi, \psi, \dot{\varphi}, \dot{\psi}) = \frac{1}{2}\left((1 + r\cos\psi)^2\dot{\varphi}^2 + r^2\dot{\psi}^2\right)$$

are the geodesics of the induced metric, parametrized proportionnally to arc length.
The Euler-Lagrange equations are

$$\frac{d}{dt}\big(1 + r\cos\psi)^2\dot{\varphi}\big) = 0, \quad \frac{d}{dt}\big(r^2\dot{\psi}\big) = -r\sin\psi(1 + r\cos\psi)\dot{\varphi}^2.$$

The first expresses the invariance under rotation around Oz and can be interpreted as the *conservation of the angular momentum around Oz*. It is the analogue of the conservation of the angle θ in the flat case. Fixing the energy is fixing the velocity and the non-zero energy levels are diffeomorphic to the unit tangent bundle $T^1\mathbb{T}^2 \equiv \mathbb{T}^3$ with global angular coordinates (φ, ψ, θ) defined by choosing as third coordinate the *Riemannian angle* θ :

$$\dot{\varphi} = \frac{\cos\theta}{1 + r\cos\psi}, \quad \dot{\psi} = \frac{\sin\theta}{r}\ .$$

The first Euler-Lagrange equation becomes the constancy of the *Clairaut integral* :

$$(1 + r\cos\psi)\cos\theta = \text{ constant.}$$

Figures 8 represents the level curves of this function in the plane (ψ, θ). Figure 7 represents the level curves of the function θ, which plays for the flat torus the role of the Clairaut integral.

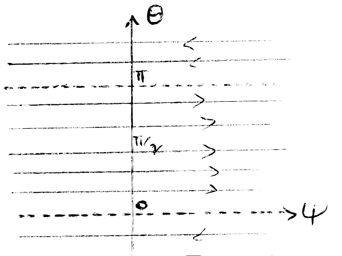

Figure 7

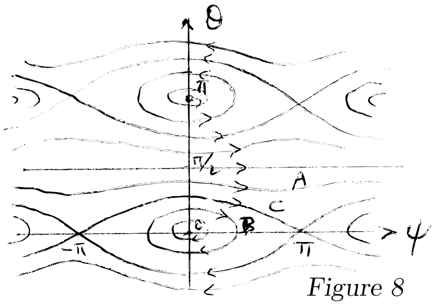

Figure 8

In the coordinates (φ, ψ, θ), the equations become

$$\frac{d\varphi}{dt} = \frac{\cos\theta}{1 + r\cos\psi}, \quad \frac{d\psi}{dt} = \frac{\sin\theta}{r}, \quad \frac{d\theta}{dt} = \frac{-\cos\theta\sin\psi}{1 + r\cos\psi}\ .$$

Because of the invariance under rotation, they are independant of φ, hence they admit a direct image in the torus (ψ, θ) which consists in ignoring the first equation. The same is of course true for the flat metric. The integral curves of this direct image are contained in the level curves of the Clairaut integral, which explains the arrows of figures 7 and 8.

In each open band $\theta \in \left]-\frac{\pi}{2} + k\pi, \frac{\pi}{2} + k\pi\right[$, $k \in \mathbb{Z}$, the flow looks qualitatively like the flow of a conservative pendulum. The rotations of the pendulum correspond to integral curves belonging to invariant tori which, as in the flat case, project biunivocally onto the configuration torus (φ, θ) (figure 9),

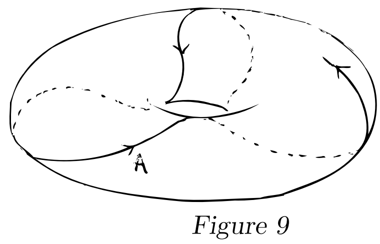

Figure 9

while oscillations correspond to integral curves belonging to invariant tori which project neither injectively nor surjectively but on an annulus whose boundary is a *caustic* (figure 10). The latter tori fill the *resonance zone.*

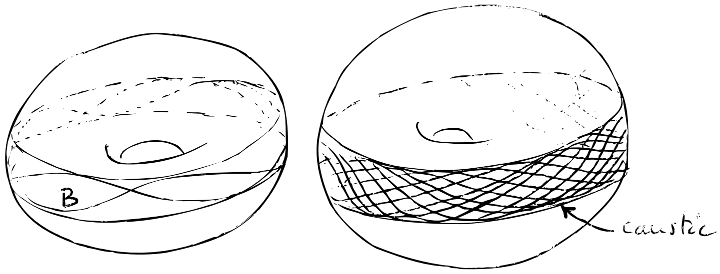

Figure 10

As in the flat case, in each of these invariant tori, integral curves are either all periodic or all dense.
A new feature is the existence in each non-zero energy level of 4 isolated periodic solutions, which correspond to the 2 geodesics defined by the intersection of the torus with the plane $z = 0$, each one with two possible directions of velocity. The inner one ($\psi = \pi, \theta = 0$ or $\psi = \pi, \theta = \pi$) is *hyperbolic* hence *unstable.* The set of integral curves with the same energy which are positively (negatively) asymptotic to it define the *stable* (*unstable*) manifold of this periodic orbit. These sets happen to coincide here. A corresponding geodesic is represented on figure 11. Their union is a surface which makes the transition between the two kinds of invariant tori oustside and inside the resonance zone. The outer one ($\psi = 0, \theta = 0$ or $\psi = 0, \theta = \pi$) is *elliptic*, hence stable. In its energy level, it is "surrounded" by invariant tori.

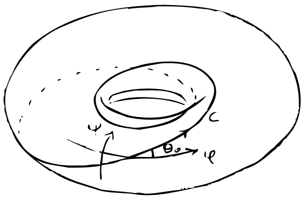

Figure 11

We have now two kinds of invariant sets dynamically defined : the invariant tori with dense integral curves and the stable = unstable manifolds of the hyperbolic periodic solutions.

2.3 Lagrangian submanifolds as geometric solutions of the Hamilton-Jacobi equation

In both exemples above, most of the *phase space* $T^*\mathbb{T}^2$ is foliated by invariant tori on which the flow of X_H^* can be shown to be a flow of translations in well chosen coordinates. This is obvious for the flat torus and a consequence of the invariance under rotation for the torus of revolution. The existence of such a foliation is a characteristic feature of the so-called *completely integrable* autonomous Hamiltonian systems. The following lemma shows that these tori are very special :

Definition 3 *A submanifold V of $T^*\mathbb{T}^n$ is called isotropic if the pull-back $j^*\omega$ of the canonical symplectic form is identically zero, where $j : V \to T^*\mathbb{T}^n$ is the canonical inclusion.*

Lemma 4 *If the restriction of the flow of X_H^* to an invariant torus $\mathcal{T}$ is a flow of translations with dense orbits, $\mathcal{T}$ is isotropic.*

The proof is a consequence of the fact that $\omega = d\lambda$, where $\lambda = \sum_{i=1}^n p_i dq_i$ is the *Liouville form* on $T^*\mathbb{T}^n$. If $j^*\omega = \sum_{i<j} a_{ij}(u_1, \cdots, u_k) du_i \wedge du_j$ in coordinates $u_1, \cdots, u_k$ on $\mathcal{T}$ such that the flow of X_H^* becomes a flow of translations $\Phi_t(u) = u + tv$, the fact that $\Phi^*\omega = \omega$ implies that the functions a_{ij} are constant along the integral curves contained in $\mathcal{T}$ *(this would not be the case if $d\Phi_t(u)$ was not the Identity)*. As these integral curves are dense, the a_{ij} are constant, hence equal to 0 because $j^*\omega = d(j^*\lambda)$ is a coboundary.

Notice that in the completely integrable cases that we studied above, an easy argument of continuity implies that all invariant tori (and not only the ones with dense integral curves), and also the stable = unstable invariant manifolds of the hyperbolic periodic solutions, share the property $j^*\omega = 0$. This property will play a fundamental role in the sequel :

Definition 5 (Definition-Proposition) *The dimension of an isotropic submanifold of $T^*\mathbb{T}^n$ is at most n. If it is equal to n, the submanifold is called Lagrangian.*

The bound on the dimension is an exercise in symplectic algebra : at each point (p, q), the bilinear form $\omega(p, q)$ is non degenerate, hence an isotropic subspace (i.e. a linear subspace contained in its ω-orthogonal) is at most of dimension $2n/2 = n$.

Each invariant Lagrangian submanifold that we found in the integrable examples is contained in a single energy level. This is a consequence of the conservation of energy when the submanifold is the closure of a single solution and the others follow by continuity. This property has a very important converse :

Proposition 6 . *let $H : T^*\mathbb{T}^n \to \mathbb{R}$ be an autonomous Hamiltonian. Every Lagrangian submanifold V of $T^*\mathbb{T}^n$ contained in a regular energy level $H^{-1}(h)$ is invariant under the flow of X_H^*.*

The proof is again an exercise in symplectic algebra : because of the maximality of the dimension of V among isotropic submanifolds, it is enough to show that at each point $m \in H^{-1}(h)$, the vector $X_H^*(m)$ belongs to (in fact generates) the kernel of $i_h^*\omega(m) = d(i_h^*\lambda)(m)$, where i_h is the canonical injection of $H^{-1}(h)$ in $T^*\mathbb{T}^n$ and $\lambda = p \cdot dq$ is the Liouville form. Indeed, if $X_H^*(m)$ was not contained in T_mV, the linear subspace generated by $X_H^*(m)$ and T_mV would be isotropic of dimension $n+1$, a contradiction.

