A SHORT INTRODUCTION TO KAM THEORY

JACQUES FÉJOZ

Contents

KAM theory is the perturbative theory, initiated by Kolmogorov, Arnold and Moser in the 1950's, of quasiperiodic motions in conservative dynamical systems. These notes are a short introduction to the subject.

References of particular value are the book [4] on Hamiltonian systems, the papers [29, 32] on KAM theory, and the book [5] for applications in celestial mechanics. More detailed accounts with various viewpoints can be found in [1, 7, 8, 11, 12, 14, 16, 27, 28, 31, 33] and references therein.

1. Hamiltonian systems

Let H be a smooth function on an open set M of $\mathbb{T}^n \times \mathbb{R}^n = \{(\theta, r)\}$, with $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. The *Hamiltonian vector field* of H is

$$X_H : \begin{cases} \dot\theta_j = \partial_{r_j} H \\ \dot r_j = -\partial_{\theta_j} H, \quad j = 1, ..., n. \end{cases}$$

Date: February 2011 (first version) and July 2013 (second version).

M is called the *phase space* for reasons coming from thermodynamics, which almost everybody has forgotten. Hamiltonian vector fields have been introduced and studied by Lagrange [25], Cauchy [9] and Hamilton [21, 22].

The projection of X_H on each plane of conjugate coordinates (θ_j, r_j) is orthogonal to the projection $(\partial_{\theta_j} H, \partial_{r_j} H)$ of the Euclidean gradient ∇H. While the Euclidean gradient points towards the direction of steepest ascent of H, X_H is tangent to the energy levels of H, or, equivalently, H is a first integral of X_H:

$$H' \cdot X_H = \frac{\partial H}{\partial \theta}\dot{\theta} + \frac{\partial H}{\partial r}\dot{r} = 0.$$

Example 1 (Pendulum) The Hamiltonian equations of

$$H(\theta, r) = \frac{r^2}{2} - \cos\theta$$

are equivalent to the classical equation $\ddot{\theta} = -\sin\theta$ of a frictionless pendulum, as given by the theorem of the angular momentum. Jacobi introduced the transcendant elliptic functions in order to solve those equations. However, since the integrale curves of X_H are contained in level sets of H, one recovers the behavior of the pendulum (up to the time-parametrization) by an algebraic computation. Of course, in higher dimension the conservation of energy is not sufficiant anymore to find the integral curves.

Example 2 (Particle in a potential) Consider a particle of position x moving in a force potential $U(x)$ in $\mathbb{R}^3$. Newton's equation

$$\ddot{x} = -\nabla U(x)$$

is equivalent to Hamilton's equations associated with the Hamiltonian

$$H(x, \dot{x}) = \frac{\dot{x}}{2} + U(x)$$

(here the phase space is rather $\mathbb{R}^3 \times \mathbb{R}^3$).

Example 3 (Hamilton-Jacobi equations) Consider a general first-order scalar partial differential equation, i.e. a relation

$$H(\theta, u'(\theta), u(\theta)) = 0,$$

where θ is the space variable (in $\mathbb{T}^n$ as well as in any other manifold of dimension n, say N), u is the unknown function, u' is the derivative of u, and H is a function over $M = T^*V \times \mathbb{R} = \{(\theta, r, u)\}$. Let V be some submanifold of M of dimension $n-1$. The theory of characteristics (see [3]) tells us that the 1-graph of a local solution can be obtained by flowing V along the integral curves of the vector field

$$\begin{cases} \dot{\theta} = \partial_r H \\ \dot{r} = -\partial_\theta H - r\partial_u H \\ \dot{u} = r \cdot \partial_r H, \end{cases}$$

provided that V satisfies some adequate transversality property. An important case is when H does not depend on the value u of the unknown, i.e. H is defined T^*V. Then the above vector field descends to the Hamiltonian vector field of H. So, solving the above PDE reduces locally to integrating X_H.

On the other hand, we will see that KAM theory yields (very particular) solutions to the Hamilton-Jacobi equation associated with a Hamiltonian.

Failing to have other first integrals than the Hamiltonian itself, generic Hamiltonian systems have some more subtle invariants. For example, they conserve the volume in phase space:

$$\operatorname{div} X_H = \sum_j \left(\frac{\partial \dot{\theta}_j}{\partial \theta_j} + \frac{\partial \dot{r}_j}{\partial r_j} \right) = \sum_j \left(\frac{\partial^2 H}{\partial \theta_j \partial r_j} - \frac{\partial^2 H}{\partial r_j \partial \theta_j} \right) = 0$$

(the volume can actually be seen as a first integral of the *variational equation* associated with X_H). We see that not only is the divergence equal zero, but each of the n terms separately are equal to zero. This is the sign that Hamiltonian vector fields have a stronger invariance property —namely they preserve a "symplectic form"—, which we will describe later.

2. Quasiperiodic motions

An important and simple class of Hamiltonians is that of *integrable* Hamiltonians, which do not depend on the angle θ. In such cases, the vector field becomes

$$\dot{\theta} = \frac{\partial H}{\partial r}(r) \equiv cst, \quad \dot{r} = 0,$$

and the flow

$$\varphi_t(\theta, r) = \left(\theta + t \frac{\partial H}{\partial r}(r), r \right).$$

The phase space is foliated in invariant tori $r = cst$, in restriction to which the flow is *quasiperiodic* (=linear), of *frequency vector* $\frac{\partial H}{\partial r}(r)$.

A vector r being fixed, let $\alpha := \frac{\partial H}{\partial r}(r) \in \mathbb{R}^n$ and consider the flow

$$\varphi_t : \mathbb{T}^n \to \mathbb{T}^n, \quad \theta \mapsto \theta + t\alpha.$$

Lemma 1. *The frequency vector α is a topological conjugacy invariant up to the action of the discrete group $GL_n(\mathbb{Z})$: if two linear flows $\theta + t\alpha$ and $\theta + t\beta$, with $\alpha, \beta \in \mathbb{R}^n$, are topologically conjugate, there exists $A \in GL_n(\mathbb{Z})$ such that $\beta = A\alpha$ (and, if the conjugacy preserves the orientation, $A \in SL_n(\mathbb{Z})$).*

Proof. Assume two linear flows $\theta + t\alpha$ and $\theta + t\beta$, with $\alpha, \beta \in \mathbb{R}^n$, are topologically conjugate: there exists a homeomorphism h of $\mathbb{T}^n$ such that $h(\theta + t\alpha) = h(\theta) + t\beta$. At the expense of substituting $h(\theta) - h(0)$ for $h(\theta)$, we may assume that $h(0) = 0$.

Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be the unique lift of h such that $H(0) = 0$. Now, the equality $H(\theta + t\alpha) = H(\theta) + t\beta$ holds for $\theta = t = 0$ and, by continuity, for $\theta \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Moreover, there exists a matrix $A \in GL_n(\mathbb{Z})$ such that $H(\theta + k) = H(\theta) + Ak$ for all $\theta \in \mathbb{R}^n$ and $k \in \mathbb{Z}^n$; A is invertible because H is. Hence $V := A^{-1}H - \mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is a $\mathbb{Z}^n$-periodic vector field. In terms of V, the conjugacy hypothesis at $\theta = 0$ asserts that

$$L(t\alpha + V(t\alpha)) = LV(0) + t\beta \quad (\forall t \in \mathbb{R}),$$

i.e.

$$L(V(t\alpha) - V(0)) = t(\beta - L\alpha).$$

Since the left hand side is bounded, necessarily $\beta = L\alpha$. □

The action of $GL_n(\mathbb{Z})$ is closely related to the arithmetic properties of frequency vectors ; see [34, 2.2.3] for $n = 2$.

