The elements of Hamiltonian perturbation theory

Giancarlo Benettin

Università di Padova, Dipartimento di Matematica P & A
INFM and GNFM. Via G. Belzoni 7, 35131 Padova, Italy.
benettin@math.unipd.it

Foreword. The purpose of these lectures is to introduce in the simplest possible way the "elements" — i.e. the basic facts, ideas, techniques, results — of Hamiltonian perturbation theory. The exposition is oriented to exponential estimates and Nekhoroshev theorem, but a few pages are devoted to KAM theory, too. Some attention is deserved to the geometric aspects of perturbation theory, which in general are less known than the analytic–arithmetic ones. Geometry plays a basic role, in particular, in the extension of perturbation theory to "degenerate systems" (the perturbed Kepler problem; the fast rotating rigid body). We shall not develop a general theory of these systems, whose study is rather recent in spite of their physical significance, but we shall discuss in some detail one of them, namely the perturbed Euler–Poinsot rigid body. We have no space, unfortunately, for the numerous applications of exponential estimates and Nekhoroshev theory to physical systems, which are rather abundant (for a survey of applications, see [11]). The lectures are mainly addressed to non expert readers, and for this reason, as well as to introduce some language and notations, we begin by a preliminary section where we recall some very elementary facts on Hamiltonian systems and canonical transformations.

Contents

1 Hamiltonian systems and canonical transformations

1.1 Local Hamiltonian systems

First of all, we recall here a few basic notions and fix notations. Given an open domain $D \subset \mathbb{R}^{2n}$, endowed with coordinates $(p,q) = (p_1,\ldots,p_n,q_1,\ldots,q_n)$, the set of differential equations (*"Hamilton equations"*)

$$\dot p_j = -\frac{\partial H}{\partial q_j} \ , \qquad \dot q_j = \frac{\partial H}{\partial p_j} \ , \qquad j = 1,\ldots n \ ,$$

shortly

$$\dot p = -\frac{\partial H}{\partial q} \ , \qquad \dot q = \frac{\partial H}{\partial p} \ , \tag{1.1}$$

is called *Hamiltonian system* of *Hamiltonian* H. Using the compact notation

$$x = (x_1,\ldots,x_{2n}) = (p_1,\ldots,p_n,q_1,\ldots,q_n)$$

system (1.1) takes the form

$$\dot x = X_H(x) \ , \tag{1.2}$$

with $X_H = (-\frac{\partial H}{\partial q}, \frac{\partial H}{\partial p})$; X_H is called the *Hamiltonian vector field* of H. The relation between H and X_H is conveniently written

$$EX_H = \partial_x H \ , \tag{1.3}$$

where $\partial_x H = (\frac{\partial H}{\partial x_1},\ldots,\frac{\partial H}{\partial x_{2n}})$ is the $2n$-tuple of derivatives of H (the coefficients of the differential $\mathrm{d}H$) and E is the $2n \times 2n$ block matrix

$$E = \begin{pmatrix} 0 & -\mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix} \ , \qquad E^{-1} = E^T = -E \ ,$$

$\mathbb{I}$ being the $n \times n$ identity; E is called the *symplectic identity.* Equations (1.2) and (1.3) together provide a compact form of Hamilton equations (1.1).

Familiar examples, actually the ones we shall mainly deal with in these lectures, are

(1) A set of n uncoupled harmonic oscillators:

$$H(p,q) = \frac{1}{2}\sum_{j=1}^{n}(p_j^2 + \omega_j^2 q_j^2) \ . \tag{1.4}$$

(2) A chain of n rotators with local interaction:

$$H(p,q) = \sum_{j=1}^{n}\frac{p_j^2}{2} + V(q) \ , \qquad V(q) = -\varepsilon\sum_{j=1}^{n}\cos(q_j - q_{j-1}) \ ,$$

$q_0 \equiv q_n$; for positive ε, V is minimal when all phases q_j are equal.

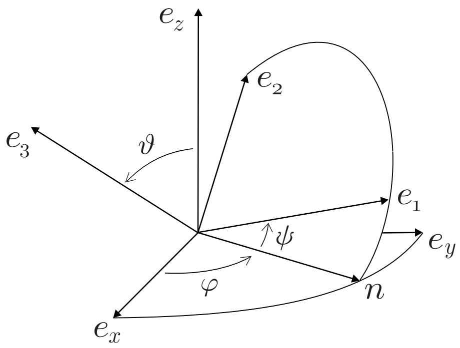

Figure 1: *The Euler angles ϑ, φ, ψ.*

(3) A point particle of unitary mass in a central force field, in two dimensions: in polar coordinates $(p, q) = (p_r, p_\vartheta, r, \vartheta)$ the Hamiltonian is

$$H(p_r, p_\vartheta, r, \vartheta) = \frac{p_r^2}{2} + \frac{p_\vartheta^2}{2r^2} + V(r) \; ; \tag{1.5}$$

$V(r) = -c/r$, $c > 0$, in the Kepler case.

(4) A rigid body with a fixed point in a positional potential: considering for simplicity the symmetric case (moments of inertia $A_1 = A_2 \neq A_3$), and using the familiar Euler angles ϑ, φ, ψ (see figure 1) and the corresponding momenta $p_\vartheta, p_\varphi, p_\psi$, H has the form

$$H(p_\vartheta, p_\varphi, p_\psi, \vartheta, \varphi, \psi) = \frac{p_\vartheta^2}{2A_1} + \frac{(p_\varphi - p_\psi \cos\vartheta)^2}{2A_1 \sin^2\vartheta} + \frac{p_\psi^2}{2A_3} + V(\vartheta, \varphi, \psi) \; . \tag{1.6}$$

We shall denote by Φ_H^t the Hamiltonian flow of H, i. e. $(p(t), q(t)) = \Phi_H^t(p^o, q^o)$, where p^o, q^o denote the initial data. The Hamiltonian H is always a constant of motion, namely $H(\Phi_H^t(p, q)) = H(p, q)$ for all p, q and all t such that $\Phi_H^t(p, q)$ does not escape the domain of definition D. If the flow is complete, i.e. $\Phi_H^t : D \to D$ is defined for all $t \in \mathbb{R}$, then $\{\Phi_H^t, t \in \mathbb{R}\}$ is a group, with composition law $\Phi_H^t \circ \Phi_H^s = \Phi_H^{t+s}$.