We have already proved this when $X_H^*|_{H^{-1}(h)}$ is replaced by Ξ_H^*, $H^{-1}(h)$ is replaced by $K^{-1}(0) \equiv T^*\mathbb{T}^n \times \mathbb{R} \subset T^*(\mathbb{T}^n \times \mathbb{R})$, and the 1-form $i_h^*\lambda$ is replaced by the Poincaré-Cartan integral invariant $p \cdot dq - Hdt$. As this is the only case that we need, we leave the general assertion as an exercise.

Remark. A Hamiltonian flow is a very particular one as it preserves the symplectic 2-form ω, hence in particular the volume. *Its restriction to a Lagrangian submanifold V, on the contrary, does not satisfy any a priori constraint : every vector-field X on V is the restriction of a Hamiltonian flow defined on a neighborhood of V.* The simplest example is obtained when $V \equiv \mathbb{T}^n$ is the zero-section $p = 0$ of $T^*\mathbb{T}^n$: if $X(q)$ is vector-field on V, the Hamiltonian $H(p, q) = p \cdot X(q)$ is such that the restriction of X_H^* to V coincides with X (but it is not convex in p !).

Lagrangian graphs and the Hamilton-Jacobi equations. All invariant tori of the geodesic flow of a flat torus are graphs of a mapping $q \mapsto p(q)$, that is sections of the projection $(p, q,) \mapsto q$. For the torus of revolution, only those not contained in the resonance zone are graphs in the same way. The invariant manifolds of the hyperbolic periodic solutions are the union of two pieces, each of which is a graph.

Lemma 7 *If the Lagrangian submanifold V of $T^*\mathbb{T}^n = (\mathbb{R}^n)^* \times \mathbb{T}^n$ is a graph, it is the graph of a mapping of the form $p = a + ds(q)$, where $a = (a_1, \cdots, a_n) \in (\mathbb{R}^n)^*$ and $s : \mathbb{T}^n \to \mathbb{R}$.*

The proof is an easy calculation : the graph V of the mapping $q \mapsto p(q)$ is Lagrangian if and only if the 2-form $\sum_{i=1}^n dp(q) \wedge dq = \sum_{i,j} \frac{\partial p_i}{\partial q_j}(q) dq_j \wedge dq_i$ on $\mathbb{T}^n$ is identically 0. But this means that $\frac{\partial p_i}{\partial q_j}(q) = \frac{\partial p_j}{\partial q_i}(q)$ for all i, j. This implies that there exists a function $\sigma : \mathbb{R}^n \to \mathbb{R}$ such that for all i, $p_i(q) = \frac{\partial \sigma}{\partial q_i}(q)$. Hence there exist constants a_i (the *periods* of σ) and a function $s : \mathbb{T}^n \to \mathbb{R}$ such that for all i, $p_i(q) = a_i + \frac{\partial s}{\partial q_i}(q)$.

Corollary 8 *A Lagrangian graph V contained in the energy level $H^{-1}(h)$ of an autonomous Hamiltonian is of the form $\{(p, q), p = a + ds\}$, where s is a solution of the partial differential equation $H(a + ds(q), q) = h$.*

Definition 9 *The time-independant Hamilton-Jacobi equations associated to the Hamiltonian $H(p, q)$ are the equations of the form $H(ds(q), q) = h$. The time-dependant Hamilton-Jacobi equation associated to the Hamiltonian $H(p, q, t)$ is the equation $\frac{\partial S}{\partial t}(q, t) + H(\frac{\partial S}{\partial q}(q, t), q, t) = 0$. After identification of $K^{-1}(0)$ with $T^*\mathbb{T}^n \times \mathbb{R}$, it is nothing but the time-independant Hamilton-Jacobi equation $K(dS(q, t), q, t) = 0$, where K is defined by $K(p, E, q, \tau) = E + H(p, q, \tau)$.*

The modified Hamiltonian and Lagrangian. According to the above Corollary, each Lagrangian graph contained in an energy level of an autonomous Hamiltonian is the graph of the derivative of a solution of the Hamilton-Jacobi equation associated to a Hamiltonian

$$H_a(p, q) = H(a + p, q)$$

where $a \in (\mathbb{R}^n)^*$ should actually be thought of as a cohomology class in $\mathcal{H}^1(\mathbb{T}^n, \mathbb{R})$. Such a Hamiltonian is easily seen to be the Legendre transform of the Lagrangian

$$L_a(q, \dot{q}) = L(q, \dot{q}) - a \cdot \dot{q} = L(q, \dot{q}) - \sum_{i=1}^{n} a_i \dot{q}_i,$$

which satisfies the same hypotheses as the original one.

A remark which will play a fundamental role in the next section is that, *while the solutions of the Euler-Lagrange equations associated to L_a are independant of a, the minimizing ones do indeed depend on a.* The simplest example is the geodesic flow of the flat torus : adding a mass to better distinguish between the tangent and cotangent sides, let us take $L(q, \dot{q}) = \frac{m}{2}||\dot{q}||^2$. The Lagrangian L_a can be written

$$L_a(q, \dot{q}) = \frac{m}{2}||\dot{q}||^2 - a \cdot \dot{q} = \frac{m}{2}||\dot{q} - \frac{1}{m}a||^2 - \frac{||a||^2}{2m},$$

and the minimizers are immediately seen to be such that $\dot{q} = \frac{1}{m}a$, that is $p = a$. We have "controlled" (the word is from Kaloshin) the velocity (or momentum) of the minimizers.

2.4 Complete solutions and complete integrability

Let us suppose that to each $a \in (\mathbb{R}^n)^*$ we can associate in a differentiable way a solution u_a of the equation

$$H(a + du_a(q), q) = \alpha(a),$$

where α is a smooth function. Setting $S(a, q) = a \cdot q + u_a(q)$, this is equivalent to

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

Definition 10 *Let $f\left(\frac{\partial S}{\partial q}, q, S\right) = 0$ be a first order partial differential equation, in a domain of $\mathbb{R}^{2n+1}$ (coordinates p, q, z) where at least one of the partial derivatives $\frac{\partial f}{\partial p_i}(p, q, z)$ does not vanish. A* (local) complete solution *of the equation is a C^1 map $S(\alpha, q)$ form an open subset $\mathcal{O}$ of $\mathbb{R}^{2n}$ to $\mathbb{R}$, such that the map from $\mathcal{O}$ to $\mathbb{R}^{2n+1}$*

$$(\alpha, q) \mapsto \left(\frac{\partial S}{\partial q}(\alpha, q), q, S(\alpha, q)\right)$$

is a (local) paramerization of the hypersurface of $\mathbb{R}^{2n+1}$ whose equation is $f = 0$.

If $\frac{\partial^2 S}{\partial a \partial q}$ is invertible, the family $f((a, q, t) = S(a, q) - \alpha(a)t$ is called a *complete solution* of the time dependant Hamilton-Jacobi equation $\frac{\partial f}{\partial t} + H(\frac{\partial f}{\partial q}, q) = 0$.

The function S is of course not $\mathbb{Z}^n$-periodic in q, that is not defined on $\mathbb{T}^n$, but its derivative is. Hence S can be used as the *generating function* of the *symplectic transformation*

$$\Phi : \mathbb{R}^n \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n, \;\; \Phi(p,q) = (a,b),$$

defined by

$$p = \frac{\partial S}{\partial q}(a,q) = a + \frac{\partial u_a}{\partial q}(q), b = \frac{\partial S}{\partial a}(a,q) = q + \frac{\partial u_a}{\partial a}(q).$$

We have $dp \wedge dq + db \wedge da = d^2 S = 0$, hence $da \wedge db = dp \wedge dq$, which is the preservation of the canonical symplectic form. This implies that in the new coordinates (a,b), the Hamiltonian vector-field X_H^* becomes $X_{H\circ\Phi^{-1}}^*$, that is X_α^*. As α does not depend on the variables b, Hamilton's equations take the particularly simple *completely integrable* form

$$\frac{da_i}{dt} = 0, \;\; \frac{db_i}{dt} = \frac{d\alpha}{da_i}(a),$$

which is similar to the one defining the geodesic flow of the flat torus.

This is not astonishing. If for each a there exists a unique solution u_a which is differentiable, the collection of the graphs of these functions defines a foliation of the phase space by Lagrangian tori. The existence of such a foliation implies in turn the existence of *action-angle* coordinates in which the flows on the invariant tori are linear.

In the last two sections, we briefly describe what features of this integrable picture do persist when the Hamiltonian is perturbed. It turns out that dynamically defined invariant objects can be proved to exist for the perturbed Hamiltonian :
1) invariant Mather sets, hyperbolic in a weak sense, to which are attached semi-invariant sets made of pieces of their stable or unstable manifolds. These should be interpreted as torn remnants of invariant tori;
2) KAM invariant tori, which are the continuation of those invariant tori of the integrable flow which, not only are densely filled by any of their trajectory (and hence dynamically defined) but are so in a sufficiently uniform way.