Proposition 1. *The following properties are equivalent :*

1. *The vector α is* non resonant*: $k \cdot \alpha \neq 0$ for all $k \in \mathbb{Z}^n \setminus \{0\}$*
2. *The flow (φ_t) of the constant vector field α is ergodic: invariant continuous functions ($f(\theta + t\alpha) \equiv f(\theta)$ for all $t \in \mathbb{R}$ and $\theta \in \mathbb{T}^n$) are constant*
3. *For every continuous function f on $\mathbb{T}^n$, the time average of f exists, is constant and equals the space average of f:*
$$\lim_{T\to+\infty} \frac{1}{T} \int_0^T f(\theta + t\alpha)\, dt = \int_{\mathbb{T}^n} f(\theta)\, d\theta.$$
4. *Every trajectory of (φ_t) is dense on $\mathbb{T}^n$.*

More general classes of functions than continuous ones can be considered, but we lazily stick here to the most convenient setting for our purpose. See [6, 20, 23] for further results on ergodicity.

Proof. (1) $\Rightarrow$ (2) Suppose that α is non resonant and let $f \in C^0(\mathbb{T}^1)$ be invariant: $f = f \circ \varphi_t$ for all t. The k-th Fourier coefficient of $f \circ \varphi_t$ is

$$\widehat{f \circ \varphi_t}(k) = \int_{\mathbb{T}^n} e^{-i2\pi k\cdot\theta} f(\theta + t\alpha)\, d\theta.$$

The change of variable $\theta' = \theta + t\alpha$ shows that

$$\widehat{f \circ \varphi_t}(k) = e^{i2\pi k\cdot\alpha t} \hat{f}(k).$$

By uniqueness, for all $k \in \mathbb{Z}^n \setminus \{0\}$ we see that $\hat{f}(k) = 0$. Hence f is constant.

(2) $\Rightarrow$ (1) Conversely, suppose that $k \cdot \alpha = 0$ for some $k \in \mathbb{Z}^n \setminus \{0\}$. Then $f(\theta) = e^{i2\pi k\cdot\theta}$ is invariant and not constant, hence the flow is not ergodic.

(1) $\Rightarrow$ (3) Call $\bar{f}$ the space-average of f. We will show the conclusion by taking more and more general functions.

– If f is constant, $\bar{f}(\theta) \equiv \bar{f}$ trivially. If $f(\theta) = e^{i2\pi k\cdot\theta}$ for some $k \in \mathbb{Z}^n \setminus \{0\}$, direct integration shows that

$$\frac{1}{T} \int_0^T f(\theta + t\alpha)\, d\theta = \frac{1}{T} e^{i2\pi k\cdot\theta} \frac{e^{i2\pi k\cdot\alpha T} - 1}{ik \cdot \alpha} \to_{T\to+\infty} 0 = \bar{f}.$$

The expression $k \cdot \alpha$ in the denominator is the first occurence of the so-called *small denominators*, which are the source of many difficulties in perturbation theory.

– If f is a trigonometric polynomial, the same conclusion holds by linearity.

– Let now f be continuous. Let $\epsilon > 0$. By the theorem of Weierstrass, there is a trigonometric polynomial P such that

$$\max_{\theta\in\mathbb{T}^n} |f(\theta) - P(\theta)| \leq \epsilon.$$

For such a P, we have shown that there is a time T_0 such that if $T \geq T_0$,

$$\left|\frac{1}{T}\int_0^T P(\theta + t\alpha)\, d\theta - \bar{P}\right| \leq \epsilon.$$

Using the two latter inequalities, we see that

$$\begin{aligned}\left|\frac{1}{T}\int_0^T f(\theta + t\alpha)\, dt - \bar{f}\right| \\ \leq \frac{1}{T}\int_0^T |f(\theta + t\alpha) - P(\theta + t\alpha)|\, dt + \left|\frac{1}{T}\int_0^T P(\theta + t\alpha)\, dt - \bar{P}\right| + |\bar{P} - \bar{f}| \leq 3\epsilon.\end{aligned}$$

So, again $\frac{1}{T}\int_0^T f(\theta + t\alpha)\, d\theta$ tends to 0.

(3) $\Rightarrow$ (1) Suppose α is resonant: $k \cdot \alpha = 0$ for some $k \in \mathbb{Z}^n \setminus \{0\}$, and let $f(\theta) = e^{i2\pi k\cdot\theta}$. The space average of f equals 0, while

$$\frac{1}{T}\int_0^T e^{i2\pi k\cdot(\theta+\alpha t)}\, dt = e^{i2\pi k\cdot\theta}.$$

So there exists a non constant continuous function whose time and space averages do not match.

(1) $\Rightarrow$ (4) Suppose that one trajectory is not dense: there exist a point $\theta \in \mathbb{T}^n$ and an open ball $B \subset \mathbb{T}^n$ such that the curve $t \mapsto \theta + t\alpha$ will never visit B. Let f be a continuous function whose support lies inside B and whose integral is > 0. The space average of f is > 0, while its time average is 0. Hence α is resonant.

(4) $\Rightarrow$ (1) Suppose α is resonant: $k \cdot \alpha = 0$ for some $k \in \mathbb{Z}^n \setminus \{0\}$. We will show that there is a small ball B centered at $\theta^o := k/2 \pmod{\mathbb{Z}^n}$ which the trajectory $t \mapsto t\alpha$ never visits. Indeed, let θ be in such a ball B of small radius. Does there exist $t \in \mathbb{R}$ such that $t\alpha = \theta$ in $\mathbb{T}^n$? Equivalently, does there exist $t \in \mathbb{R}$ and $\ell \in \mathbb{Z}^n$ such that $\alpha = \theta + \ell$? Taking the dot product of the equation with k yields $0 = k \cdot \theta + k \cdot \ell$. But $k \cdot \ell \in \mathbb{Z}$, while $k \cdot \theta \in]0, 1[$ provided the radius of B is small enough (depending on k). This shows that there is no such $t \in \mathbb{R}$. □

If we think for instance to two planets revloving around the Sun with frequencies α_1 and α_2, that the frequency vector $\alpha = (\alpha_1, \alpha_2)$ be resonant means that the two planets will regularly find themselves in the same relative position. Hence, their mutual attraction, which is small due to their small masses compared to the mass of the Sun, instead of averaging out, will pile up. This is all the more true that the order $|k| := |k_1| + \cdots + |k_n|$ of the resonance is small. As a general rule, perturbation theory rather studies what happens away from resonances, and at some distance away from them in the phase space (all the farther that they have low order).

3. A more geometric viewpoint

In the study of differential equations, Poincaré has shifted the interest of mathematicians from particular solutions to geometric properties of the phase flow, considering all solutions simultaneously. Technically, the latter strategy often consists in computing *normal forms*, i.e. simple expressions of the vector field in well chosen coordinates.

One of the primary interests of the Hamiltonian formalism is that all the information on a Hamiltonian vector field is contained in a function. It is easier to compute changes of coordinates for functions than for vector fields. But in order to preserve the simple relation between the Hamiltonian function and its vector field, only some special changes of coordinates should be used, namely those diffeomorphisms $\phi : M \to M$ such that the direct image by ϕ of the Hamiltonian vector field of $H \circ \phi$ equals the Hamiltonian vector field of H:

$$\phi_* X_{H\circ\phi} = X_H.$$

In dimension 2, we have seen that X_H preserves the area. So, certainly ϕ should preserve the area form $\omega = d\theta \wedge dr$.

Let us introduce a coordinate-free definition of X_H. Let

$$\omega = \sum_{1\le j\le n} d\theta_j \wedge dr_j.$$

This geometric structure is called the *symplectic form* of the phase space M. It is the field of 2-forms (antisymmetric bilinear forms) which maps two velocities $(\dot\theta, \dot r)$ and $(\dot\Theta, \dot R)$ (tangent vectors of M at some point (θ, r)) to

$$\omega((\dot\theta, \dot r), (\dot\Theta, \dot R)) = \sum_{1\le j\le n} \det \begin{pmatrix} \dot\theta_j & \dot\Theta_j \\ \dot r_j & \dot R_j \end{pmatrix},$$

i.e. to the sum of the oriented areas of the projections on planes of conjugate coordinates (θ_j, r_j), of the parallelogram generated by the two velocity vectors. An excellent and straightforward introduction to differential forms can be found in Arnold's book [4].