1.2 Canonical transformations

The above definition of Hamiltonian system uses coordinates, and it is very important, both conceptually and practically, to understand the class of coordinates changes which leave invariant the form of Hamilton equations. Such transformations are usually called *canonical.*

A. The definition of canonical transformation. Let us denote the change of variables by

$$(p, q) = w(p', q') \; , \qquad (p', q') \in D' \subset \mathbb{R}^{2n}$$

(w is assumed to be a local diffeomorphism $D' \to w(D') \subset D$), and let H' denote the pull-back of H:

$$H' = H \circ w \ , \qquad \text{i.e.} \quad H'(p',q') = H(w(p',q')) \ .$$

DEFINITION 1.1 *The diffeomorphism $(p,q) = w(p',q')$ is said to be canonical, if for any H the Hamilton equations (1.1) are equivalent to the corresponding primed equations $\dot p' = -\frac{\partial H'}{\partial q'}$, $\dot q' = \frac{\partial H'}{\partial p'}$, the new Hamiltonian being $H' = H \circ w$.*

If this happens, then solutions are mapped into solutions, namely

$$\Phi^t_H \circ w = w \circ \Phi^t_{H'} \ . \tag{1.7}$$

In general, a differential equation $\dot x = X(x)$ transforms, under a coordinates change $x = w(x')$, into $\dot x' = X'(x')$, the vector fields X and X' being related by

$$J(x')X'(x') = X(w(x')) \ ,$$

where J denotes the Jacobian matrix of w, $J_{ij} = \frac{\partial w_i}{\partial x_j}$. According to the above definition, w is canonical if for any H

$$\big(X_H\big)' = X_{H'} \ ; \tag{1.8}$$

using (1.3), as well as the trivial relation $\partial_{x'} H' = J^T \partial_x H \circ w$, one immediately concludes that

PROPOSITION 1.2 *The transformation w is canonical iff its Jacobian matrix J satisfies the algebric relation*

$$J^T E J = E \ . \tag{1.9}$$

Matrices J satisfying (1.9) are called *symplectic* and form a group, called the symplectic group; correspondingly, canonical transformations also form a group. Matrix E itself is symplectic and if J is symplectic, so is also J^T.

All point transformations $q = u(q')$ extend canonically to momenta, by setting

$$p = v(p',q') = \Big(\frac{\partial u}{\partial q'}(q')\Big)^{-T} p'$$

(the superscript $-T$ denotes transposition and inversion); the proof is trivial for linear u, a little tedious for generic u. For $n = 1$, the elementary rescaling

$$p = \alpha p' \ , \qquad q = q'/\alpha \ ,$$

is canonical and for $\alpha = \sqrt{\omega}$ it changes the Hamiltonian of the harmonic oscillator $H = \frac{1}{2}(p^2 + \omega^2 q^2)$ into $H'(p',q') = \frac{1}{2}\omega(p'^2 + q'^2)$. The level lines pass from ellipses to circles, see figure 2, and the motion is uniform on them, with angular velocity $-\omega$. The further canonical change of coordinates from p', q' to I, φ:

$$p' = \sqrt{2I}\cos\varphi \ , \qquad q' = \sqrt{2I}\sin\varphi \ , \tag{1.10}$$

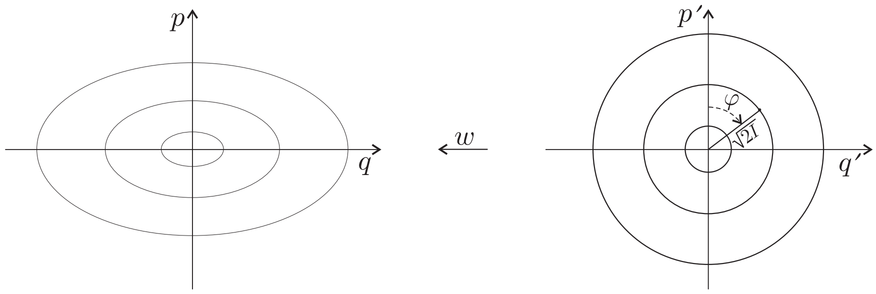

Figure 2: *The canonical transformation from* (p,q) *to* (p',q') *to* (I,φ).

defined for positive I and all φ, changes H' into

$$H''(I,\varphi) = \omega I \ ;$$

correspondingly the motion gets linear,

$$I(t) = I^o \ , \qquad \varphi(t) = \varphi^o + \omega t \ . \tag{1.11}$$

This elementary description of the motion provides a good starting point to investigate the (difficult) problem of several coupled oscillators. As is remarkable, the polar–like coordinates I, φ mix together the usual coordinate and momentum, and do not have any equivalent in Lagrangian mechanics where only point transformations are meaningful.

B. Canonical transformations and Poisson brackets. It is useful to introduce other conditions equivalent to (1.9). A particularly interesting one makes use of Poisson brackets. Let us recall that for any pair of (regular) functions $f, g : \mathbb{R}^{2n} \to \mathbb{R}$, their Poisson bracket $\{f,g\}$ is a function: $\mathbb{R}^{2n} \to \mathbb{R}$ defined as

$$\{f,g\} = \frac{\partial f}{\partial q} \cdot \frac{\partial g}{\partial p} - \frac{\partial f}{\partial p} \cdot \frac{\partial g}{\partial q} = E\partial_x f \cdot \partial_x g \ , \tag{1.12}$$

the dot denoting the usual Euclidean scalar product. The operation is bilinear and antisymmetric. For any vector field X, the linear operator $L_X = X \cdot \partial_x$ is called the *Lie derivative* associated to X, and if Φ_X^t denotes the flow of X, it is

$$(L_X f)(x) = \frac{\mathrm{d}}{\mathrm{d}t} f(\Phi_X^t(x))\Big|_{t=0} \ .$$

For any H, it is clearly

$$\{\,.\,, H\} = L_{X_H} \ , \qquad \text{shortly} \quad L_H \ .$$

By iteration it immediately follows

$$\frac{\mathrm{d}^k f}{\mathrm{d}t^k} = L_H^k f \ ,$$

and so formally

$$f \circ \Phi_H^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} L_H^k f = e^{tL_H} f \ . \tag{1.13}$$

These series, called Lie series, are frequently used in perturbation theory. For analytic functions and small t, they are easily seen to converge [59]: we however shall not discuss convergence and use instead truncated expansions with remainders (see section 2.4). The definition (1.12) makes explicit use of coordinates, and it is obviously important to know which changes of coordinates $(p,q) = w(p',q')$ "preserve the Poisson brackets", namely are such that

$$\{f \circ w, g \circ w\} = \{f, g\} \circ w \qquad \forall\, f, g \tag{1.14}$$

(drivatives with respect to p', q' on the left, with respect to p, q on the right). The answer, easy and expected, is

PROPOSITION 1.3 *The change of variables w preserves the Poisson bracket iff it is canonical. Moreover, the Poisson bracket of any pair of functions is preserved iff the elementary Poisson brackets are preserved, namely iff, denoting shortly $w(p',q') = (p(p',q'), q(p',q'))$, it is*

$$\{p_i, p_j\} = 0 \ , \qquad \{q_i, q_j\} = 0 \ , \qquad \{q_i, p_j\} = 1$$

for any i, j.

C. Canonical transformations and differential forms. A further condition for canonicity, equivalent to (1.9) and (1.14), is easily stated using the language of differential forms. Two differential forms play indeed a relevant role in the geometric formulation of Hamiltonian mechanics, namely the *canonical 1-form*

$$\vartheta = p \cdot \mathrm{d}q$$

and its exterior derivative $\Omega = \mathrm{d}\vartheta$, calld *canonical 2-form:*

$$\Omega = \mathrm{d}p \wedge \mathrm{d}q = \sum_{i<j} E_{ij}\, \mathrm{d}x_i \wedge \mathrm{d}x_j \ .$$

Quite generally, if a bilinear form is represented in coordinates by a matrix A, after a change of coordinates $x = w(x')$ the new representative matrix is

$$A' = J^T A J \ .$$

But the matrix representative of Ω is E, which is invariant for the above operation whenever w is canonical. It immediately follows

PROPOSITION 1.4 *The transformation w is canonical iff the canonical 2-form is preserved, i.e. iff*

$$\mathrm{d}p' \wedge \mathrm{d}q' = \mathrm{d}p \wedge \mathrm{d}q \ ; \tag{1.15}$$

equivalently, iff $\vartheta' - \vartheta$ is closed, i.e. there exists locally f such that

$$p' \cdot \mathrm{d}q' - p \cdot \mathrm{d}q = \mathrm{d}f \ .$$

The language of differential forms, though not mandatory, is neverteless occasionally useful in Hamiltonian perturbation theory.