3 Remnants of complete integrability (1) : weak KAM solutions

A non-zero energy surface of the completely integrable geodesic flows that we studied is filled, possibly up to a singular set of codimension one, by invariant Lagrangian tori. Moreover, the connected components of the complement of the singular set are symplectically diffeomorphic to $T^*\mathbb{T}^n$ by a diffeomorphism which transforms the tori into Lagrangian graphs (we did not prove that for the resonance zone but it is true). Finding invariant tori under X_H^* which are Lagrangian graphs in the energy level $H^{-1}(h)$ of an autonomous Hamiltonian H is the same as finding GLOBAL solutions of the Hamilton-Jacobi equations associated to the Hamiltonians H_a (recall that the invariant tori are the graphs of the derivative of such global solutions). Global solutions do not exist in general but the K.A.M. theory, which will be quickly described in section 4), asserts that a Cantor set of global solutions exists when H is a small C^k-perturbation of a completely integrable Hamiltonian (k not too small). On the other hand, the weak K.A.M. theorem of Fathi asserts that under the Tonelli convexity hypotheses and without any condition of proximity to a completely integrable Hamiltonian, global solutions u exist *in a weak sense*, that is as Lipshitz functions which are *viscosity* solutions of the equation in the sense of Lions, Papanicolaou, Varadhan (who had already proved the theorem in the case of the torus). In the non-regular case, the graphs of the derivatives of these weak solutions are semi-invariant sets made of pieces of stable (resp. unstable) manifolds of weakly hyperbolic invariant sets, the so-called *Mather sets*. They can be thought of as the (generally) broken remnants of the invariant tori of the integrable case.

Minimization of the action plays a central role in weak KAM theory. To give a flavor of its role, let us give another look at the trivial case of the flat torus : lifted to the universal cover (the flat plane), the geodesics become staight lines which form parallel families corresponding to Lagrangian submanifolds of the cotangent bundle. In addition to the local minimizing property which defines them, they do minimize the length (or the action) *globally* in $\mathbb{R}^2$, that is between any two of their points. In his classical 1932 work on the geodesic flow of an arbitrary metric on the 2-torus, G. Hedlund showed the existence in complete generality of a kind of "integrable skeleton", made of geodesics which have the same property of minimizing the length between any two points of their lift to the universal cover $\mathbb{R}^2$. For the torus of revolution, the "globally minimizing" geodesics are the ones deprived of caustics, that is the ones which do not belong to the resonance zone. In the general case, these geodesics form families which define invariant sets called to day *Aubry-Mather sets*, which generalize the families of parallel straight lines of the flat case. They are the first example of non trivial Mather sets. In the case of more degrees of freedom, the Mather sets are defined as the union of the supports of probability measures invariant under the Lagrange flow and minimizing the averaged action. An example already given by Hedlund himself shows that the step from minimizing trajectories to minimizing measures is a necessary one.

The basic technical tools are Tonelli and Weierstrass theories, which give respectively the existence and the regularity of curves which minimize the Lagrangian action under the fixed-ends condition. In the autonomous case, the weak KAM solutions can all be obtained by appropriately truncating the image of a Lagrangian manifold under the Hamiltonian flow when the time grows indefinitely. This makes the whole theory a vast generalization of the classical λ-*lemma.*

3.1 A minimizing property of the local solution of the Cauchy problem for the Hamilton-Jacobi equation

The geometric solution. We shall be at first interested in the time-dependant equation. Solving the Cauchy problem with a function $u : \mathbb{T}^n \to \mathbb{R}$ as initial data at a given time t_0, amounts to understanding the evolution of the (Lagrangian) graph of $du(q)$ under the flow of X_H^*. The solution is contained in the figure 12 which explains how singularities (*caustics*) do occur which prevent the existence of a global solution as a function but allow for the existence of a "multiform" solution. The graph of the derivative of this multiform solution is the Lagrangian submanifold of $K^{-1}(0) \equiv T^*T^n \times \mathbb{R} \subset T^*(T^n \times \mathbb{R})$ which is defined as the union of the images of the graph of $du(q)$ under the flow of Ξ_H^*.

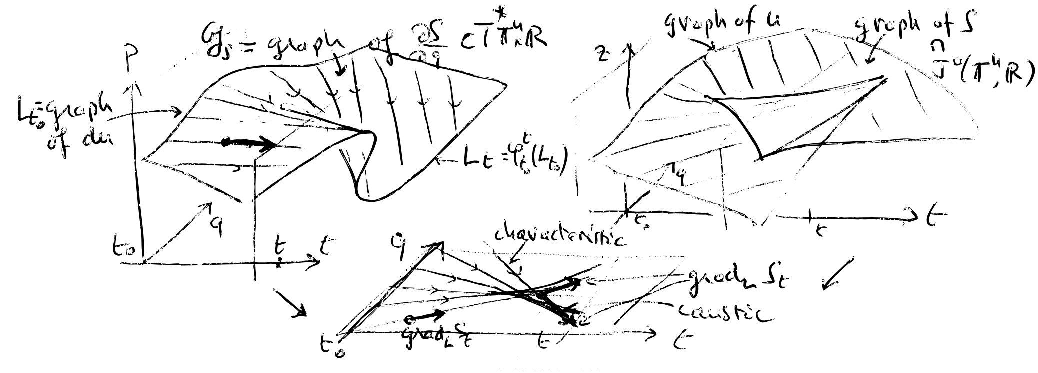

Figure 12

Characteristics. Let $S(q,t)$ be a geometrical (i.e. a priori *multivalued*) solution. The projections on space-time $\mathbb{T}^n \times \mathbb{R}$ of the integral curves of Ξ_H^* contained in the "graph" $\mathcal{G}_S \subset T^*\mathbb{T}^n \times \mathbb{R}$ of the space derivative $(q,t) \mapsto \frac{\partial S}{\partial q}(q,t)$ are called the *characteristics* associated to S. If we identify $T^*\mathbb{T}^n \times \mathbb{R}$ with $K^{-1}(0)$, where K is the extended Hamiltonian on $T^*(\mathbb{T}^n \times \mathbb{R})$, $\mathcal{G}_S$ becomes a Lagrangian submanifold of $T^*(\mathbb{T}^n \times \mathbb{R})$. The characteristics are the graphs of the solutions $t \mapsto q(t)$ of the *multivalued* differential equation

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}\left(\frac{\partial S}{\partial q}(q,t), q, t\right) := \mathrm{grad}_L S_t(q), \qquad (\mathcal{C})$$

which, by the Legendre diffeomorphism, is equivalent to $\frac{\partial S}{\partial q}(q,t) = \frac{\partial L}{\partial \dot q}(q, \frac{dq}{dt}, t)$. The multivaluedness is a reflection of the fact that beyond the caustic, several characteristics pass through a given point (q,t) of space-time.

The Weierstrass theory. We consider now a (*true, univalued*) solution $S(q,t)$, defined on a certain interval of time $[t_0, t_1]$, of the time-dependant Hamilton-Jacobi equation $\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial q}, q, t\right) = 0$. The differential equation $(\mathcal{C})$ is now univalued and the graphs of its integral curves, i.e. the characteristics associated to S, form a *field of extremals.* Let $c : [t_0, t_1] \to \mathbb{T}^n$ be a segment of extremal whose lift to $T^*\mathbb{T}^n \times \mathbb{R}$, $C^*(t) = \left(\frac{\partial L}{\partial \dot q}\big(c(t), \dot c(t), t\big), c(t), t\right)$, is contained in the Lagrangian graph $\mathcal{G}_S$ defined by the equation $p = \frac{\partial S}{\partial q}(q,t)$.

Proposition 11 *The segment of extremal c minimizes the action among absolutely continuous paths $\gamma : [t_0, t_1] \to \mathbb{T}^n$ with the same extremities, whose graph remains in the domain of definition of S.*

The idea of the proof is to lift to $\Gamma^*(t) = \left(\frac{\partial S}{\partial q}(\gamma(t), t), \gamma(t), t\right) \subset \mathcal{G}_S$ any curve $\gamma(t)$ that we want to compare to $c(t)$ (figure 13) and to use the fact that $\mathcal{G}_S$ is exact Lagrangian, to get

$$\mathcal{A}_L(c) = \int_{\Gamma^*} (p \cdot dq - H(p,q,t)dt).$$

Indeed, on $\mathcal{G}_S$, we have $p \cdot dq - H(p,q,t)dt = dS(q,t)$ because $p = \frac{\partial S}{\partial q}(q,t)$ and $H(p,q,t) = -\frac{\partial S}{\partial t}(q,t)$. This implies that

$$\mathcal{A}_L(c) = \int_{C^*} (p \cdot dq - H(p,q,t)dt) = \int_{t_0}^{t_1} d\big(S(q(t),t)\big) = \big[S\big(q(t),t\big)\big]_{t_0}^{t_1}$$

does not depend on the path on which one integrates as long as this path is contained in $\mathcal{G}_S$.
The Young-Fenchel inequality then implies that the difference of the actions

$$\mathcal{A}_L(c) - \mathcal{A}_L(\gamma) = \int_{t_0}^{t_1} \left[\pi(t) \cdot \dot\gamma(t) - H\big(\pi(t), \gamma(t), t\big) - L\big(\gamma(t), \dot\gamma(t), t\big)\right] dt,$$

where $\pi(t) = \frac{\partial S}{\partial q}\big(\gamma(t), t\big)$, is the integral of an everywhere ≤ 0 function.