If $X = (\dot\theta, \dot r)$ is a vector field,

$$\omega(X, \cdot) = \sum_{1\le j\le n} (\dot\theta_j \, dr_j - \dot r_j \, d\theta_j),$$

so the Hamiltonian vector field can be defined by the following equation.

Lemma 2. *The Hamiltonian vector field of H is characterized by the implicit equation $\omega(X_H, \cdot) = dH$.*

Hence the only eligible transformations ϕ are be the ones which preserve ω, in the sense that

$$\omega = \phi^*\omega,$$

where $\phi^*\omega(X, Y) := \omega(\phi' \cdot X, \phi' \cdot Y)$ for all pairs of tangent vectors X and Y at a point. Such transformations are called *symplectic* or *canonical*.

A fundamental operation on differential forms is the *exterior derivative*. It extends the usual differential of functions to differential forms of any degree p:

$$d \sum_{i_1<\cdots<p} \rho_{i_1,\ldots,i_p}(\theta) d\theta_{i_1} \wedge \cdots \wedge d\theta_{i_p} = \sum_{i_1<\cdots<p} d\rho_{i_1,\ldots,i_p}(\theta) \wedge d\theta_{i_1} \wedge \cdots \wedge d\theta_{i_p}.$$

It can be defined intrinsically (and implicitely) by the Stokes formula

$$\int_V d\rho = \int_{\partial V} \rho,$$

where V is an oriented manifold with boundary of dimension $\deg \rho + 1$, and ∂ is the boundary oparator. That $\partial^2 = \emptyset$, entails that d is a cohomology operator: $d^2 = 0$. Again, see [4] for a self-contained introduction to differential forms.

Remark 4 Using the exterior derivative, Maxwell's first two equations of electromagnetism boil down to

$$dF = 0,$$

where F is the electro-magnetic 2-form in the 4-dimensional space-time [19].

Example 5 Let $\rho = \sum_{1\le i\le n} \rho_i(\theta)\, d\theta_i$ be a closed 1-form on $\mathbb{T}^n$, *closed* meaning $d\rho = 0$. (If $n = 3$ and ρ is identified to a vector field, $d\rho$ is an intrinsic version of the curl of ρ.) The diffeomorphism

$$\phi : (\theta, r) \mapsto (\theta, r + \rho(\theta))$$

satisfies

$$\phi^*\omega - \omega = \sum_{1\le i\le n} d\theta_i \wedge d\rho_i(\theta) = -d\rho = 0,$$

and thus is symplectic.

Example 6 Let φ be a diffeomorphism of $\mathbb{T}^n$. Define its lift to $\mathbb{T}^n \times \mathbb{R}^n$ by

$$\phi : (\theta, r) \mapsto (\varphi(\theta), r \cdot \varphi'(\theta)^{-1}).$$

This diffeomorphism preserves the 1-form $\lambda = r \cdot d\theta$:

$$\phi^*\lambda = r \cdot \varphi'(\theta)^{-1} \cdot \varphi'(\theta) \cdot d\theta = \lambda,$$

hence the symplectic form $\omega = -d\lambda$ also:

$$\phi^*\omega = -\phi^* d\lambda = -d\phi^*\lambda = -d\lambda = \omega.$$

Proposition 2. *If (ϕ_t) is the flow of a Hamiltonian vector field X_H, $\phi_t^*\omega = \omega$ for all $t \in \mathbb{R}$ (wherever the flow is defined).*

This property is an essential feature of Hamiltonian flows. It implies the the volume $d\theta_1 \wedge \cdots \wedge d\theta_n \wedge dr_1 \wedge \cdots \wedge dr_n$ (= the n-th exterior power of ω, up to a multiplicative constant) is preserved. Yet it is only in the 1980's that Gromov's celebrated non-squeezing theorem pointed out some specifically symplectic properties [18, 26].

We will use proposition 2 in order to build symplectic diffeomorphisms close to the identity.

Proof. The proof is straightforward with the standard toolbox of exterior calculus:

$$\begin{aligned}\phi_t^*\omega - \omega &= \int_0^t \frac{d}{ds}(\phi_s^*\omega)\, ds \quad \text{by the fundamental formula of calculus}\\ &= \int_0^t \phi_s^*(\mathcal{L}_{X_H}\omega)\, ds \quad \text{by definition of the Lie derivative } \mathcal{L}_X.\end{aligned}$$

The Cartan homotopy formula says that $\mathcal{L}_{X_H}\omega = di_{X_H}\omega + i_{X_H}d\omega$, where $i_X\omega := \omega(X,\cdot)$. Since ω has constant coefficients, $d\omega = 0$. Since $i_{X_H}\omega = dH$ and $d^2 = 0$, $di_{X_H}\omega = d^2 = 0$. Hence $\phi_t^*\omega = \omega$. □

We have seen some of the ergodic properties of quasiperiodic flows with a non resonant frequency vector. Let us mention here an important property of invariant tori carrying ergodic quasiperiodic flows. This property says how such tori are embedded in the phase space with respect to the symplectic structure.

Proposition 3 (Herman)**.** *Let T be an invariant embedded torus in M, carrying an ergodic quasiperiodic flow. Then T is* isotropic*, i.e. the 2-form induced on T vanishes.*

Proof. Let $j : \mathbb{T}^n \hookrightarrow M$ be a parametrization of $T = j(\mathbb{T}^n)$ such that the induced flow on $\mathbb{T}^n$ is $\phi_t(\theta) = \theta + t\alpha$, $\alpha \in \mathbb{R}^n$ non resonant. Let Ω be the induced 2-form on $\mathbb{T}^n$:

$$\Omega = j^*\omega = \sum_{1\le k<l\le n} \Omega_{kl}(\theta)\, d\theta_k \wedge d\theta_l.$$

We want to show that $\Omega = 0$. Since (ϕ_t) is a translation,

$$\phi_t^*\Omega(\theta) = \sum \Omega_{kl}(\theta + t\alpha)\, d\theta_k \wedge d\theta_l.$$

Since all trajectories are dense and $\phi_t^*\Omega = \Omega$ for all $t \in \mathbb{R}$, the functions Ω_{kl} are constant on $\mathbb{T}^n$.

But ω has a primitive, and so has Ω: $\Omega = d\Lambda$, whith $\Lambda := -j^*(\sum_k r_k\, d\theta_k)$. Integrate Ω on 2-tori $T_{kl} \subset \mathbb{T}^n$ obtained by fixing all coordinates θ_m, $m = 1, ..., n$, but θ_k and θ_l:

$$\int_{T_{kl}} \Omega = \int_{\mathbb{T}^2} \Omega_{kl}\, d\theta_k\, d\theta_l = \Omega_{kl} \qquad (\forall k, l).$$

On the other hand, by Stokes formula, this integral equals 0. So $\Omega = 0$. □

If in addition T is a perturbation of $\mathbb{T}^n \times \{0\}$, it is the graph of a 1-form ρ over $\mathbb{T}^n$ (up to the identification of the cotangent bundle of $\mathbb{T}^n$ to $\mathbb{T}^n \times \mathbb{R}^n$). The proposition then asserts that ρ is closed.

Exercise 7 Let T be an isotropic submanifold of dimension n in $\mathbb{T}^n \times \mathbb{R}^n$ (T is then said *Lagrangian*). Show that it is invariant by the flow of a Hamiltonian H if and only if it lies in a level of H.