1.3 Hamiltonian systems on a manifold

We started with the local notion (1.1) of Hamiltonian system, defined in an open domain $D \subset \mathbb{R}^{2n}$. But once we know how to transform the Hamilton equations from a set of coordinates to another, we have in our hands more, and it gets easy to pass to the deeper notion of Hamiltonian system on a manifold. (A minimal familiarity is here assumed with the language of differential manifolds and the notion of fiber bundle; for a nice introduction, oriented towards mechanics, see for example [88]. A classical reference on the subject is [1].) Such a notion is especially important for constrained systems, like the sperical pendulum of the rigid body, for which the configuration space is not an Euclidean space but an n-dimensional manifold Q. In such a case Lagrangian and Hamiltonian mechanics are properly defined, respectively, on the so–called tangent and cotangent fiber bundles of Q, commonly denoted TQ and T^*Q. For example, for a rigid body with a fixed point, it is $Q = SO(3)$, namely the group of orthogonal 3×3 matrices with determinant one. Euler angles are local coordinates on $SO(3)$, i.e. provide a *chart* for it, and two charts of Euler angles (with different polar axis) cover $SO(3)$, i.e. provide an *atlas* of the manifold. Correspondingly, a set of Euler coordinates $(p_\vartheta, p_\varphi, p_\psi, \vartheta, \varphi, \psi)$ provides a chart for $T^*SO(3)$ and two charts provide an atlas (for a closer investigation on the structure of $T^*SO(3)$, see section 6).

Let M be a manifold of even dimension $2n$, endowed by an atlas $\mathcal{A}$. For any pair of charts of local coordinates $x = (p,q) \in D$ and $x' = (p',q') \in D'$ (see figure 3), such that the pre–images of D and D' on the manifold do overlap, the transition map $x = w(x')$ is well defined and the differential structure of M just requires, as the compatibility condition between charts, that w is a diffeomorphism.

Definition 1.5 *The atlas $\mathcal{A}$ is said to be symplectic if, as a further compatibility condition, all transition maps are canonical.*

Having at disposal a symplectic atlas[1] is clearly enough to define Hamiltonian systems on M. Indeed, given any function $\mathcal{H} : M \to \mathbb{R}$, its local representatives H are well defined, and by definition are such that, for overlapping domains, $H' = H \circ w$. In each chart the local vector fields X_H are then also well defined, via (1.1) or (1.3), and thanks to canonicity, such vector fields are compatible, namely satisfy (1.8); so there exists a vector field $\mathcal{X}_\mathcal{H}$ on M, of which they are local representatives. Correspondingly, thanks to (1.7), the local flows Φ_H^t glue so as to define a flow $\varPhi_\mathcal{H}$ on M. Also observe that, thanks to (1.9), a 2-form $\varOmega$

[1] Or sub–atlas: indeed, once a symplectic atlas exists, it can always be enlarged in a non symplectic way, though of course in the additional charts the equations of motion lose the form (1.1). The use of non–symplectic local coordinates in Hamiltonian dynamics is not common but occasionally useful.

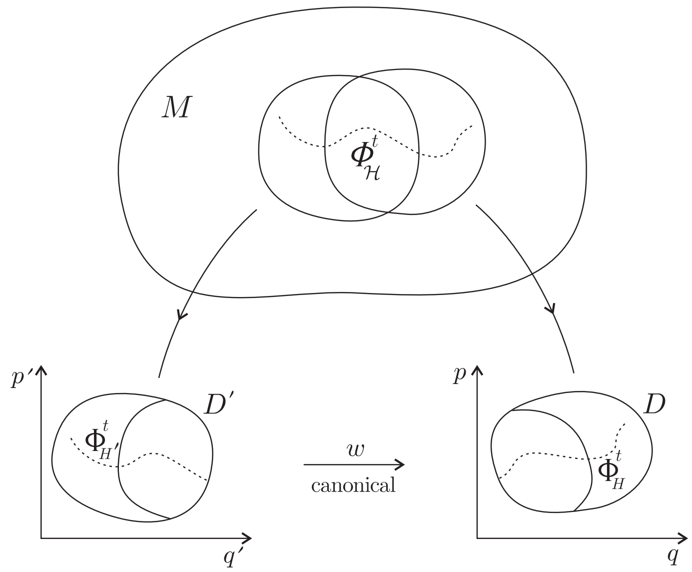

Figure 3: *Illustrating the notion of symplectic manifold.*

is naturally defined on M, namely the form represented locally, in all charts, by matrix E. The global relation corresponding to the local relation (1.3), relating $\mathcal{H}$ to its vector field $\mathcal{X}_{\mathcal{H}}$, is then

$$\Omega(\mathcal{X}_{\mathcal{H}},\,.\,) = \mathrm{d}\mathcal{H}\ ; \tag{1.16}$$

as is quite remarkable, this is a coordinate–free expression, which can be used as the intrinsic definition of the Hamiltonian vector field $\mathcal{X}_{\mathcal{H}}$ for any given $\mathcal{H}$. Note that the 2-form Ω is nondegenerate (that is, if $\Omega(\mathcal{X},\mathcal{Y}) = 0$ for any $\mathcal{Y}$, then $\mathcal{X} = 0$; correspondingly, $\det E \neq 0$), as is necessary for (1.16) to be a good definition, and is closed (it is $\mathrm{d}\Omega = 0$, as trivially follows from the fact that E is a constant matrix).

A differential manifold M of even dimension $2n$, endowed with a symplectic atlas $\mathcal{A}$, is called a *symplectic manifold.* We naturally arrived to this notion, starting from the local notion of Hamiltonian system, just observing that, if transition functions are canonical, then all local objects — the Hamiltonian, its vector field, its flow, the canonical 2-form — are local representatives of well defined objects on the manifold. In the literature, however, it is customary to proceed more axiomatically, and to define symplectic manifolds in a different though equivalent way:

DEFINITION 1.6 *A symplectic manifold is a pair (M,Ω), where M is a differential manifold of even dimension and Ω is a non degenerate closed 2-form on it.*

The 2-form is indeed all one needs to define Hamiltonian systems through (1.16), without even mentioning coordinates. (Poisson brackets and all quantities entering Hamiltonian mechanics are immediately defined as well.) Quite obviously, if one is given a differential manifold M endowed with a symplectic atlas, then Ω is immediately defined as the 2-form represented in all charts by E, so (M, Ω) is a symplectic manifold according to the axiomatic definition. But the converse is also true, namely any symplectic manifold (M, Ω) admits a symplectic atlas:

PROPOSITION 1.7 (Darboux theorem). *Given a symplectic manifold (M, Ω), there always exists a symplectic atlas $\mathcal{A}$ of M, such that in each chart of $\mathcal{A}$ the 2-form Ω is represented by the symplectic identity E.*

For a proof, see [1].