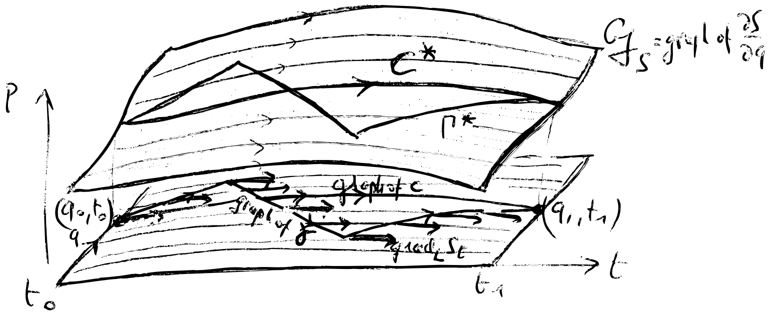

Figure 13 (in $T^*\mathbb{T}^n \times \mathbb{R} \equiv K^{-1}(0)$)

Corollary. The solution $S(q,t)$ of the time-dependant Hamilton-Jacobi equation defined on the (small enough) interval $[t_0, t]$ with initial condition $S(q, t_0) = u(q)$ is given by :

$$S(q,t) = \min_{\gamma, \gamma(t) = q} \left[u(\gamma(t_0)) + \int_{t_0}^{t} L\big(\gamma(s), \dot\gamma(s), s\big) ds\right],$$

where the min is taken over all absolutely continuous paths $\gamma : [t_0, t] \to \mathbb{T}^n$ such that $\gamma(t) = q$.

The proof is the same as for the Proposition : c is replaced by the unique extremal whose graph is the characteristic associated to S such that $c(t) = q$ and γ by a path defined on the interval $[t_0, t]$ and such that $\gamma(t) = q$. Then, as above,

$$\mathcal{A}_L(c) = \int_{C^*} (p \cdot dq - Hdt) = \int_{\Delta^*} (p \cdot dq - Hdt),$$

where Δ^* is composed of the lift to $\mathcal{G}_S$ of a path in $\mathbb{T}^n \times \{t_0\}$ joining $c(t_0)$ to $\gamma(t_0)$, followed by the lift Γ^* of γ (figure 14). One concludes because the part of $\mathcal{G}_S$ above $t = t_0$ coincides with the graph of du.

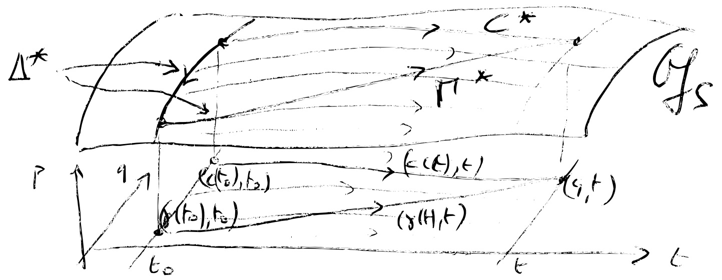

Figure 14

3.2 Global Lipschitz solutions of the Hamilton-Jacobi equation : the Lax-Oleinik semi-group

The Lax-Oleinik semi-group (autonomous case). The solution of the Cauchy problem defined above is in general only *local in time* : caustics appear as soon as the extremals in the corresponding field start intersecting each other. We now globalize it at the expense of regularity by taking in the global situation the *same* formula as in the local one : this amounts to cutting the *swallowtails* of the graph of the multiform function S. By keeping only for each (q, t) the lowest of the values of $S(q, t)$, one obtains the (discontinuous) graph of a Lipschitz solution of the Hamilton-Jacobi equation (figure 15) .

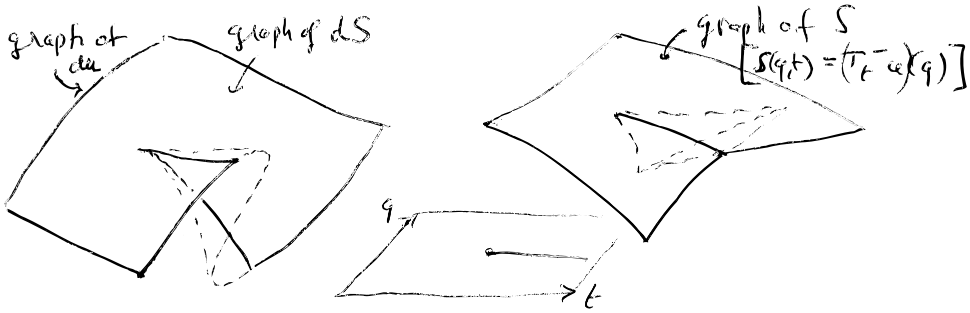

Figure 15

An astonishing feature of the result is that we get a global (weak) solution of the Cauchy problem *even when the initial condition $u(q)$ is only continuous.* If we approach u by C^1 functions u_n, the behaviour of the derivatives du_n may become wild as n tends to infinity but still the truncated global solutions corresponding to initial conditions u_n have a nice limit.

The complete statement *(in the autonomous case)* is the following :

Theorem 12 (Existence of the Lax-Oleinik semi-group) *1) The formula (for $t \geq 0$)*

$$\left(T_t^- u\right)(q) = \inf_\gamma \left[u\big(\gamma(s)\big) + \int_0^t L\big(\gamma(s), \dot\gamma(s)\big) ds \right],$$

where the inf *is taken over all absolutely continuous paths $\gamma : [0,t] \to \mathbb{T}^n$ such that $\gamma(t) = q$, defines a semi-group $\{T_t^-\}_{t\geq 0}$ of mappings from the space of continuous functions $C^0(\mathbb{T}^n, \mathbb{R})$ to itself;*
2) For all q, t, there exists a minimizing extremal $\gamma : [0,t] \to \mathbb{T}^n$ such that $\gamma(t) = q$ and

$$(T_t^- u)(q) = u(\gamma(0)) + \int_0^t L\big(\gamma(s), \dot\gamma(s)\big) ds; \qquad (*)_{u,q,t}$$

3)$||T_t^- u - T_t^- v||_0 \leq ||u - v||_0$;
4) $T_t^-(u + c) = (T_t^- u) + c$;
5) At each point where $S(q,t) = (T_t^- u)(q)$ has a derivative, it is a true solution of the time-dependant Hamilton-Jacobi equation;
6) The same is true for the semi-group $\{T_t^+\}_{t\geq 0}$ defined by

$$(T_t^+ u)(q) = \sup_\gamma \left[u\big(\gamma(0)\big) - \int_{-t}^0 L\big(\gamma(s), \dot\gamma(s)\big) ds \right],$$

where the sup *is taken over absolutely continuous paths $[-t, 0] \to \mathbb{T}^n$ such that $\gamma(-t) = q$.*

We refer to [C1] for a sketch of the proof and to [Fa] for the details. The main tools are
1) the existence, thanks to our convexity hypotheses, of minimizing extremals among absolutely continuous curves with given end-points (*Tonelli's theorem*);
2) the regularity of minimizing extremals, based on the minimizing property of small enough extremals and on their uniqueness (*Weierstrass local theory*);
3) an easy but fundamental *lemma of a priori compactness* asserting that there is an a priori bound on the velocities of a minimizing extremal, which depends only on the length of its interval of definition.

Differentiability of $T_t^- u$ and the unicity of characteristics. The point (q,t) is a point of differentiability of $S(q,t) = (T_t^- u)(q)$ if and only if it is an endpoint of a unique characteristic.
Figure 16 illustrates this :

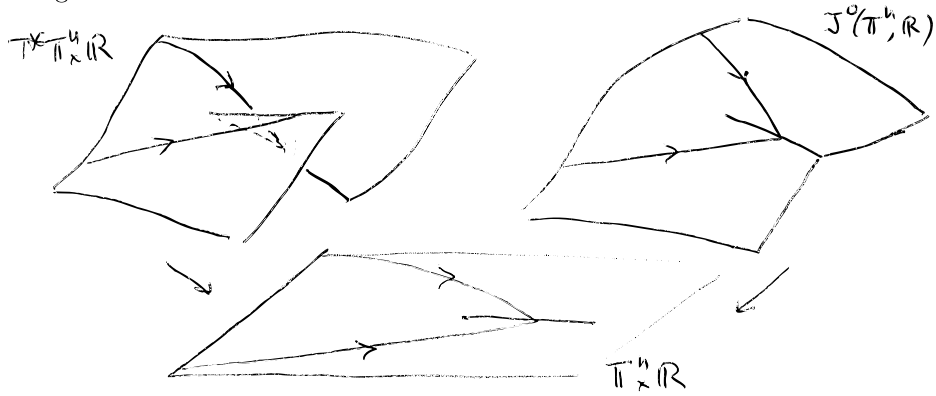

Figure 16

3.3 Fathi's weak KAM solutions and Aubry Mather-Mañé invariant sets as substitutes of invariant tori

Following Fathi, we deduce from the existence of the Lax-Oleinik semi-group the existence of weak KAM solutions (in fact viscosity solutions) of the *time-independant* Hamilton-Jacobi equation $H(du(q), q) = c$ for a well-chosen c (equal to $\alpha(0)$ in Mather's notation).