4. Perturbation series and the averaging principle

Consider a Hamiltonian $H(\theta, r)$ on a neighborhood of $\mathbb{T}^n \times \{0\}$ in $\mathbb{T}^n \times \mathbb{R}^n$. We will assume that H depends formally on some parameter ϵ and that, when $\epsilon = 0$, H does not depend on the angles:

$$H(\theta, r) = H_0(r) + \epsilon H_1(\theta, r) + \epsilon^2 H_2(\theta, r) + \cdots .$$

Can we eliminate the dependance of H_1 on θ by a change of coordinates ϵ-close to the identity, and can we then similarly deal with higher order terms?

In order to try to do so, let us consider some auxiliary Hamiltonian ϵF, with flow ϕ_t. We would like to choose F so that $\phi_1^* H = H \circ \phi_1$ does not depend on θ, up to second order terms in ϵ.

Recall that

$$\left.\frac{d}{dt}\right|_{t=0} \phi_t^* H = H' \cdot X_F = X_F \cdot H,$$

where X_F is seen as a derivation operator, and that more generally

$$\frac{d}{dt}\phi_t^* H = \left.\frac{d}{ds}\right|_{s=0} \phi_{t+s}^* H = \phi_t^*(X_F \cdot H).$$

By Taylor's formula (applied to the function $t \mapsto \phi_t^* H$ between $t = 0$ and $t = 1$),

$$\phi_1^* H = H + \epsilon X_F \cdot H + \epsilon^2 \int_0^1 (1-t)\phi_t^*(X_F^2 \cdot H)\, dt.$$

Expanding H in powers of ϵ yields

$$\phi_1^* H = H_0(r) + \epsilon\,(H_1 + X_F \cdot H_0) + O(\epsilon^2).$$

Split H_1 into

$$H_1(\theta, r) = \bar{H}_1(r) + \tilde{H}_1(\theta, r), \quad \bar{H}_1 = \int_{\mathbb{T}^n} H(\theta, r)\, d\theta.$$

We would like to find F so that

$$\tilde{H}_1 + X_F \cdot H_0 = 0,$$

or, equivalently, since $X_F = \partial_r F \cdot \partial_\theta - \partial_\theta F \cdot \partial_r$,

$$\partial_r H_0 \cdot \partial_\theta F = \tilde{H}_1.$$

In general $\bar{H}_1(r)$ is not equal to 0. This means that the frequency vector on the torus $\mathbb{T}^n \times \{r\}$ is modified by terms of order 1 in ϵ. Since it is a conjugacy invariant, it is hopeless to try to eliminate $\bar{H}_1$ (and, indeed, $X_F \cdot H_0$ has zero average).

Among the partial derivatives of the unknown F, the above equation involves only the derivatives with respect to θ. So r can be considered as a fixed parameter. The equation then becomes a first order linear partial differential equation with constant coefficients. Let $\alpha = \partial_r H_0(r) \in \mathbb{R}^n$. Let $\mathcal{L}_\alpha$ be the Lie derivative operator in the direction of the constant vector field α :

$$\mathcal{L}_\alpha : f \mapsto \mathcal{L}_\alpha f = \alpha \cdot \partial_\theta f = \sum_{1 \le j \le n} \alpha_j \frac{\partial f}{\partial \theta_j},$$

defined for functions f on $\mathbb{T}^n$ of various possible classes of regularity.

Let $\mathcal{F}$ be the set of formal Fourier series on $\mathbb{T}^n$ with no constant term.

Lemma 3. *If α is non resonant and $g \in \mathcal{F}$, there is a unique $f \in \mathcal{F}$ such that $\mathcal{L}_\alpha f = g$.*

Proof. By asumption g is a formal series of the form

$$g = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} g_k e^{i2\pi k \cdot \theta}$$

and we look for a series f of the same form, satisfying

$$\sum_k i2\pi k \cdot \alpha f_k e^{i2\pi k \cdot \theta} = \sum_k g_k e^{i2\pi k \cdot \theta}.$$

The unique solution is given by the coefficients

$$f_k = \frac{g_k}{i2\pi k \cdot \alpha} \quad (\forall k \in \mathbb{Z}^n \setminus \{0\}).$$

□

For $s > 0$, let

$$\mathbb{T}^n_s := \{\theta \in \mathbb{C}^n/\mathbb{Z}^n, \ \max_{1 \le j \le n} |\operatorname{Im} \theta_j| \le s\}$$

be the complex extension of $\mathbb{T}^n$ of width s. Let $\mathcal{A}(\mathbb{T}^n_s)$ be the set of real holomorphic functions from (a neighborhood of) $\mathbb{T}^n_s$ to $\mathbb{C}$. Endowed with the supremum norm

$$|f|_s := \sup_{\theta \in \mathbb{T}^n_s} |f(\theta)|,$$

it is a Banach space [15, 6.3].

In order for the operator $\mathcal{L}_\alpha^{-1}$ to send analytic function to analytic functions, one needs some quantitative arithmetic condition preventing α from being too close to any low order resonance —how close depending of the order.

Definition 8 For $\gamma, \tau > 0$, $\alpha \in \mathbb{R}^n$ is (γ, τ)-*Diophantine* if

$$\forall k \in \mathbb{Z}^n \setminus \{0\} \quad |k \cdot \alpha| \ge \frac{\gamma}{|k|^\tau}, \quad |k| := |k_1| + \cdots + |k_n|.$$

Let $D_{\gamma,\tau}$ be the set of all such vectors, and $D_\tau = \cup_{\gamma > 0} D_{\gamma,\tau}$.

The following facts hold:

- Dirichlet's theorem: $D_\tau \neq \emptyset \Leftrightarrow \tau \ge n - 1$.
- If $\tau = n - 1$, D_τ is locally uncountable, has Hausdorf dimension n, but has n-dimensional Lebesgue measure zero.
- If $\tau > n - 1$, $\mathbb{R}^n \setminus D_\tau$ has n-dimensional Lebesgue measure zero. So, the measure of $D_{\gamma,\tau}$ tends to the full measure as γ tends to 0. On the other hand, the trace of $D_{\gamma,\tau}$ on the unit sphere is a Cantor set.

See [2, 29, 30, 34] and references therein for proofs and additional facts.

Proposition 4. *Assume that $\alpha \in D_{\gamma,\tau}$ and let $0 < s < s + \sigma$. If $g \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$, there is a unique function $f \in \mathcal{A}(\mathbb{T}^n_s)$ such that $\mathcal{L}_\alpha f = g$. Besides,*

$$|f|_s \leq C\gamma^{-1}\sigma^{-n-\tau}|g|_{s+\sigma},$$

where the number C depends only on the dimension n and the exponent τ.

This estimate calls for a comment. We have already mentionned Cauchy's Mémoire presented to the Accademia delle Scienze di Torino on October 11, 1831, where he introduced and studied the so-called equations of Hamilton [9]. In the same Mémoire in Celestial Mechanics [10] (!), Cauchy proved the remarkable formula

$$f(z) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f(\zeta)}{\zeta - z}\, d\zeta,$$

where f is a holomorphic function in some complex domain containing a disc centered at z and bounded by the circle $\mathcal{C}$. This formula plays an essential rôle here. By differentiating with respect to z, we get

$$f'(z) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f(\zeta)}{(\zeta - z)^2}\, d\zeta.$$

It follows that if $f \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$, then

$$|f'|_s \leq \sigma^{-1}|f|_{s+\sigma}.$$

More generally, any differential operator of the first order will satisfy a similar kind of estimate. In particular,

$$|\mathcal{L}_\alpha f|_s \leq C|\alpha|\sigma^{-1}|f|_{s+\sigma}, \quad \text{with} \quad |\alpha| := \max_{1\leq j\leq n} |\alpha_j|.$$

The operator $\mathcal{L}_\alpha$ is typical of KAM theory in that both $\mathcal{L}_\alpha$ and its inverse behave like differential operators, due to small denominators.