1.4 Weakly canonical transformations

According to our definition, a transformation $(p, q) = w(p', q')$ is canonical iff for any H (1.8) holds, H' *being the pull-back* $H \circ w$ *of* H. A possible weaker notion is the following:

DEFINITION 1.8 *A diffeomorphism $(p, q) = w(p', q')$ is said to be weakly canonical, if for any H there exist H' such that the Hamilton equations (1.1) are equivalent to the corresponding primed equations $\dot{p}' = -\frac{\partial H'}{\partial q'}$, $\dot{q}' = \frac{\partial H'}{\partial p'}$.*

(That is: Hamiltonian vector fields are transformed by w into Hamiltonian vector fields, with no need that H and H' are related by $H' = H \circ w$.) The notion is effectively weaker, as is shown by the "homogeneous rescaling"

$$p_j = \alpha_j p'_j \ , \qquad q_j = \beta_j q'_j \ , \qquad \alpha_j \beta_j = \gamma \qquad \forall i \ , \tag{1.17}$$

which for $\gamma \neq 1$ is canonical only in the weaker sense: it is an easy matter to check that the above change of coordinates associates to any Hamiltonian vector field X_H the new Hamiltonian vector field $X_{H'}$, with

$$H' = \gamma^{-1} \, H \circ w \ .$$

Adding homogeneous rescalings enlarges the group of canonical transformations to a wider group. But there is nothing else:

PROPOSITION 1.9 *Such a larger group includes all weakly canonical transformations.*

(The proof is a tedious exercise.) So, the generalization from canonical to weakly canonical transformation, though not empty, is trivial. By the way, let us observe that multiplying the Hamiltonian by a constant is equivalent to rescaling time: for any $c \neq 0$ it is easy to check that

$$\Phi^t_{cH} = \Phi^{ct}_H \ .$$

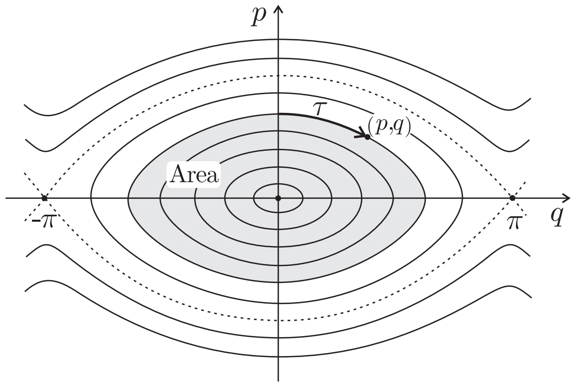

Figure 4: *The construction of the action–angle variables for the pendulum.*

So, if the homogeneous rescaling (1.17) is completed by the rescaling of time $t = c^{-1}t'$, then the new Hamiltonian is $H' = \frac{c}{\gamma} H \circ w$. Rescaling coordinates and time is occasionally useful to eliminate spurious constants, due to the use of inappropriate scales (see sections 4.1,4.3).

2 Integrable and nearly–integrable Hamiltonian systems

2.1 Examples of integrable systems

Let us preliminarily introduce a few familiar examples of integrable systems.

(1) The harmonic oscillator,

$$H = \frac{1}{2}(p^2 + \omega^2 q^2) \ .$$

The phase portrait is as in figure 2, left. If one excludes the origin, then the phase space of the system is foliated into invariant (curves homeomorphic to) circles. The canonical transformation

$$p = \sqrt{2I\omega}\cos\varphi \ , \qquad q = \sqrt{2I/\omega}\sin\varphi \ , \qquad (I,\varphi) \in \mathbb{R}^+ \times S^1 \ , \tag{2.1}$$

introduces coordinates adapted to the foliation: namely I selects the circle while φ is a phase on it. Correspondingly, as we already know, H is turned into $H'(I,\varphi) = \omega I$ and the motion in the new coordinates, see (1.11), appears linear. The coordinates I, φ are called the *action–angle variables* of the system. Due to the square root of the action in (2.1), such variables cannot be used in a neighborhood of the equilibrium point.

(2) The pendulum,

$$H(p,q) = \frac{p^2}{2} - \omega^2 \cos q \ .$$

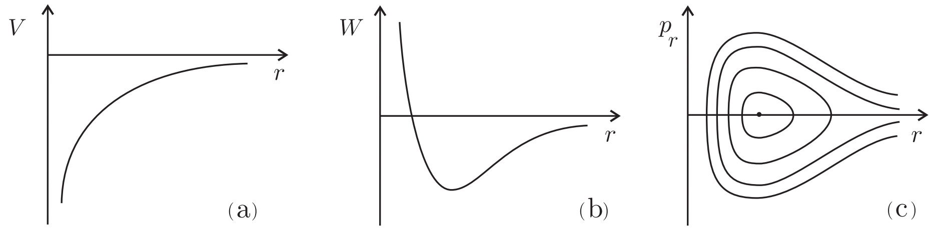

Figure 5: *(a) A typical attracting potential; (b) the corresponding potential W of the reduced system; (c) the corresponding phase portrait (the set of curves* $H_{p_\vartheta} = \mathrm{const}$*).*

The phase portrait is in figure 4. Exclude the two equilibrium points and the separatrices issuing from the unstable one. The remaining phase space is divided in three disjoint regions, each of them being foliated into circles (circles outside the separatrices exploit the periodicity of q). In each region it is possible to introduce action–angle coordinates, that is to introduce a canonical transformation $(p,q) = w(I,\varphi)$ such that

$$H'(I,\varphi) = h(I)$$

and correspondingly

$$I(t) = I^o \ , \qquad \varphi(t) = \varphi^o + \omega(I^o)t \ , \qquad \omega = \frac{\partial h}{\partial I} \ . \tag{2.2}$$

For the motions inside the separatrices (oscillating pendulum) the good choice of the new coordinates turns out to be

$$I(p,q) = \frac{1}{2\pi} A(p,q) \ , \qquad \varphi(p,q) = 2\pi \frac{\tau(p,q)}{T(p,q)} \ , \tag{2.3}$$

where $A(p,q)$ is the area inside the curve through (p,q), while $\tau(p,q)$ is the time necessary to reach (p,q) from the positive p axis and $T(p,q)$ is the period. The fact that in such coordinates the motion gets linear, is trivial; the canonicity of w is checked, for example, by testing that $\{\varphi, I\} = 1$ (it might be useful to introduce intermediate canonical variables E, τ, where E is the energy and τ is as above). For motions above or below the the separatrices (rotating pendulum), the definition of I, φ is identical to (2.3), up to the fact that $A(p,q)$ is now the area between the curve through (p,q) and the q–axis, in the interval $(-\pi, \pi)$.