For this, one notices that the following properties are equivalent to each other:

$$(1)\ \exists c,\ H(du(q), q) = c$$

$$(2)\ \exists c,\ \frac{\partial S}{\partial t} + H\big(\frac{\partial S}{\partial q}, q\big) = 0, \text{ where } S(q,t) = u(q) - ct;$$

$$(3)\ u \text{ is a fixed point of the semi-group } u \mapsto T_t^- u + ct;$$

$$(4)\ u \text{ represents a fixed point of } T_t^- u \text{ in } C^0(\mathbb{T}^n, \mathbb{R})/\mathbb{R}.$$

One then proves the existence of a fixed point by a Leray-Shauder type fixed point argument. To conveniently state the theorem, we introduce the following definitions :

Domination. Given a real number c, we say that the function $u : \mathbb{T}^n \to \mathbb{R}$ is dominated by $L + c$ (and we write $u \prec L + c$) if for any interval $[a,b]$, $a \leq b$ and any absolutely continuous curve $\gamma : [a,b] \to \mathbb{T}^n$ we have

$$u\big(\gamma(b)\big) - u\big(\gamma(a)\big) \leq \int_a^b L\big(\gamma(s), \dot\gamma(s)\big)ds + c(b-a).$$

One can prove that $u \prec L + c$ if and only if u is locally Lipschitzian and $H(du(q), q) \leq c$ at each point q where u has a derivative.

Calibration. We suppose that $u \prec L + c$. The curve $\gamma : [a, b] \to \mathbb{T}^n$ is said to be (u, L, c)-calibrated if

$$u\big(\gamma(b)\big) - u\big(\gamma(a)\big) = \int_a^b L\big(\gamma(s), \dot{\gamma}(s)\big)ds + c(b - a).$$

Weak KAM theorem. *Let $L(q, \dot{q})$ be a time-independant Lagrangian on $T\mathbb{T}^n$ of class (at least) C^3, which satisfies the general convexity hypotheses. There exist Lipschitz functions $u_-, u_+ : \mathbb{T}^n \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ (the Mañé's critical energy) such that*
1) $u_-, u_+ \prec L + c$,
2) $\forall q \in \mathbb{T}^n$, there exists $\gamma_-^q :]-\infty, 0] \to \mathbb{T}^n$ and $\gamma_+^q :]0, +\infty] \to \mathbb{T}^n$ with $\gamma_-^q(0) = \gamma_+^q(0) = q$, such that, for all $t \geq 0$,

$$u_-(q) - u_-(\gamma_-^q(-t)) = ct + \int_{-t}^0 L\big(\gamma(s), \dot{\gamma}(s)\big)ds,$$

$$u_+(\gamma_+^q(t)) - u_+(q) = ct + \int_0^t L\big(\gamma(s), \dot{\gamma}(s)\big)ds,$$

3) $u_\pm$ satisfies $H(du_\pm(q), q) = c$ at each point q where it has a derivative.

Properties of weak KAM solutions. $u_\pm$ are not necessarily unique (the simplest example is the "two-fold covering" of the pendulum).
The following property is the analogue of the one that we discussed for $T_t^- u$: the discontinuities of the derivative of a weak KAM solution are the intersection points of at least two rays (characteristics).

Proposition 13 (differentiability and unicity of calibration) *A weak KAM solution u_- has a derivative at q if and only if there exists a UNIQUE (L, c, u_-)-calibrated path $\gamma_-^q :]-\infty, 0] \to \mathbb{T}^n$ such that $\gamma_-^q(0) = q$ (in other words, there is a unique characteristic which arrives at q).*

A consequence is that u_- is differentiable at every point of a characteristic except possibly at its extremity. This comes from the fact that, because they are necessarily regular, *two characteristics cannot meet except at a common end-point.*

Invariant measures and unicity of Mañé's critical energy. Let M_L be the set of all Borel probability measures μ on TM which are invariant under the flow of X_L. If $(q_s, \dot{q}_s) = \varphi_s(q_0, \dot{q}_0) \in T\mathbb{T}^n$ is an extremal, i.e. an integral curve of X_L, integrating the domination inequality against an invariant measure $\mu \in M_L$, we get

$$\int_{TT^n} \big(u_-(q_t) - u_-(q_0)\big)d\mu \leq \int_{TT^n} d\mu \left[\int_0^t L(\varphi_s(q_0, \dot{q}_0))ds + ct\right].$$

Dividing both sides by t and using the invariance of μ we get, when $t \to \infty$,

$$c = -\inf_{\mu \in M_L} \int L d\mu.$$

Minimizing measures can indeed be constructed which are supported in the α-limit sets of the minimizing extremals γ_-^q (figure 17) or the ω-limit sets of the γ_+^q : let t_n be a sequence of times tending to $+\infty$, μ_n be defined by

$$\mu_n(f) = \frac{1}{t_n} \int_{-t_n}^{0} f\big(\gamma_-^q(s), \dot{\gamma}_-^q(s)\big) ds$$

and μ be the weak limit of a subsequence of the μ_n's. One easily checks that $\int L d\mu = -c$.

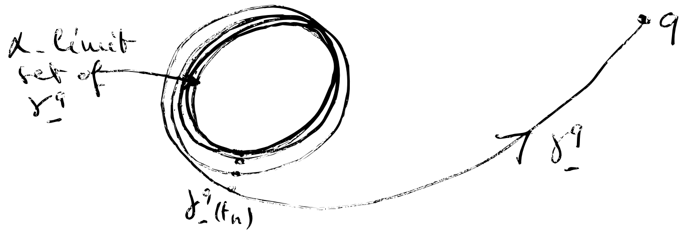

Figure 17

Mañé's critical energy and Hill's region for classical systems. When $L(q,\dot{q}) = \frac{1}{2}||\dot{q}||^2 - V(q)$, $c = \max_q V(q)$, i.e. the value of the energy under which the *Hill's region* (projection on $\mathbb{T}^n$ of an energy level) is not the whole configuration space $\mathbb{T}^n$.

The fundamental Lipschitz estimates. A well known property of solutions of the Hamilton-Jacobi equation is that as soon as they are of class C^1, their derivative is automatically Lipshitz. Classical examples are the so-called Buseman functions, that is the solutions of the time independant Hamilton-Jacobi equation associated to the geodesic flow of the flat metric on $\mathbb{R}^n$:

$$||\mathrm{grad}\, u|| = 1.$$

Such functions cannot be C^1 without being $C^{1,1}$. Another one is Birkhoff's theorem which asserts that a homologically non trivial continuous curve invariant under a "monotone twist map" of the annulus is necessarily a Lipschitz graph (see [Mo] and section 3.4). The following proposition gives a general statement of this kind (for a sketch of proof, see [C1] and for a complete one, see [Fa]) :

Proposition 14 *The following assertions are equivalent :*
1) $u : \mathbb{T}^n \to \mathbb{R}$ is of class C^1 and belongs to $\mathcal{S}_-$;
2) $u : \mathbb{T}^n \to \mathbb{R}$ is of class C^1 and belongs to $\mathcal{S}_+$;
3) $u : \mathbb{T}^n \to \mathbb{R}$ belongs to $\mathcal{S}_- \cap \mathcal{S}_+$;
4) $u : \mathbb{T}^n \to \mathbb{R}$ is of class C^1 and $\exists c \in \mathbb{R}$ such that $H(du(q), q) = c$.
IN ALL CASES, $du(q)$ IS LOCALLY LIPSCHITZ (i.e., u is not only C^1 but $C^{1,1}$).

The Mather set. The graph of the derivative of a weak KAM solution is semi-invariant under the flow of X_H. Mather's theory shows that it is, in a generalized sense, made of pieces of invariant manifolds of a "weakly hyperbolic" fully invariant set, the (image under the Legendre diffeomorphism of the) *Mather set* :

Definition 15 *The Mather set $\tilde{\mathcal{M}}_0 \subset T\mathbb{T}^n$ is the closure of the union of the supports of all invariant Borel probability measures which minimize $\int L d\mu$, that is such that $\int L d\mu = -c$, where c is the Mañé energy. $\tilde{\mathcal{M}}_0$ is invariant under the flow φ_t of X_L.*

Integrating the inequality $u \prec L + c$ against an invariant measure, one gets

Proposition 16 (universal calibration) *Let $(q, \dot{q}) \in \tilde{\mathcal{M}}_0$ and let $\gamma(t)$ be the extremal with initial conditions $(q, \dot{q})$, that is $\varphi_t(q, \dot{q}) = (\gamma(t), \dot{\gamma}(t)) \in \tilde{\mathcal{M}}_0$. Then, for any $u \prec L + c$ and any $t \leq t'$, $\gamma|_{[t,t']}$ is (L, c, u)-calibrated.*

As such a u exists (for instance a weak KAM solution), this proposition implies that the extremals contained in the projection $\mathcal{M}_0$ of $\tilde{\mathcal{M}}_0$ on $\mathbb{T}^n$ are minimizing.