Proof. Let $g(\theta) = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} g_k\, e^{ik\cdot\theta}$ be the Fourier expansion of g. The unique formal solution to the equation $\mathcal{L}_\alpha f = g$ is given by $f(\theta) = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{g_k}{i\,k\cdot\alpha}\, e^{i\,k\cdot\theta}$.

Since g is analytic, its Fourier coefficients decay exponentially: we find

$$|g_k| = \left| \int_{\mathbb{T}^n} g(\theta)\, e^{-ik\cdot\theta}\, \frac{d\theta}{2\pi} \right| \leq |g|_{s+\sigma} e^{-|k|(s+\sigma)}$$

by shifting the torus of integration to a torus $\operatorname{Im}\theta_j = -\operatorname{sign}(k_j)(s+\sigma)$.

Using this estimate and replacing the small denominators $k\cdot\alpha$ by the estimate defining the Diophantine property of α, we get

$$\begin{aligned} |f|_s &\leq \frac{|g|_{s+\sigma}}{\gamma} \sum_k |k|^\tau\, e^{-|k|\,\sigma} \\ &\leq \frac{2^n|g|_{s+\sigma}}{\gamma} \sum_{\ell\geq 1} \binom{\ell+n-1}{\ell} \ell^\tau\, e^{-\ell\,\sigma} \leq \frac{4^n|g|_{s+\sigma}}{\gamma\,(n-1)!} \sum_\ell (\ell+n-1)^{\tau+n-1}\, e^{-\ell\,\sigma}, \end{aligned}$$

where the latter sum is bounded by

$$\begin{aligned}
\int_1^\infty (\ell+n-1)^{\tau+n-1} e^{-(\ell-1)\sigma}\, d\ell &= \sigma^{-\tau-n} e^{n\sigma} \int_{n\sigma}^\infty \ell^{\tau+n-1} e^{-\ell}\, d\ell \\
&< \sigma^{-\tau-n} e^{n\sigma} \int_0^\infty \ell^{\tau+n-1} e^{-\ell}\, d\ell \\
&= \sigma^{-\tau-n} e^{n\sigma} \Gamma(\tau+n).
\end{aligned}$$

Hence f belongs to $\mathcal{A}(\mathbb{T}^n_s)$ and satisfies the wanted estimate. □

So, we may define $F(\theta,r) := \mathcal{L}_\alpha^{-1} \tilde{H}_1(\theta,r)$ for a fixed value of r chosen so that $\alpha = \partial H_0/\partial r(r) \in D_{\gamma,\tau}$. As well, we may define partial derivatives of F with respect to r at any order, so as to define not only the trace of a function F on $\mathbb{T}^n \times \{r\}$, but the whole infinite jet of a function along this torus; for instance at the first order, we may set

$$\frac{\partial F}{\partial r}(r) := \mathcal{L}_\alpha^{-1} \frac{\partial \tilde{H}_1}{\partial r}(\theta,r).$$

Borel's lemma asserts that such an infinite jet along $\mathbb{T}^n \times \{r\}$ extends to a smooth function. Better, one can show using Whitney's extension theorem that all such jets taken together with r varying among values for which the frequency is (γ,τ)-Diophantine:

$$\frac{\partial H_0}{\partial r}(r) \in D_{\gamma,\tau},$$

extend to a smooth function F. We have thus eliminated the dependence of H_1 on θ along all (γ,τ)-Diophantine tori.

By repeating the procedure, we may do so at any finite order in ϵ. The theorem of Kolmogorov consists in showing the existence of a similar analyic normalization at the infinite order, under some non-degeneracy asumption, as we will now see.

5. Statement of the invariant torus theorem of Kolmogorov

Let $\mathcal{H}$ be the space of germs along $\mathrm{T}^n_0 := \mathbb{T}^n \times \{0\}$ of real analytic Hamiltonians in $\mathbb{T}^n \times \mathbb{R}^n = \{(\theta,r)\}$ ($\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$), endowed with the usual, inductive limit topology (see section 6). The vector field associated with $H \in \mathcal{H}$ is

$$\vec{H}: \quad \dot\theta = \partial_r H, \quad \dot r = -\partial_\theta H.$$

For $\alpha \in \mathbb{R}^n$, let $\mathcal{K}^\alpha$ be the affine subspace of Hamiltonians $K \in \mathcal{H}$ such that $K|_{\mathrm{T}^n_0}$ is constant (i.e. T^n_0 is invariant) and $\vec{K}|_{\mathrm{T}^n_0} = \alpha$:

$$\mathcal{K}^\alpha = \{K \in \mathcal{H},\ \exists c \in \mathbb{R},\ K(\theta,r) = c + \alpha\cdot r + O(r^2)\}, \quad \alpha\cdot r = \alpha_1 r_1 + \cdots + \alpha_n r_n,$$

where $O(r^2)$ are terms of the second ordrer in r, which depend on θ.

Let also $\mathcal{G}$ be the space of germs along T^n_0 of real analytic symplectomorphisms G in $\mathbb{T}^n \times \mathbb{R}^n$ of the following form:

$$G(\theta,r) = (\varphi(\theta), (r+\rho(\theta))\cdot\varphi'(\theta)^{-1}),$$

where φ is an isomorphism of $\mathbb{T}^n$ fixing the origin (meant to straighten the flow on an invariant torus), and ρ is a closed 1-form on $\mathbb{T}^n$ (meant to straighten an invariant torus).

In the whole paper we fix $\alpha \in \mathbb{R}^n$ Diophantine ($0 < \gamma \ll 1 \ll \tau$; see [29]):

$$|k \cdot \alpha| \geq \gamma |k|^{-\tau} \quad (\forall k \in \mathbb{Z}^n \setminus \{0\}), \quad |k| = |k_1| + \cdots + |k_n|$$

and

$$K^o(\theta, r) = c^o + \alpha \cdot r + Q^o(\theta) \cdot r^2 + O(r^3) \in \mathcal{K}^\alpha$$

such that the average of the quadratic form valued function Q^o be non-degenerate:

$$\det \int_{\mathbb{T}^n} Q^o(\theta)\, d\theta \neq 0.$$

Theorem 9 (Kolmogorov [24, 13])**.** *For every $H \in \mathcal{H}$ close to K^o, there exists a unique $(K, G) \in \mathcal{K}^\alpha \times \mathcal{G}$ close to (K^o, id) such that $H = K \circ G$ in some neighborhood of $G^{-1}(\mathrm{T}^n_0)$.*

See [29, 32] and references therein for background. The functional setting below is related to [17].

6. The action of a group of symplectomorphisms

Define complex extensions $\mathbb{T}^n_{\mathbb{C}} = \mathbb{C}^n/\mathbb{Z}^n$ and $\mathrm{T}^n_{\mathbb{C}} = \mathbb{T}^n_{\mathbb{C}} \times \mathbb{C}^n$, and neighborhoods ($0 < s < 1$)

$$\mathbb{T}^n_s = \{\theta \in \mathbb{T}^n_{\mathbb{C}},\ \max_{1 \leq j \leq n} |\mathrm{Im}\,\theta_j| \leq s\} \quad \text{and} \quad \mathrm{T}^n_s = \{(\theta, r) \in \mathrm{T}^n_{\mathbb{C}},\ \max_{1 \leq j \leq n} \max\left(|\mathrm{Im}\,\theta_j|, |r_j|\right) \leq s\}.$$

For complex extensions U and V of real manifolds, denote by $\mathcal{A}(U, V)$ the Banach space of real holomorphic maps from the interior of U to V, which extend continuously on U; $\mathcal{A}(U) := \mathcal{A}(U, \mathbb{C})$.

• Let $\mathcal{H}_s = \mathcal{A}(\mathrm{T}^n_s)$ with norm $|H|_s := \sup_{(\theta,r) \in \mathrm{T}^n_s} |H(\theta, r)|$, such that $\mathcal{H} = \cup_s \mathcal{H}_s$ be their inductive limit.