(3) The next more interesting example is a point mass in a central force field, in two dimensions; the Hamiltonian is (1.5), V being a typical attractive potential as in figure 5a. We restrict the attention to bounded motions (energy $E < 0$), with angular momentum $p_\vartheta \neq 0$, and denote by M^+, M^- the subsets with, respectively, positive and negative p_ϑ. Quite clearly, p_ϑ is constant and the angle ϑ advances monotonically on a circle, see figure 6. Concerning p_r and

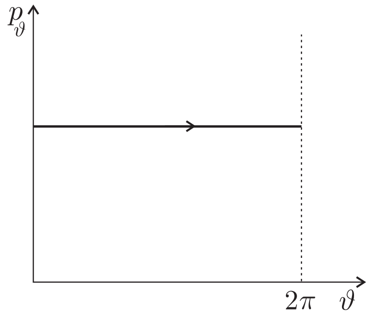

Figure 6: *The variables p_ϑ, ϑ run on a circle.*

r, for each fixed p_ϑ their motion is determined by the one–degree of freedom Hamiltonian

$$H_{p_\vartheta}(p_r, r) = \frac{p_r^2}{2} + W_{p_\vartheta}(r) \ , \qquad W_{p_\vartheta}(r) = V(r) + \frac{p_\vartheta^2}{2r^2} \ .$$

The shape of W, for V as above,[2] is in figure 5b, while the corresponding phase portrait is in figure 5c. For negative energy the trajectories of the reduced system are circles, with the exception of the equilibrium where, as for the harmonic oscillator and the pendulum, circles shrink to a point. Coming back to the original two–dimensional system, it turns out that the set of motions with given momentum p_ϑ and given energy $E < 0$ form a two dimensional torus, actually the Cartesian product of the two circles. So, $M^\pm$ are foliated into tori $\mathbb{T}^2$, the foliation being singular in correspondence with circular motions: tori $\mathbb{T}^2$ there shrink to circles. A family of tori — say those with given p_ϑ and variable energy — is represented in figure 7. The periods T_r and T_ϑ of the motions of, respectively, (p_r, r) and (p_ϑ, ϑ) on their circles, depend on p_r and E, and for most choices of V it turns out that the matrix $\frac{\partial(T_r, T_\vartheta)}{\partial(p_\vartheta, E)}$ has nonvanishing determinant, so that on most tori the periods are incommensurable, all trajectories fill tori densely, and a finer decomposition is impossible. That is: *the foliation into tori $\mathbb{T}^2$ is intrinsically defined by the dynamics.* Well known exceptions are the Kepler potential $V \sim r^{-1}$ and the harmonic potential $V \sim r$, for which T_r and T_ϑ are always equal, and so all (bounded) motions are periodic: tori $\mathbb{T}^2$ are futher decomposed into circles. For such systems, a closer inspection to the structure of the phase space is mandatory (the image of the phase space as foliated into tori $\mathbb{T}^2$ turns out to be misleading). Motions on tori, if projected onto the coordinate space, have the well known shape outlined in figure 8: r oscillates between a minimal and maximal value, corresponding respectively to the "perihelion" and the "aphelion" (or "pericenter" and "apocenter"), with period T_r, and during such time the perihelion advances by some angle Θ (ϑ advances by $2\pi + \Theta$). In general Θ is incommensurate with 2π and orbits do

[2]To have such a shape, it is necessary that $-V(r) \ll r^{-2}$ for small r and $-V(r) \gg r^{-2}$ for large r.

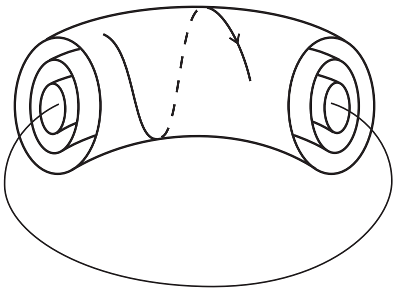

Figure 7: *A family of tori for the central motions.*

not close.

Inside $M^{\pm}$ it is possible to introduce action–angle variables $(I, \varphi) = (I_1, I_2, \varphi_1, \varphi_2)$, in such a way that the new Hamiltonian H' depends only on the actions,

$$H'(I_1, I_2, \varphi_1, \varphi_2) = h(I_1, I_2) \ ,$$

and the motion appears linear:

$$I_j(t) = I_j^o \ , \qquad \varphi_j(t) = \varphi_j^o + \omega_j(I_1^o, I_2^o)t \ , \qquad j = 1, 2 \ ,$$

with $\omega_j = \frac{\partial h}{\partial I_j}$. As a matter of fact, the good choice of coordinates turns out to be

$$\begin{aligned}
I_1(p_\vartheta, p_r, \vartheta, r) &= \frac{1}{2\pi} A(p_r, p_\vartheta, r) \\
I_2(p_\vartheta, p_r, \vartheta, r) &= p_\vartheta \\
\varphi_1(p_\vartheta, p_r, \vartheta, r) &= \frac{2\pi}{T_r} \tau(p_\vartheta, p_r, \vartheta, r) \\
\varphi_2(p_\vartheta, p_r, \vartheta, r) &= \Psi + \frac{\Theta}{T_t} \tau(p_\vartheta, p_r, \vartheta, r) \ ,
\end{aligned}$$

where $A(p_r, p_\vartheta, r)$ is the area inside the (closed) curve $H_{p_\vartheta} = \text{const}$ passing through (r, p_r) in figure 5c, while $\tau(p_\vartheta, p_r, \vartheta, r)$, see figure 8, is the time necessary to reach $(p_\vartheta, p_r, \vartheta, r)$ from the last perihelion (in turn proportional, for Kepler's second law, to the shadowed area in figure 8) and Ψ is the angle from some fixed direction to the last perihelion; φ_1 and φ_2 are called, respectively, the *mean anomaly* and the *anomaly of the perihelion.* In the Kepler case it is $\Theta = 0$, so all orbits close and the angle φ_2 is an additional constant of motion. Correspondingly h depends only on I_1, more precisely it is $h = -c/I_1^2$, $c > 0$.

(4) Finally, a few words on a system with three degrees of freedom, namely the Lagrange top (the symmetric rigid body with a fixed point in the gravity). The system has three independent constants of motion, namely the energy and, due

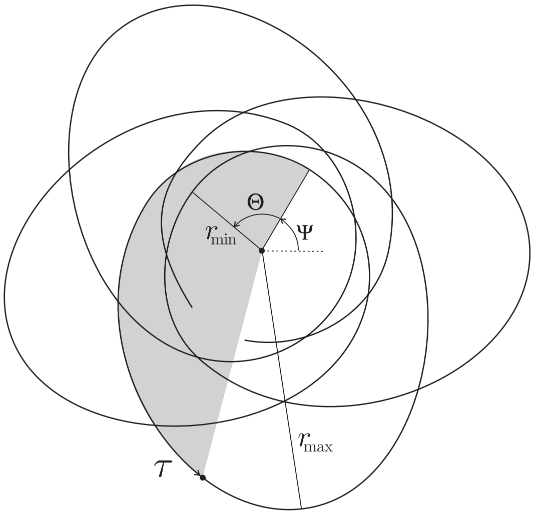

Figure 8: *A typical trajectory of a central motion.*

to the symmetries of the problem, the components of the angular momentum in the direction of the gravity and in the direction of the symmetry axis of the body. Taking the z–axis as the direction of the gravity and using as coordinates the Euler angles, the Hamiltonian turns out to be

$$H(p_\vartheta, p_\varphi, p_\psi, \vartheta, \varphi, \psi) = \frac{p_\vartheta^2}{2A_1} + \frac{(p_\varphi - p_\psi \cos\vartheta)^2}{2A_1 \sin^2\vartheta} + \frac{p_\psi^2}{2A_3} + c\cos\vartheta \ , \qquad c > 0 \ ,$$