Finally, we state without proof the generalization of the theorem of Birkhoff on Lipschitz graphs that we just recalled :

Theorem 17 (The structure theorem) *The Mather set $\tilde{\mathcal{M}}_0 \in T\mathbb{T}^n$ is a Lipschitz graph over its projection $\mathcal{M}_0$ in $\mathbb{T}^n$. Its image under the Legendre diffeomorphism $\Lambda : T\mathbb{T}^n \to T^*\mathbb{T}^n$ (which is well defined in the autonomous case) is contained in the critical energy level $H^{-1}(c)$.*

Conjugate weak KAM solutions. Weak KAM solutions are determined by their restriction to the Mather set $\mathcal{M}_0$. This allows to pair weak KAM solutions u_+, u_- so that the graphs of their (almost everywhere defined) derivatives play the role of (generalized) stable and unstable manifolds to $\tilde{\mathcal{M}}_0$.

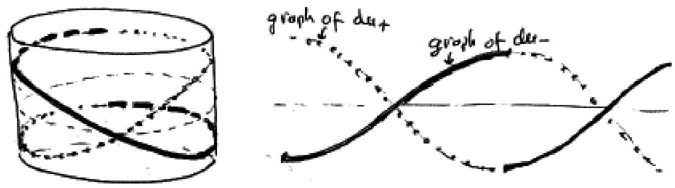

Figure 18 the trivial case of the pendulum

Convergence of the Lax-Oleinik semi-group in the autonomous case. This is a kind of a generalized λ-*lemma*, which states that (*in the autonomous case only*), for any $u \in C^0(M, \mathbb{R})$, the limits when $t \to \infty$ of $T_t^- + ct$ and $T_t^+ - ct$ exist and are weak KAM solutions u_- or u_+. The following picture illustrates this theorem in the simple case of a constant function u and the flow of the pendulum.

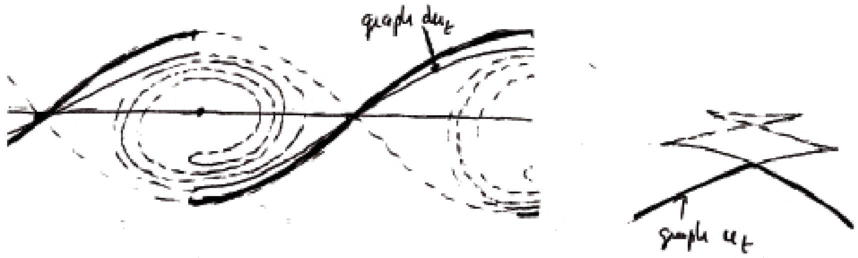

Figure 19 (convergence of the semi-group for the pendulum)

Mather's alpha function as an averaged Hamiltonian.
Using a control, that is replacing $L(q, \dot{q})$ by $L_a(q, \dot{q}) = L(q, \dot{q}) - a \cdot \dot{q}$, one defines a critical energy c_a and a Mather set $\mathcal{M}_a$. We noted already that, if replacing L by L_a does not change the Euler-Lagrange equations, it *does change the minimizers* of the action integral.

Definition 18 (Definition-Proposition) *Mather's alpha function is the function*

$$\alpha : \mathcal{H}^1(\mathbb{T}^n, \mathbb{R}) \equiv (\mathbb{R}^n)^* \text{ defined by } \alpha(a) = c_a.$$

It is convex and superlinear.

It can be checked that for any compact manifold M, α is naturally defined on the first cohomology group of M, that is : the Mañé energy $c(L - \varpi)$ depends only on the cohomology class of the closed 1-form ϖ.

In the case of the torus, we can interpret α as an averaged Hamiltonian in the following sense : let us pretend that to each $a \in (\mathbb{R}^n)^*$ we can associate in a differentiable way (even continuity is false because of non-unicity !) a weak KAM solution u_a. At each point q where u_a is differentiable, we have

$$H(a + du_a(q), q) = \alpha(a),$$

and, as we explained in section 2.4, the function $S(a, q) = a \cdot q + u_a(q)$ would be the *generating function* of a coordinate change leading to *action-angle coordinates* a, b which transforms the system into the particularly simple *completely integrable* form

$$\frac{da_i}{dt} = 0, \ \ \frac{db_i}{dt} = \frac{d\alpha}{da_i}(a).$$

This is not astonishing. If for each a there exists a unique weak KAM solution u_a which is differentiable, the collection of the graphs of these functions defines a foliation of the phase space by Lagrangian tori.

In general, this foliation is neither uniquely nor eveywhere defined but nevertheless, it can be thought of as a kind of (non uniquely defined) *integrable skeleton* made of KAM tori which are graphs (if they exist) and (non uniquely defined) pieces of stable and unstable manifolds of "weakly hyperbolic" Mather sets. In the case of the geodesic flow of a torus of revolution, this amounts to forgetting the whole (open) resonance zone. In the case of a monotone twist map (see section 3.4), we get the union of the invariant curves and (non uniquely defined) pieces of stable and unstable manifolds of the Aubry-Mather sets.

Mather's theory of minimal measures as a generalization of Hedlund's theory. The following theorem of Mather relates the Mather sets and the $\tilde{\mathbb{T}}^n$-minimizers, that is the extremals which, when lifted to the universal covering $\mathbb{R}^n$ of $\mathbb{T}^n$, minimise the action integral between any two of their points. They are the natural generalization to higher dimensions of Hedlund's *class A geodesics* on the 2-torus.

Theorem 19 *For any $a \in (\mathbb{R}^n)^*$, an extremal which is contained in $\mathcal{M}_a$ is a $\tilde{\mathbb{T}}^n$-minimizer.*

This does not mean that any vector ρ can be the rotation vector of a $\tilde{\mathbb{T}}^n$-minimizer. In fact, if Hedlund had shown that for any Riemannian metric on the 2-torus and any real number ρ, there exist class A geodesics which, in the universal covering $\mathbb{R}^2$, stay at a bounded distance of a straight line of slope ρ and hence have the rotation number ρ, he had also given an example of a Riemannian metric on $\mathbb{T}^3$ for which class A geodesics exist only for three rotation vectors. Indeed, to achieve a rotation vector, one needs in general to take averages on a set of extremals, not on a single one. This example was generalized by Bangert.

Gaps ? Mather sets can in general have gaps, which means that they do not cover the whole of $\mathbb{T}^n$. To get an idea of why this is so, one can notice that if we deform the flat metric by making a localized bump, $\tilde{\mathbb{T}}^n$-minimizers will avoid the bump to stay minimizing and this will create a gap. On the other hand, the KAM theorem, described in section 4, will tell us that such gaps do not exist if the perturbation is small enough and the Mather set in question is a perturbation of an invariant torus of the flat geodesic flow, which is filled by the solutions in a "sufficiently uniform way".

3.4 History: monotone distortions of the annulus

The *planar circular restricted 3-body problem* is the model of a non-integrable 2 degrees of freedom autonomous Hamiltonian system whose study, for well chosen values of the energy (the *Jacobi constant*) can be completely reduced to the study of the *Poincaré return map* of the flow in a *surface of section*, which is an annulus. The Hamiltonian character of the flow translates in the preservation of the area by the return map while the convexity of the Hamiltonian is translated into the *monotone twist property* of the map (see [Mo]). The equations are similar to those of the geodesic flow on a slightly deformed 2-sphere, a fact of which Poincaré was well aware. Such area preserving twist maps of the annulus were intensely studied by Birkhoff. In the eighties, Serge Aubry and John Mather independently discovered the existence of the so-called Aubry-Mather invariant sets which are action-minimizing and replace in some sense the invariant curves which exist in the integrable case (see [Mo, C2]). These invariant sets share with the invariant curves the property of being Lipschitz graphs of functions on the circle (for the invariant curves, this property was discovered by Birkhoff, see [H]). It was then pointed out by Jurgen Moser that Aubry and Mather's works had considerable overlap with Gustav Hedlund's paper in 1932 which studied arbitrary metric on $\mathbb{T}^2$. All the theories coincide when the geodesic flow of the torus admits a global surface of section, which is for instance the case when the metric is a small perturbation of the flat metric or more generally of a metric whose geodesic flow is integrable. Also, it was shown by Moser that a monotone twist map is always the Poincaré return map of the Euler-Lagrange flow of a periodic Lagrangian satisfying the general convexity hypotheses [Mo]. Finally, the weak KAM theory extends to such maps and puts both Hedlund's and Aubry-Mather's works in a common framework.