Fix s_0. There exist ϵ_0 such that $K^o \in \mathcal{H}_{s_0}$ and, for all $H \in \mathcal{H}_{s_0}$ such that $|H - K^o|_{s_0} \leq \epsilon_0$,

$$\left|\det \int_{\mathbb{T}^n} \frac{\partial^2 H}{\partial r^2}(\theta, 0)\, d\theta\right| \geq \frac{1}{2} \left|\det \int_{\mathbb{T}^n} \frac{\partial^2 K^o}{\partial r^2}(\theta, 0)\, d\theta\right| \neq 0. \tag{1}$$

Hereafter we assume that s is always $\geq s_0$. Set $\mathcal{K}^\alpha_s = \{K \in \mathcal{H}_s \cap \mathcal{K}^\alpha,\ |K - K^o|_{s_0} \leq \epsilon_0\}$, and let $\vec{\mathcal{K}}_s \equiv \mathbb{R} \oplus O(r^2)$ be the vector space directing $\mathcal{K}^\alpha_s$.

• Let $\mathcal{D}_s$ be the space of isomorphisms $\varphi \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{T}^n_{\mathbb{C}})$ with $\varphi(0) = 0$ and $\mathcal{Z}_s$ be the space of bounded real holomorphic closed 1-forms on $\mathbb{T}^n_s$. The semi-direct product $\mathcal{G}_s = \mathcal{Z}_s \rtimes \mathcal{D}_s$ acts faithfully and symplectically on the phase space by

$$G : \mathrm{T}^n_s \to \mathrm{T}^n_{\mathbb{C}}, \quad (\theta, r) \mapsto (\varphi(\theta), (\rho(\theta) + r) \cdot \varphi'(\theta)^{-1}), \quad G = (\rho, \varphi), \tag{2}$$

and, to the right, on $\mathcal{H}_s$ by $\mathcal{H}_s \to \mathcal{A}(G^{-1}(\mathbb{T}^n_s))$, $K \mapsto K \circ G$.

• Let $\mathfrak{d}_s := \{\dot\varphi \in \mathcal{A}(\mathbb{T}^n_s)^n,\ \dot\varphi(0) = 0\}$ with norm $|\dot\varphi|_s := \max_{\theta \in \mathbb{T}^n_s} \max_{1 \leq j \leq n} |\dot\varphi_j(\theta)|$, be the space of vector fields on $\mathbb{T}^n_s$ which vanish at 0. Similarly, let $|\dot\rho|_s = \max_{\theta \in \mathbb{T}^n_s} \max_{1 \leq j \leq n}, |\dot\rho_j(\theta)|$

on $\mathcal{Z}_s$. An element $\dot G = (\dot\rho, \dot\varphi)$ of the Lie algebra $\mathfrak{g}_s = \mathcal{Z}_s \oplus \mathfrak{d}_s$ (with norm $|(\dot\rho,\dot\varphi)|_s = \max(|\dot\rho|_s, |\dot\varphi|_s))$ identifies with the vector field

$$\dot G : \mathrm{T}^n_s \to \mathbb{C}^{2n}, \quad (\theta, r) \mapsto (\dot\varphi(\theta), \dot\rho(\theta) - r\cdot \dot\varphi'(\theta)), \tag{3}$$

whose exponential is denoted by $\exp \dot G$. It acts infinitesimally on $\mathcal{H}_s$ by $\mathcal{H}_s \to \mathcal{H}_s$, $K \mapsto K' \cdot \dot G$.

Constants $\gamma_i, \tau_i, c_i, t_i$ below do not depend on s or σ.

Lemma 0. *If $\dot G \in \mathfrak{g}_{s+\sigma}$ and $|\dot G|_{s+\sigma} \le \gamma_0\sigma^2$, then $\exp \dot G \in \mathcal{G}_s$ and $|\exp \dot G - \mathrm{id}\,|_s \le c_0\sigma^{-1}|\dot G|_{s+\sigma}$.*

Proof. Let $\chi_s = \mathcal{A}(\mathrm{T}^n_s)^{2n}$, with norm $\|v\|_s = \max_{\theta\in\mathbb{T}^n_s}\max_{1\le j\le 2n}|v_j(\theta)|$. Let $\dot G \in \mathfrak{g}_{s+\sigma}$ with $|\dot G|_{s+\sigma} \le \gamma_0\sigma^2$, $\gamma_0 := (36n)^{-1}$. Using definition (3) and Cauchy's inequality, we see that if $\delta := \sigma/3$,

$$\|\dot G\|_{s+2\delta} = \max\left(|\dot\varphi|_{s+2\delta}, |\dot\rho + r\cdot\dot\varphi'(\theta)|_{s+2\delta}\right) \le 2n\delta^{-1}|\dot G|_{s+3\delta} \le \delta/2.$$

Let $D_s = \{t \in \mathbb{C},\ |t| \le s\}$ and $F := \left\{f \in \mathcal{A}(D_s\times\mathbb{T}^n_s)^{2n},\ \forall (t,\theta)\in D_s\times\mathbb{T}^n_s,\ |f(t,\theta)|_s \le \delta\right\}$. By Cauchy's inequality, the Lipschitz constant of the Picard operator

$$P : F \to F, \quad f \mapsto Pf, \quad (Pf)(t,\theta) = \int_0^t \dot G(\theta + f(s,\theta))\, ds$$

is $\le 1/2$. Hence, P possesses a unique fixed point $f \in F$, such that $f(1,\cdot) = \exp(\dot G) - \mathrm{id}$ and $|f(1,\cdot)|_s \le \|\dot G\|_{s+\delta} \le c_0\sigma^{-1}|\dot G|_{s+\sigma}$, $c_0 = 6n$.

Also, $\exp \dot G \in \mathcal{G}_s$ because at all times the curve $\exp(t\dot G)$ is tangent to $\mathcal{G}_s$, locally a closed submanifold of $\mathcal{A}(\mathbb{T}^n_s, \mathbb{T}^n_{\mathbb{C}})$ (the method of the variation of constants gives an alternative proof). □

7. A property of infinitesimal transversality

We will show that locally $\vec{\mathcal{K}}_s$ is tranverse to the infinitesimal action of $\mathfrak{g}_s$ on $\mathcal{H}_{s+\sigma}$.

Lemma 1. *For all $(K, \dot H) \in \mathcal{K}^\alpha_{s+\sigma}\times\mathcal{H}_{s+\sigma}$, there exists a unique $(\dot K, \dot G) \in \vec{\mathcal{K}}_s \times \mathfrak{g}_s$ such that*

$$\dot K + K'\cdot\dot G = \dot H \quad \textit{and} \quad \max(|\dot K|_s, |\dot G|_s) \le c_1\sigma^{-t_1}\left(1 + |K|_{s+\sigma}\right)|\dot H|_{s+\sigma}.$$

Proof. We want to solve the linear equation $\dot K + K'\cdot\dot G = \dot H$. Write

$$\begin{cases} K(\theta,r) = c + \alpha\cdot r + Q(\theta)\cdot r^2 + O(r^3) \\ \dot K(\theta,r) = \dot c + \dot K_2(\theta,r), & \dot c\in\mathbb{R}, \quad \dot K_2 \in O(r^2) \\ \dot G(\theta,r) = (\dot\varphi(\theta), R + S'(\theta) - r\cdot\dot\varphi'(\theta)), & \dot\varphi\in\chi_s, \quad \dot R\in\mathbb{R}^n, \quad \dot S\in\mathcal{A}(\mathbb{T}^n_s). \end{cases}$$

Expanding the equation in powers of r yields

$$\left(\dot c + (\dot R + \dot S')\cdot\alpha\right) + r\cdot\left(-\dot\varphi'\cdot\alpha + 2Q\cdot(\dot R + \dot S')\right) + \dot K_2 = \dot H =: \dot H_0 + \dot H_1\cdot r + O(r^2), \tag{4}$$

where the term $O(r^2)$ on the right hand side does not depend on $\dot K_2$.