and the conserved quantities are H, p_φ, p_ψ. A detailed description of the system can be found, for example, in [3]. We limit ourselves to observe that, generically, motions take place on three–dimensional invariant tori. Indeed, for given p_φ and p_ψ the behavior of p_ϑ, ϑ is determined by the one–degree of freedom Hamiltonian

$$\begin{aligned} H_{p_\varphi,p_\psi}(p_\vartheta, \vartheta) &= \frac{p_\vartheta^2}{2A_1} + W_{p_\varphi,p_\psi}(\vartheta) \ , \\ W_{p_\varphi,p_\psi}(\vartheta) &= \frac{(p_\varphi - p_\psi \cos\vartheta)^2}{2A_1 \sin^2\vartheta} + \frac{p_\psi^2}{2A_3} + c\cos\vartheta \ , \end{aligned}$$

and generically (excluding equilibria and separatrices) motions are on circles. On the same time, the pairs (p_φ, φ) and (p_ψ, ψ) also run on circles, in general with nonvanishing average velocity. The product of the three circles gives an invariant torus $\mathbb{T}^3$. The foliation in tori has several singularities, among them the equilibria of the reduced system, which are uniform precessions of the complete system with constant $\dot\varphi$, $\dot\psi$. The passage to action–angle variables for such a system is discussed in [43], [44]. In the familiar case of a fast rotating top, the three circular motions composing the torus $\mathbb{T}^3$ roughly correspond to the three well known motions of the top, called precession, nutation and proper rotation (see figure 9).

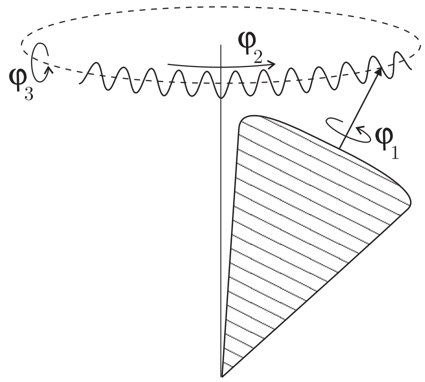

Figure 9: *The motion of the fast top. (The identification of the angles is only approximate.)*

2.2 The Liouville–Arnol'd theorem

The feature common to these examples is the presence of a foliation of the phase space into n–dimensional tori, n being the number of the degrees of freedom, on which the motions are quasi periodic. Moreover, there exist canonical coordinates (I,φ) adapted to the foliation, such that the actions identify the torus and stay constant, while the angles are coordinates on the torus and advance uniformly in time. Correspondingly, the new Hamiltonian depends only on the actions. The new coordinates are defined in domains of the form $B\times\mathbb{T}^n$, B being some open ball in $\mathbb{R}^n$.

Roughly speaking, the Liouville–Arnol'd theorem, that we are going to introduce, states that such a picture holds whenever a system with n degrees of freedom has n independent constants of motions $F_1,\ldots,F_n$, such that their Hamiltonian flows pairwise commute, namely for all i,j

$$\Phi^t_{F_i}\circ\Phi^s_{F_j}=\Phi^s_{F_j}\circ\Phi^t_{F_i}\qquad\forall s,t\in\mathbb{R}\ .$$

A necessary and sufficient condition for this, in turn, is that all Poisson brackets $\{F_i,F_j\}$ vanish [3], and this is the assumption commonly entering the theorem. A possible statement is the following:

PROPOSITION 2.1 (Liouville–Arnol'd theorem) *Let D be an open subset of $\mathbb{R}^n$ and assume there exist n functions $F_1,\ldots,F_n:D\to\mathbb{R}$, such that $\{F_i,F_j\}=0$ in D for all i,j. Assume $F_1,\ldots,F_n$ are linearly independent, i.e.*

$$\mathrm{Rk}\left(\frac{\partial(F_1,\ldots,F_n)}{\partial(p_1,\ldots,p_n,q_1,\ldots,q_n)}\right)=n\ ,\tag{2.4}$$

on a level set

$$\Sigma_c=\{(p,q)\in D:F_i(p,q)=c_i\}\ ,\qquad c=(c_1,\ldots,c_n)\in\mathbb{R}^n\ ,$$

so that Σ_c is an n–dimensional submanifold of D. Further assume that Σ_c is compact (or contains a compact component, and restrict the attention to it). Then:

i. *Σ_c is diffeomorphic to the n–dimensional torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$.*

ii. *There exists a neighborhood $\mathcal{C}$ of c, such that the set*

$$D_{\mathcal{C}} = \bigcup_{c' \in \mathcal{C}} \Sigma_{c'}$$

("tubolar nbhd of Σ_c") is diffeomorphic to $\mathcal{C} \times \Sigma_c$, and in $D_{\mathcal{C}}$ there exist action–angle variables, namely there exists a canonical transformation $(p,q) = w(I,\varphi)$, defined for $I \in B \subset \mathbb{R}^n$ and $\varphi \in \mathbb{T}^n$, $D_{\mathcal{C}} = w(B \times \mathbb{T}^n)$, such that the actions are constant on tori and, conversely, each pull–back $F'_i = F_i \circ w$ depends only on the actions:

$$F'_i(I,\varphi) = f_i(I) \ , \qquad i = 1,\ldots,n \ .$$

The use of the theorem for Hamiltonian dynamics is immediate. Indeed if $H = F_1$ is the Hamiltonian we are interested in, and $F_2,\ldots,F_n$ are independent constants of motion for H, which are pairwise in involution, then the theorem applies, and consequently:

The level sets Σ_c of the n constants of motions, if (2.4) holds on them, are invariant tori $\mathbb{T}^n$, and around any of them the phase space is foliated into invariant tori $\mathbb{T}^n$. In a neighborhood of Σ_c there exist action–angle coordinates, such that the new Hamiltonian $H' = H \circ w$ depends only on the actions,

$$H'(I,\varphi) = h(I) \ ,$$

and correspondingly the actions are constant while the angles advance linearly in time. The motion on each torus is quasi–periodic.

Hamiltonian systems satisfying the assumptions of the Liouville–Arnol'd theorem are called *Liouville–integrable,* or simply integrable.

Remark. Among integrable systems, there is special class of systems which, in addition to the n constants of motion in involution which are necessary to accomplish the assumptions of the Liouville–Arnol'd theorem, have some extra constants of motion. An important example is the Euler–Poinsot rigid body (the rigid body with a fixed point, in absence of external torques), which has three degrees of freedom but four constants of motion, namely the Hamiltonian H and (the three components of) the angular momentum vector m. A good choice of the constants of motion, in view of the Liouville–Arnol'd theorem, is H, $\|m\|$, m_z, $m_{z'}$, where m_z and $m_{z'}$ denote the components of m along any two different directions, and it is easy to check that $(H,\|m\|,m_z)$, as well as $(H,\|m\|,m_{z'})$, are triples in involution. A similar example is the central motion in three dimensions: in this case too, the system has three degrees of freedom but four constants of motion, actually the same quantities as in the previous example; in the particular case of the Kepler potential there is a further constant of motion, namely, as already commented, the anomaly of the perihelion in the plane of the motion. Such systems are called *superintegrable,* or also *properly degenerate* integrable systems. In section 6 we shall study in some detail one of them, namely the Euler–Poinsot rigid body. We shall see, in that example, that the Arnold–Liouville theorem, though applicable, does not provide a deep description of degenerate systems and can be even misleading.