4 Remnants of complete integrability (2) : KAM tori

In this section, M is still supposed to be the n-torus. What is the structure of a typical Mather set ? The KAM theory gives a partial answer for Hamiltonians close enough in C^k topology (k big enough, say ≥ 4) to a completely integrable one. It asserts that if some non-degeneracy hypotheses are satisfied (and it is the case for the convex Hamiltonians that we are considering) there exists a big (in the sense of measure theory) set of "regular" Mather sets : these are invariant tori which are each the closure of any orbit they contain (i.e. which are "dynamically defined"). *In other words, in the convex case where Mather's theory applies, the KAM theorem can be viewed a posteriori as a regularity result about Mather sets.*

From the resolution above of the Cauchy problem for Hamilton-Jacobi equations, it is clear that looking in this purely geometric way for a solution which is at the same time global and regular is hopeless. Writing the equation in global coordinates $q \in \mathbb{R}^n$, one looks indeed for a geometric solution which is not only caustic-free (that is the graph of the derivative, up to a constant, of a univalued function) but also $\mathbb{Z}^n$-periodic in q. It is Kolmogorov who first understood that instead of solving the Cauchy problem one must specify a priori the dynamics on the Lagrangian manifold (torus) whose existence is seeked for.

4.1 Small denominators : the example of flows on the torus

For the completely integrable Hamiltonian $H(a,b) = \alpha(a)$, each torus $\mathcal{T}_{\hat{a}}$ of equation $a = \hat{a}$ (where $\hat{a} = (\hat{a}_1, \dots, \hat{a}_n)$ and the $\hat{a}_i$ are constants) is invariant under the Hamiltonian flow and the restriction to such a torus of the vector-field X_H^* is constant:

$$a = \hat{a}, \quad \frac{db_i}{dt} = \frac{\partial \alpha}{\partial a_i}(\hat{a}) = \hat{\omega}_i.$$

Exercise: each integral curve on $\mathcal{T}_{\hat{a}}$ is dense in $\mathcal{T}_{\hat{a}}$ if and only if the *frequencies* $\hat{\omega}_i$ are *non resonant*, in the sense that

$$\forall k = (k_1, \dots, k_n) \in \mathbb{Z}^n, \quad \sum_{i=1}^n k_i \hat{\omega}_i = 0 \quad \text{implies} \quad k = 0.$$

As Poincaré had already noticed, an invariant 2-torus filled in by a family of periodic trajectories is nothing but the collection of these trajectories, and is usually destroyed by a perturbation. The opening of a resonance that we noticed when replacing the flat torus by a torus of revolution can be turned into an example of this. What remains is a (usually) finite set of periodic trajectories, the invariant manifolds of the "hyperbolic ones" containing the graphs of the derivatives of weak KAM solutions. The higher dimensional situation is more complicated but the underlying intuition is the same. One could think that for a torus $\mathcal{T}_{\hat{a}}$ to resist to a small perturbation the density of the trajectories would be enough but it is not so. The original KAM theorem asserts that there is indeed persistance, but only for those tori $\mathcal{T}_{\hat{a}}$ which are filled in in a sufficiently uniform way, namely those whose frequencies $\hat{\omega}_i$ satisfy a *diophantine condition* $HD_{\gamma,\tau}$ of the form:

$$\forall k = (k_1, \dots, k_n) \in \mathbb{Z}^n \setminus 0, \quad |\sum_{i=1}^n k_i \hat{\omega}_i| \geq ||k||^{-\tau},$$

where $||.||$ is some norm on $\mathbb{R}^n$.

Let us give a simple, but paradigmatic, example where such diophantine conditions appear naturally (generalization to arbitrary dimension n is straightforward): suppose that we want to find a C^∞ function $f : \mathbb{T}^2 \to \mathbb{R}$ such that

$$\omega_1 \frac{\partial f}{\partial q_1} + \omega_2 \frac{\partial f}{\partial q_2} = g,$$

where $g : \mathbb{T}^2 \to \mathbb{R}$ is given (in other words, the Lie derivative $L_X f$ of f along the constant vector-field $X = (\omega_1, \omega_2)$ must be equal to g). A necessary condition is obviously that g be of zero mean on the torus. Moreover, the Fourier expansions $\sum_{(k_1,k_2)\in\mathbb{Z}^2} f_{k_1k_2} e^{k_1q_1+k_2q_2}$ and $\sum_{(k_1,k_2)\in\mathbb{Z}^2} g_{k_1k_2} e^{k_1q_1+k_2q_2}$ of f and g must satisfy the equations:

$$(k_1\omega_1 + k_2\omega_2) f_{k_1k_2} = g_{k_1k_2}, \qquad (\mathcal{E})$$

which even in the absence of resonance (i.e. when ω_1/ω_2 is irrational) leads to arbitrarily small denominators in the Fourier coefficients of f, a problem well known to astronomers since the eighteen century. In order to control the growth of these coefficients, henceforth the regularity of f, one must control the rational approximations of the ratio ω_1/ω_2. The following lemma (see [B]) is (or should be) a classical (but not completely obvious) exercise on Fourier series:

Lemma 20 *The frequency vector $\omega = (\omega_1, \omega_2)$ satisfies the diophantine condition $HD_{\gamma,\tau}$ if and only if equation $(\mathcal{E})$ has a C^∞ solution (unique up to the addition of a constant) f for any given C^∞ function g with zero mean.*

This fact is crucial in the proof of the following normal form theorem of Arnold and Moser: among the perturbations of a constant vector-field (ω_1, ω_2) on $\mathbb{T}^2$, are constant vector-fields with different frequencies (ω'_1, ω'_2), and in particular whith a different ratio $\omega'_1/\omega'_2 \neq \omega_1/\omega_2$ hence a completely different dynamics. This can be corrected trivially by adding a correction $(\Delta\omega_1, \Delta\omega_2)$ of the frequencies. In other words, a given constant vector-field is of codimension 2 in the set of all constant vector fields The normal form theorem of Arnold and Moser states that among all C^∞ vector-fields on $\mathbb{T}^2$ close enough to a constant vector-field whose frequencies satisfy $HD_{\gamma,\tau}$, the ones which are C^∞-conjugated to it form a submanifold of codimension 2. More precisely, the mapping

$$\Phi_\omega : Diff^\infty(\mathbb{T}^2, 0) \times \mathbb{R}^2 \to \mathcal{X}^\infty(\mathbb{T}^2)$$

defined by $\Phi_\omega(h, \lambda) = h_*\omega + \lambda$ is a C^∞ (i.e. "tame" in the sense of Hamilton) diffeomorphism of a neighborhood of $(Id, 0)$ onto a neighborhood of ω. In other words, a mere translation by λ close to zero in $\mathbb{R}^2$ is enough to transform any vector-field close enough to the constant vector field ω into one which is C^∞-conjugated to ω. The proof, which works for tori of any dimension n, uses a "hard" implicit function theorem, that is one valid in a scale of Fréchet spaces. The key feature of such theorems is the necessity of inverting (or inverting approximately) the differential of the mapping Φ_ω not only at the given point $(Id, 0)$ but in a whole neighborhood (invertibility is not an open property in Fréchet spaces). A simple example of this (serious) difficulty is the mapping $f \mapsto e^f$ from the space of C^0 real functions on $\mathbb{R}$ to itself. The derivative at $f = 0$ is the Identity but the image of the mapping, which is the set of positive functions, is not a neighborhood of the constant function 1 in the compact-open topology. This nasty phenomenon indeed disappears when one restricts the functions to a compact interval.

4.2 Herman's normal form theorem

Herman's normal form theorem follows exactly the same scheme as the Arnold-Moser theorem. Indeed, the Arnold-Moser theorem is but a special case of it. We study hamiltonians $H(r, \theta)$ on $T^*\mathbb{T}^n \equiv \mathbb{T}^n \times \mathbb{R}^n$. The role of the constant vector field of frequencies ω on the torus is now held by the set $\mathcal{N}_\omega$ of normal forms $N(r, \theta) = N_\omega(r) + O(r^2)$, where $N_\omega(r) = \omega \cdot r$. This is the set of Hamiltonians whose Hamiltonian vector-field leaves invariant the torus $r = 0$ and induces on it the constant vector-field with frequency vector ω. Let also $\mathcal{G}$ be some space (which we will not describe, see [Fe]) of Hamiltonian diffeomorphisms close to Identity, defined on a neighborhood of $\mathbb{T}^n \times \{0\}$ in $\mathbb{T}^n \times \mathbb{R}^n$. Let $C^\infty_+(\mathbb{T}^n \times \mathbb{R}^n)$ be the quotient of the space of Hamiltonians by the real constants.

Theorem 21 (Herman's normal form) *For every $\omega \in HD_{\gamma,\tau}$ and for every $N^o \in \mathcal{N}_\omega$, the map*

$$\begin{array}{rccl} \Phi_\omega : & \mathcal{N}_\omega \times \mathcal{G} \times \mathbb{R}^n & \to & C^\infty_+(\mathbb{T}^n \times \mathbb{R}^n) \\ & (N, G, \Delta\omega) & \mapsto & H = N \circ G + N_{\Delta\omega}, \end{array}$$

is a local C^∞-diffeomorphism in a neighborhood of $(N^o, id, 0)$. Moreover, the inverse map Φ_ω^{-1} depends smoothly in the sense of Whitney of $\omega \in HD_{\gamma,\tau}$.