Fourier series and Cauchy's inequality show that if $g \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$ has zero average, there is a unique function $f \in \mathcal{A}(\mathbb{T}^n_s)$ of zero average such that $L_\alpha f := f' \cdot \alpha = g$, and $|f|_s \leq c\sigma^{-t}|g|_{s+\sigma}$ [29].

Equation (4) is triangular in the unknowns and successiveley yields:

$$\begin{cases} \dot{S} & = L_\alpha^{-1}\left(\dot{H}_0 - \int_{\mathbb{T}^n} \dot{H}_0(\theta)\, d\theta\right) \\ \dot{R} & = \frac{1}{2}\left(\int_{\mathbb{T}^n} Q(\theta)\, d\theta\right)^{-1} \int_{\mathbb{T}^n} \left(\dot{H}_1(\theta) - 2Q(\theta)\cdot \dot{S}'(\theta)\right) d\theta \\ \dot{\varphi} & = L_\alpha^{-1}\left(\dot{H}_1(\theta) - 2Q(\theta)\cdot(\dot{R} + \dot{S}'(\theta))\right) \\ \dot{c} & = \int_{\mathbb{T}^n} \dot{H}_0(\theta)\, d\theta - \dot{R}\cdot\alpha \\ \dot{K}_2 & = O(r^2), \end{cases}$$

and, together with Cauchy's inequality, the wanted estimate. □

8. The local transversality property

Let us bound the discrepancy between the action of $\exp(-\dot{G})$ and the infinitesimal action of $-\dot{G}$.

Lemma 2. *For all $(K, \dot{H}) \in \mathcal{K}^\alpha_{s+\sigma} \times \mathcal{H}_{s+\sigma}$ such that $(1 + |K|_{s+\sigma})|\dot{H}|_{s+\sigma} \leq \gamma_2\sigma^{\tau_2}$, if $(\dot{K}, \dot{G}) \in \vec{\mathcal{K}} \times \mathfrak{g}_s$ solves the equation $\dot{K} + K' \circ \dot{G} = \dot{H}$ (lemma 1), then $\exp \dot{G} \in \mathcal{G}_s$, $|\exp \dot{G} - \mathrm{id}\,|_s \leq \sigma$ and*

$$|(K + \dot{H}) \circ \exp(-\dot{G}) - (K + \dot{K})|_s \leq c_2\sigma^{-t_2}(1 + |K|_{s+\sigma})^2|\dot{H}|^2_{s+\sigma}.$$

Proof. Set $\delta = \sigma/2$. Lemmas 0 and 1 show that, under the hypotheses for some constant γ_2 and for $\tau_2 = t_1 + 1$, we have $|\dot{G}|_{s+\delta} \leq \gamma_0\delta^2$ and $|\exp \dot{G} - \mathrm{id}\,|_s \leq \delta$.

Let $H = K + \dot{H}$. Taylor's formula says

$$\mathcal{H}_s \ni H \circ \exp(-\dot{G}) = H - H' \cdot \dot{G} + \left(\int_0^1 (1-t)\, H'' \circ \exp(-t\dot{G})\, dt\right) \cdot \dot{G}^2$$

or, using the fact that $H = K + \dot{K} + K' \cdot \dot{G}$,

$$H \circ \exp(-\dot{G}) - (K + \dot{K}) = -(\dot{K} + K' \cdot \dot{G})' \cdot \dot{G} + \left(\int_0^1 (1-t)\, H'' \circ \exp(-t\dot{G})\, dt\right) \cdot \dot{G}^2.$$

The wanted estimate thus follows from the estimate of lemma 1 and Cauchy's inequality. □

Let $B_{s,\sigma} = \{(K, \dot{H}) \in \mathcal{K}^\alpha_{s+\alpha} \times \mathcal{H}_{s+\sigma},\ |K|_{s+\sigma} \leq \epsilon_0,\ |\dot{H}|_{s+\sigma} \leq (1+\epsilon_0)^{-1}\gamma_2\sigma^{\tau_2}\}$ (recall (1)).

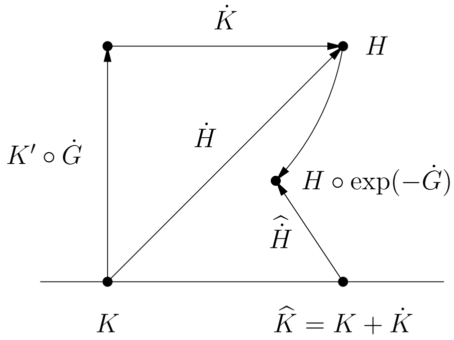
According to lemmas 1-2, the map $\phi : B_{s,\sigma} \to \mathcal{K}^\alpha_s \times \mathcal{H}_s$,

$$\phi(K, \dot{H}) = (K + \dot{K}, (K + \dot{H}) \circ \exp(-\dot{G}) - (K + \dot{K})),$$

satisfies, if $(\hat{K}, \hat{\dot{H}}) = \phi(K, \dot{H})$,

$$|\hat{K} - K|_s \leq c_3\sigma^{-t_3}|\dot{H}|_{s+\sigma}, \quad |\hat{\dot{H}}|_s \leq c_3\sigma^{-t_3}|\dot{H}|^2_{s+\sigma}.$$

Theorem 10 applies and shows that if $H - K^o$ is small enough in $\mathcal{H}_{s+\sigma}$, the sequence $(K_j, \dot{H}_j) = \phi^j(K^o, H - K^o)$, $j \geq 0$, converges towards some $(K, 0)$ in $\mathcal{K}_s^\alpha \times \mathcal{H}_s$.



Let us keep track of the $\dot{G}_j$'s solving with the $\dot{K}_j$'s the successive linear equations $\dot{K}_j + K_j' \cdot \dot{G}_j = \dot{H}_j$ (lemma 1). At the limit,

$$K := K^o + \dot{K}_0 + \dot{K}_1 + \cdots = H \circ \exp(-\dot{G}_0) \circ \exp(-\dot{G}_1) \circ \cdots .$$

Moreover, lemma 1 shows that $|\dot{G}_j|_{s_{j+1}} \leq c_4 \sigma_j^{-t_4} |\dot{H}_j|_{s_j}$, hence the isomorphisms $\gamma_j := \exp(-\dot{G}_0) \circ \cdots \circ \exp(-\dot{G}_j)$, which satisfy

$$|\gamma_n - \mathrm{id}\,|_{s_{n+1}} \leq |\dot{G}_0|_{s_1} + ... + |\dot{G}_n|_{s_{n+1}},$$

form a Cauchy sequence and have a limit $\gamma \in \mathcal{G}_s$. At the expense of decreasing $|H - K^o|_{s+\sigma}$, by the inverse function theorem, $G := \gamma^{-1}$ exists in $\mathcal{G}_{s-\delta}$ for some $0 < \delta < s$, so that $H = K \circ G$.

Appendix. A fixed point theorem

Let $(E_s, |\cdot|_s)_{0<s<1}$ and $(F_s, |\cdot|_s)_{0<s<1}$ be two decreasing families of Banach spaces with increasing norms. On $E_s \times F_s$, set $|(x,y)|_s = \max(|x|_s, |y|_s)$. Fix $C, \gamma, \tau, c, t > 0$.

Let

$$\phi : B_{s,\sigma} := \{(x,y) \in E_{s+\sigma} \times F_{s+\sigma},\ |x|_{s+\sigma} \leq C,\ |y|_{s+\sigma} \leq \gamma\sigma^\tau\} \to E_s \times F_s$$

be a family of operators commuting with inclusions, such that if $(X, Y) = \phi(x, y)$,

$$|X - x|_s \leq c\sigma^{-t}|y|_{s+\sigma} \quad \text{and} \quad |Y|_s \leq c\sigma^{-t}|y|_{s+\sigma}^2.$$

In the proof of theorem 9, "$|x|_{s+\sigma} \leq C$" allows us to bound the determinant of $\int_{\mathbb{T}^n} Q(\theta)d\theta$ away from 0, while "$|y|_{s+\sigma} \leq \gamma\sigma^\tau$" ensures that $\exp \dot{G}$ is well defined.