Remark. It should be stressed that while the angles are global, i.e. run on $\mathbb{T}^n$, instead the actions in general are only local. Obvious obstructions to the existence of global action–angle coordinates, are the singularities of the foliation into invariant tori, which are found in connection with level sets Σ_c on which condition (2.4) is not satisfied (an example is the case of circular orbits in the central motion). But even inside a connected region of the phase space which is free of singularities, global action–angle coordinates in general do not exist, due to topological obstructions (more or less as on a manifold, though regular, global coordinates in general do not exist). Such local character of the actions should be taken into account in the study of perturbed systems: quite clearly, if the output of the perturbation study is the constancy of all actions (KAM and Nekhoroshev for non–degenerate systems, see section 5), then it is possible to work consistently inside a single chart of action–angle coordinates, and no difficulty arises. If instead one or more actions are out of control, then it is not possible to work locally in the actions, and to carry on the perturbation study it is mandatory to have in mind the global structure of the phase space. As we shall see in section 6, this is precisely the case of properly degenerate systems.

Remark. The choice of the action–angle variables is not univoque. If $(I,\varphi) \in B \times \mathbb{T}^n$ are action–angle variables for a system, then the change of variables

$$I = AI' \ , \qquad \varphi = A^{-T}\varphi' \ , \tag{2.5}$$

where A is an integer matrix such that $|\det A| = 1$, is canonical, and the new coordinates (I',φ') are also action–angle variables: namely they are defined in a domain of the form $B' \times \mathbb{T}^n$ and I is a function only of I', so that any integrable Hamiltonian $h(I)$ saves the integrable form. Integer matrix means a matrix with integer entries. In addition to (2.5), translations of the form

$$I = I' + I^* \ , \qquad \varphi = \varphi' + \varphi^*(I') \tag{2.6}$$

(I^* needs to be a constant, while φ^* can depend on I') also turn action–angle into action–angle variables. It could be seen that the combination of (2.5) and (2.6) exhaust all possibilities.

The property that A is integer with dererminant one (equivalently: both A and A^{-1} are integer) is necessary and sufficient in order for $A : \mathbb{T}^n \to \mathbb{T}^n$ to be a diffeomorphism (in both direction the 2π–periodicity of the angles is saved).

2.3 Nearly integrable systems

Nearly integrable Hamiltonian systems are systems of the form

$$H(I,\varphi) = h(I) + \varepsilon f(I,\varphi) \ , \tag{2.7}$$

where ε is a small parameter. The study of such systems plays a crucial role in Hamiltonian mechanics; Poincaré himself, at the beginning of his *Mèthodes Nouvelles de la Méchanique Céleste,* (Vol. 1, Chapter 1) introduces such a study as a very basic one, using for them the expression *"problème général de la dynamique"*. From the Hamilton equations corresponding to (2.7),

$$\dot I = -\varepsilon \frac{\partial f}{\partial \varphi}(I,\varphi) \ , \qquad \dot\varphi = \omega(I) + \varepsilon \frac{\partial f}{\partial I}(I,\varphi) \ ,$$

the *"a priori"* estimate

$$\|I(t) - I^o\| < (\mathrm{const})\, \varepsilon\, t \tag{2.8}$$

immediately follows. Such an estimate is however pessimistic: indeed, denoting by $\langle\, . \,\rangle$ the average on φ, it is

$$\Big\langle \frac{\partial f}{\partial \varphi} \Big\rangle = 0 \ ,$$

and since for small ε the angles run much faster than the actions, there is a good chance for the presence of compensations in $\dot I$, with possibly better estimates of the form

$$\|I(t) - I^o\| \ll 1 \qquad \text{for} \quad |t| \le 1/\varepsilon \ .$$

The question we are facing, in fact, is not strictly Hamiltonian, and the rules of the game are perhaps better understood in a more general framework. To this end, let $(I,\varphi) \in \mathbb{R}^m \times \mathbb{T}^n$ and assume the equations of motion have the form

$$\dot I = \varepsilon F(I,\varphi) \ , \qquad \dot\varphi = \omega(I) + \varepsilon G(I,\varphi) \ , \tag{2.9}$$

with some F, G. Denote $\mathcal{F}(I) = \langle F(I,\,.\,)\rangle$ and consider the "averaged equation"

$$\dot J = \varepsilon \mathcal{F}(J) \ .$$

The hope, also called "averaging principle", is that such a simplified equation represents rather well the evolution of the actions. More precisely:

Definition 2.2 *We say that the system (2.9) satisfies the averaging principle, if for any initial datum (I^o, φ^o), taking $J(0) = I^o$, it is*

$$\|I(t) - J(t)\| \quad \stackrel{\varepsilon\to 0}{\longrightarrow} \quad 0 \qquad \text{for} \quad |t| \le 1/\varepsilon \ . \tag{2.10}$$

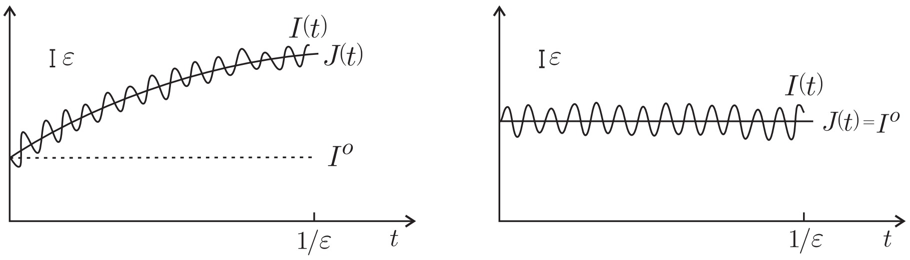

Figure 10: *Illustrating the averaging principle, in the general case (left) and in the Hamiltonian case (right).*

In the Hamiltonian case it is trivially $J(t) = I^o$ and correspondingly the averaging principle claims that the actions stay almost constant for the whole time scale ε^{-1}. Figure 10 illustrates the averaging principle, both in the general and in the Hamiltonian case. Let us stress that the averaged equation is *much* simpler than the original system (2.9), not only because there are less variables, but mainly because for the averaged equation the time ε^{-1} is a natural unit of time, i.e., it is a short time, while in the system (2.9), due to the equations governing the angles, the time unit is one and thus ε^{-1} is a long time.

Remark. The time scale ε^{-1} appearing in (2.10) is the minimal one for which the question is nontrivial: for shorter time scales, thanks to the *a priori* estimate (2.8), both $I(t)$ and $J(t)$ stay close to I^o. It is of course interesting, and possible in several cases, to look at larger time scales: for example polynomial in ε, $|t| \le \varepsilon^{-r}$, or exponential, $|t| \le \exp 1/\varepsilon$, or perpetual, $|t| < \infty$.

It is not difficult to find systems, Hamiltonian or general, for which the averaging principle is satisfied, as well as systems for which it is not. Let us produce two easy Hamiltonian examples.