As in the Arnold-Moser theorem, this theorem asserts that the set of Hamiltonians which are conjugated to a normal form with a diophantine frequency vector (i.e. those of the form $H = N \circ G$ with $N = N_\omega + O(r^2)$) form a submanifold of codimension n of the set of Hamiltonians modulo constants. Herman's theorem is in fact more general (see [Fe]) in that it works also with normal forms which leave invariant tori of dimension lower than n. For the proof, one needs as above inverting the $d\Phi_\omega$ on a whole neighborhood of $(N^0, Id, 0)$.

4.3 From Herman's normal form to Kolmogorov's theorem: the non-degeneracy as a source of parameters

The above normal form is certainly beautiful but a priori not very useful to understand the dynamics of X_H^*, ... unless in case $\Delta\omega = 0$. In order to achieve this, it is natural to seek for parameters which allow a control of the frequency correction $\Delta\omega$. In the non-degenerate case studied by Kolmogorov, the control parameters are the actions.

In order to get some understanding of how it can be done, let us apply the theorem to the trivial case of a completely integrable hamiltonian in action-angle variables, i.e. a hamiltonian $H^0(r)$ depending only on the action variables $r = (r_1, \dots, r_n)$. To H^0, we associate the n-parameter family of normal forms

$$N^s(r) = H^0(s + r) - H^0(s).$$

(Recall that we focus only on the invariant torus $r = 0$.) It follows from Taylor formula that

$$N^s \in \mathcal{N}_{\omega_s^0}, \quad \text{where} \quad \omega_s^0 = \frac{\partial H^0}{\partial r}(s).$$

Now, for a given s, and a given frequency vector ω close to ω_s^0, we can write in a unique way

$$N^s = N + N_{\Delta\omega}, \quad \text{with} \quad N \in \mathcal{N}_\omega \quad \text{and} \quad \Delta\omega = \omega_s^0 - \omega,$$

that is $N^s = \Phi_\omega(N, Id, \omega_s^0 - \omega)$, which relates N^s to the normal forms of frequency vector $\omega = \omega_s^0 + \delta\omega$ instead of ω_s^0.

In other words, *in the completely integrable case, we can trade the "a posteriori" frequency correction $\Delta\omega$ for an "a priori" frequency correction $\delta\omega = -\Delta\omega$.*

Let us consider now a Hamiltonian $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ defining what Poincaré, at the beginning of the *New Methods of Celestial Mechanics*, called "the general problem of dynamics", that is

$$H(r, \theta) = H^0(r) + H^1(\theta, r),$$

where H^0, depending only on the action variables $r = (r_1, \ldots, r_n)$, is completely integrable and H^1 is a small perturbation (in a well chosen topology) which depends on both angle and action variables $(\theta, r) = (\theta_1, \ldots, \theta_n, r_1, \ldots, r_n)$. To H we can also associate the n-parameter family of Hamiltonians

$$H^s(\theta, r) = H(\theta, s + r) - H^0(s) = N^s(r) + H^1(\theta, r).$$

Let us first pretend that Herman's normal form theorem is true for *any* frequency vector ω. This is not true of course but, thanks to the Whitney-smooth dependance of $\Phi_\omega^{-1}(H^s) = (N, G, \Delta\omega)$ in ω, the Whitney extension theorem allows us to extend to all ω (in a smooth but highly non unique way) the mapping $(s, \omega) \mapsto \Delta\omega$, where $\Delta\omega$ is the projection on $\mathbb{R}^n$ of $\Phi_\omega^{-1}(H^s)$. But H^s is a small perturbation of N^s for which $\Delta\omega = \omega_s^0 - \omega$, whose derivative with respect to ω is $-Id$; one then deduces from the usual implicit function theorem in $\mathbb{R}^n$ that the equation $\Delta\omega = 0$ defines the graph of a mapping $s \mapsto \bar{\omega}_s$ close to the mapping $s \mapsto \omega_s^0$.

The problem is of course that, for a given s, $\bar{\omega}_s$ has no reason to be a diophantine frequency vector, which seems to deprive our result of any meaning. But s is a parameter that we can choose. If we suppose that H^0 is *Kolmogorov non-degenerate* in the neighborhood of $r = \hat{s}$, that is if

$$\frac{\partial^2 H^0}{\partial r^2}(\hat{s}) \quad \text{is invertible,} \qquad (N.D.)$$

it follows that the frequency map $s \mapsto \omega_s^0$ and also $s \mapsto \bar{\omega}_s$ covers a full neighborhood of the frequency vector $\omega_{\hat{s}}^0$. Hence, we have a set of large relative Lebesque measure of values of s close to $\hat{s}$ such that $\bar{\omega}_s$ is indeed diophantine and Herman's theorem applies (figure 20).

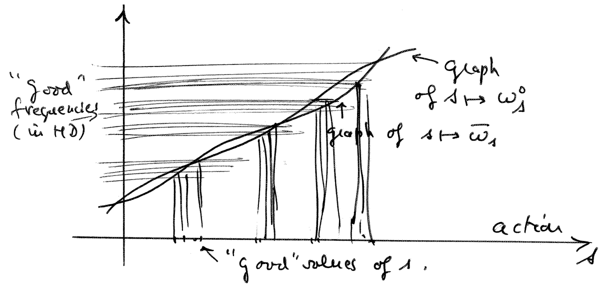

Figure 20

This implies the famous

Theorem 22 (Kolmogorov theorem) *Let $H(\theta, r) = H^0(r) + H^1(\theta, r)$ be a small perturbation (in a smooth enough topology) of the completely integrable Hamiltonian $H^0(r)$. If H^0 is non-degenerate in the neighborhood of $r = \hat{s}$, that is if $\frac{\partial^2 H^0}{\partial r^2}(\hat{s})$ is invertible, the hamiltonian vector-field X_H^* leaves invariant, close to $r = \hat{s}$, a set of large relative Lebesgue measure of tori, on each of which it is conjugate to a diophantine constant vector-field.*

Using the normal form theorem, the non-degeneracy hypothesis in Kolmogorov's theorem can be greatly weakened in the analytic case. Indeed, thanks to the following beautiful result, it is enough to suppose that the image of the mapping $s \mapsto \omega_s^0$ does not lie in a proper vector subspace of $\mathbb{R}^n$:

Theorem 23 (Arnold, Margulis, Pyartli) *If some real-analytic map $s \mapsto \omega_s^o$ from a domain of $\mathbb{R}^p$ to $\mathbb{R}^n$ is non-planar in the sense that its image is nowhere locally contained in some proper vector space of $\mathbb{R}^n$, the Lebesgue measure of $\{s,\ \omega_s^o \in HD_{\gamma,\tau}\}$ is positive provided that γ is small enough and τ large enough.*

4.4 The problem of Arnold diffusion

I just mention this important problem to which much research activity was devoted in recent years. This problem appears only for systems with three and more degrees of freedom. For one degree of freedom, a Lagrangian invariant torus, which has dimension n, is a component of an energy level curve. For two degrees of freedom, an energy manifold has dimension three and it is seprated into two regions by a torus of dimension two. For three degrees of freedom, a 3-torus no more separates a five dimensional energy manifold and this leaves place a priori for a leaking (diffusion) of trajectories through the (Cantor) set of invariant tori given for example by Kolmogorov theorem. The theorem of Nekhoroshev, a refined version of classical estimates on divergent series, says that if such a diffusion occurs, it must occur very (exponentially) slowly for systems close to integrable. Arnold was the first to build an ad hoc perturbation of an integrable hamiltonian for which this phenomenon does occur. More recently, a genericity result for three degrees of freedom hamiltonians was announced by John Mather.

References

[A] V.I. Arnold, *Méthodes mathématiques pour la mécanique classique*, Editions Mir 1976

[B] J.B. Bost *Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnol'd, Moser, Rüssmann, Zehnder, Herman, Pöschel,...)*, (séminaire Bourbaki 639) Astérisque 133-134, p. 113-157, Société Mathématique de France (1986)

[C1] A. Chenciner, *Calculus of variations in the convex case: an introduction to Fathi's weak KAM theorem and Mather's theory of minimal invariant measures*, 5 lectures at the Universitat Politècnica de Catalunya, Barcelona (july 2004). Notes available at http://www.imcce.fr/Equipes/ASD/person/chenciner/chen_preprint.html

[C2] A. Chenciner, *La dynamique au voisinage d'un point fixe ellptique conservatif : de Poincaré et Birkhoff à Aubry et Mather* (séminaire Bourbaki 622), Astérisque 121-123, p. 147-170, Société Mathématique de France (1985)

[Fa] A. Fathi, *Weak KAM theorem in Lagrangian dynamics*, book to appear

[Fe] J. Féjoz, *Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman)*, Michael Herman Memorial Issue, *Ergodic Theory and Dynamical Systems* **24:5** (2004) 1521–1582. Updated version at http://www.institut.math.jussieu.fr/~fejoz/.

[H] M. Herman, *Sur les courbes invariantes par les difféomorphismes de l'anneau*, volume 1, Astérisque 103-104, Société Mathématique de France (1983)

[L] M. Levi, *Lectures on Geometrical Methods in Mechanics*, in "Classical and Celestial Mechanics, the Recife lectures", H. Cabral and F. Diacu ed., Princeton University Press 2002

[Mo] J. Moser, *Selected Chapters in the Calculus of Variations*, Birkhauser (2003)