Theorem 10. *Given $s < s + \sigma$ and $(x, y) \in B_{s,\sigma}$ such that $|y|_{s+\sigma}$ is small, the sequence $(\phi^j(x, y))_{j\geq 0}$ exists and converges towards a fixed point $(\xi, 0)$ in $B_{s,0}$.*

Proof. It is convenient to first assume that the sequence is defined and $(x_j, y_j) := F^j(x, y) \in B_{s_j,\sigma_j}$, for $s_j := s + 2^{-j}\sigma$ and $\sigma_j := s_j - s_{j+1}$. We may assume $c \geq 2^{-t}$, so that $d_j := c\sigma_j^{-t} \geq 1$. By induction, and using the fact that $\sum 2^{-k} = \sum k 2^{-k} = 2$,

$$\begin{aligned} |y_j|_{s_j} &\leq d_{j-1}|y_{j-1}|^2_{s_{j-1}} \\ &\leq \cdots \\ &\leq |y|^{2^j}_{s+\sigma} \prod_{0\leq k\leq j-1} d_k^{2^{k+1}} \\ &\leq \Big(|y|_{s+\sigma} \prod_{k\geq 0} d_k^{2^{-k-1}}\Big)^{2^j} = \big(c4^t\sigma^{-t}|y|_{s+\sigma}\big)^{2^j}. \end{aligned}$$

Given that $\sum_{n\geq 0}\mu^{2^n} \leq 2\mu$ if $2\mu \leq 1$, we now see by induction that if $|(x,y)|_{s+\sigma}$ is small enough, (x_j, y_j) exists in B_{s_j,σ_j} for all $j \geq 0$, y_j converges to 0 in F_s and the series $x_j = x_0 + \sum_{0\leq k\leq j-1}(x_{k+1} - x_k)$ converges normally towards some $\xi \in E_s$ with $|\xi|_s \leq C$. □

These lecture notes were written for the Ciclo di Lezioni organized by Vivina Barutello, Susanna Terracini and Veronica Felli at the Università di Bicocca in 2011, and for the Jornades on interactions between dynamical systems and partial differential equations organized by Amadeu Delshams, Tere Seara and Xavier Cabre at the Universitat Politècnica de Catalunya in 2013. Thank you to the organizers.

References

[1] V. I. Arnold. Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Uspehi Mat. Nauk*, 18(5 (113)):13–40, 1963.

[2] V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963.

[3] V. I. Arnold. *Geometrical methods in the theory of ordinary differential equations*, volume 250 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. Springer-Verlag, New York, 1983. Translated from the Russian by Joseph Szücs, Translation edited by Mark Levi.

[4] V. I. Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1989. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[5] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical aspects of classical and celestial mechanics*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, third edition, 2006. Translated from the Russian original by E. Khukhro.

[6] V.I. Arnold and A. Avez. *Ergodic problems of classical mechanics*. Advanced book classics. Addison-Wesley, 1989.

[7] J.-B. Bost. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel,...). *Astérisque*, 1984/85(133-134):113–157, 1986. Séminaire Bourbaki.

[8] H. W. Broer, G. B. Huitema, and M. B. Sevryuk. *Quasi-periodic motions in families of dynamical systems*, volume 1645 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Order amidst chaos.

[9] A. Cauchy. Extrait du Mémoire présenté à l'Académie de Turin (sans titre), October 11, 1831. `http://www.accademiadellescienze.it/TecaViewer/index.jsp?RisIdr=TECA0000014562&keyworks=Cauchy`.

[10] A. Cauchy. Sur la mécanique céleste et sur un nouveau calcul qui s'applique à un grande nombre de questions diverses. Mémoire présenté à l'Académie de Turin le 11 octobre, 1831. Œuvres complètes, série II, tome 15.

[11] A. Celletti and L. Chierchia. KAM stability and celestial mechanics. *Mem. Amer. Math. Soc.*, 187(878):viii+134, 2007.

[12] L. Chierchia. KAM lectures. In *Dynamical Systems. Part I*, Pubbl. Cent. Ric. Mat. Ennio Giorgi, pages 1–55. Scuola Norm. Sup., Pisa, 2003.

[13] L. Chierchia. A. N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. A comment on: "On conservation of conditionally periodic motions for a small change in Hamilton's function". *Regul. Chaotic Dyn.*, 13(2):130–139, 2008.

[14] R. de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.

[15] R. E. Edwards. *Functional analysis*. Dover Publications Inc., New York, 1995. Theory and applications, Corrected reprint of the 1965 original.

[16] J. Féjoz. Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). *Ergodic Theory Dynam. Systems*, 24(5):1521–1582, 2004.

[17] J. Féjoz and M. Garay. Un théorème sur les actions de groupes de dimension infinie. *C. R. Math. Acad. Sci. Paris*, 348(7-8):427–430, 2010.

[18] M. Gromov. Pseudoholomorphic curves in symplectic manifolds. *Invent. Math.*, 82(2):307–347, 1985.

[19] V. Guillemin and S. Sternberg. *Symplectic techniques in physics*. Cambridge University Press, Cambridge, second edition, 1990.

[20] P.R. Halmos. *Lectures on ergodic theory*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2006.

[21] W. R. Hamilton. On a general method in dynamics. *Philosophical Transactions of the Royal Society*, pages 95–144, 1834. Lu le 10 avril 1834. Mathematical Works (Cambridge University Press), volume IV.

[22] W. R. Hamilton. Second essay on a general method in dynamocs. *Philosophical Transactions of the Royal Society*, pages 247–308, 1834. Lu le 15 janvier 1835. Mathematical Works (Cambridge University Press), volume IV.

[23] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by A. Katok and L. Mendoza.

[24] A. N. Kolmogorov. On the conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR (N.S.)*, 98:527–530, 1954.

[25] J.-L. Lagrange. *Mécanique analytique*. Gauthier-Villars, 1888. Quatrième édition en deux volumes. Œuvres complètes (Gauthier-Villars), volumes XI et XII. (Première édition en 1808).

[26] D. McDuff and D. Salamon. *Introduction to symplectic topology*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1998.

[27] J. Moser. *Stable and random motions in dynamical systems*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2001. With special emphasis on celestial mechanics, Reprint of the 1973 original, With a foreword by Philip J. Holmes.

[28] R. Pérez-Marco. KAM techniques in PDE. *Astérisque*, 290:Exp. No. 908, ix, 307–317, 2003. Séminaire Bourbaki. Vol. 2001/2002.

[29] J. Pöschel. A lecture on the classical KAM theorem. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 707–732. Amer. Math. Soc., Providence, RI, 2001.

[30] H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 598–624. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.

[31] D. A. Salamon. The Kolmogorov-Arnold-Moser theorem. *Math. Phys. Electron. J.*, 10:Paper 3, 37 pp. (electronic), 2004.

[32] M. B. Sevryuk. The classical KAM theory at the dawn of the twenty-first century. *Mosc. Math. J.*, 3(3):1113–1144, 1201–1202, 2003. Dedicated to V. I. Arnold on the occasion of his 65th birthday.

[33] M. B. Sevryuk. KAM tori: persistence and smoothness. *Nonlinearity*, 21(10):T177–T185, 2008.

[34] J.-C. Yoccoz. An introduction to small divisors problems. In *From number theory to physics (Les Houches, 1989)*, pages 659–679. Springer, Berlin, 1992.

E-mail address: jacques.fejoz@dauphine.fr

Université Paris-Dauphine – Paris VI & Observatoire de Paris