Example: Let $n = 2$,

$$H(I,\varphi) = \omega \cdot I + \varepsilon \cos(\varphi_2 - \varphi_1) \ .$$

The general solution is $\varphi_i(t) = \varphi_i^o + \omega_i t$, $i = 1, 2$, and

$$I_1(t) = I_1^o + \varepsilon \mathfrak{I}(t) \ , \qquad I_2(t) = I_2^o - \varepsilon \mathfrak{I}(t) \ ,$$

with

$$\mathfrak{I}(t) = \frac{\cos(\varphi_2(t) - \varphi_1(t))}{\omega_2 - \omega_1}$$

for $\omega_1 \neq \omega_2$ and

$$\mathfrak{I}(t) = t \ \sin(\varphi_1^o - \varphi_2^o)$$

for $\omega_1 = \omega_2$. So the principle is satisfied (and moreover, the actions are perpetually stable) for $\omega_1 \neq \omega_2$, while it is not for $\omega_1 = \omega_2$.

Example: Let $n = 2$,

$$H(I, \varphi) = \frac{I_1^2}{2} - \frac{I_2^2}{2} + \varepsilon \sin(\varphi_2 - \varphi_1) \ . \tag{2.11}$$

The behavior is nontrivial and the general solution cannot be written. But there exists a family of motions which does not satisfy the averaging principle, namely

$$I_1(t) = I_2(t) = I^o + \varepsilon t \ , \qquad \varphi_1(t) = -\varphi_2(t) = \varphi^o + I^o t + \tfrac{1}{2}\varepsilon\, t^2 \ .$$

We shall come back on this example in section 4.

2.4 Lindstedt method versus Von Zeipel method

There are basically two methods to start the study of nearly integrable systems. A very elementary idea is to correct solutions and write, for example,

$$I(t) = I_0(t) + \varepsilon I_1(t) + \cdots \ , \qquad \varphi(t) = \varphi_0(t) + \varepsilon \varphi_1(t) + \cdots \ , \tag{2.12}$$

with of course (in the Hamiltonian case) $I_0(t) = I^o$, $\varphi_0(t) = \varphi^o + \omega(I^o)t$; inserting such expressions into the equations of motion gives rise to a hyerarchy of equations, which in principle determine the deverlopment (2.12) at any order. Such a procedure is called *Lindstedt method* and series (2.12) are known as *Lindstedt series.* Though conceptually simple, however, the method is not so easy to use in practice (at least not to obtain mathematically rigorous results), due to the huge amount of terms which are rapidly generated by raising the order. As a matter of fact, Lindstedt series were soon abandoned in rigorous studies. Only recently, after the pioneering work [38], the method has been revisited, and succesfully implemented, in connection with KAM theorem and splitting of separatrices [50] [51] [33], for which it appeared to perform excellently. For an elementary physical application see [12]. We shall not discuss Lindstedt's method in these lectures.

An alternative idea to study nearly integrable systems, looking for conditions such that the averaging principle is satisfied, is to produce small canonical transformations, i.e. transformations near to the identity, which move the perturbation to higher orders in ε. A near the identity transformation is a transformation of the form $(I, \varphi) = w_\varepsilon(I', \varphi')$, with

$$w_\varepsilon(I', \varphi') = (I', \varphi') + \varepsilon \mathcal{W}(I, \varphi, \varepsilon) \tag{2.13}$$

and small ε. The idea is to look for $\mathcal{W}$ such that the new Hamiltonian $H' = H \circ w_\varepsilon$ is "more integrable" than H, say is of the form

$$H'(I', \varphi') = h(I') + \varepsilon g(I') + \varepsilon^2 f'(I', \varphi') \ . \tag{2.14}$$

The consequence is immediate (see figure 11): by applying the *a priori* estimate to H', we get $\|\dot{I}'\| < (\text{const})\varepsilon^2$ and so

$$\|I'(t) - I'(0)\| < (\text{const})\varepsilon \qquad \text{for} \quad |t| \le 1/\varepsilon \ .$$

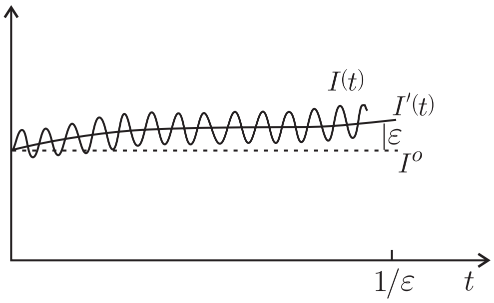

Figure 11: *Illustrating the Von Zeipel method.*

But since w_ε is ε–near the identity, at any time I and I' stay close, so that

$$\|I(t) - I^o\| < (\text{const})\varepsilon \qquad \text{for} \quad |t| \leq 1/\varepsilon \ ,$$

in agreement with the averaging principle. There is also hope that the method can be iterated and the perturbation is pushed to higher orders in ε, with the result that the actions are stable for longer times. Such a method, called *Von Zeipel method,* is very widely used. It was implemented in a variety of slightly different ways and led, in particular, to celebrated results like the KAM and Nekhoroshev theorems.

2.5 Generating near the identity canonical transformations

To implement the above illustrated program, we must be able to produce, in some efficient way, near the identity canonical transformations. Two methods are widely used in the literature: one is based on the so–called "mixed variable generating functions", the other on the Lie series that we already mentioned in section 1.2.

A. Mixed variables generating functions. Given a function

$$S(p', q) \ , \qquad S : \mathbb{R}^{2n} \to \mathbb{R} \ ,$$

such that

$$\det\left(\frac{\partial^2 S}{\partial p'_i \partial q_j}\right) \neq 0 \ , \tag{2.15}$$

the equations

$$p = \frac{\partial S}{\partial q}(p', q) \ , \qquad q' = \frac{\partial S}{\partial p'}(p', q) \ , \tag{2.16}$$

define implicitly a local diffeomorphism $(p, q) = w(p', q')$.

PROPOSITION 2.3 *The transformation w is canonical.*

PROOF. Quite trivially, from (2.16) it follows $\mathrm{d}S = p \cdot \mathrm{d}q + q' \cdot \mathrm{d}p'$. By differentiating, one then gets $\mathrm{d}p \wedge \mathrm{d}q + \mathrm{d}q' \wedge \mathrm{d}p' = 0$, so that (1.15) is satisfied. □

The function $S(p', q) = p' \cdot q$ generates the identity, while functions of the form

$$S_\varepsilon(p', q) = p' \cdot q + \varepsilon\, \mathcal{S}(p', q, \varepsilon) \tag{2.17}$$

generate transformations which are ε–near the identity (for small ε, condition (2.15) is automatically satisfied).

PROPOSITION 2.4 *Any near the identity canonical transformation (2.13) admits a generating function of the form (2.17).*

The proof is an inversion of the previous one, and is left as an exercise. The proposition is interesting in connection with non–existence theorems.

B. The Lie method. For any given $\chi : \mathbb{R}^{2n} \to \mathbb{R}$, the Hamiltonian flow Φ^t_χ at any fixed time t can be regarded as a transformation of variables:

$$(p, q) = \Phi^t_\chi(p', q') \ , \tag{2.18}$$

in which the new variables p', q' play the role of initial data. The following proposition is among the classical ones in Hamiltonian mechanics:

PROPOSITION 2.5 *The transformation (2.18) is canonical.*

The proof can be found in any textbook of theoretical mechanics. Near to the identity canonical transformations are produced by either taking $t = \varepsilon$ small, or equivalently by taking $t = 1$ and χ small, say replacing χ by $\varepsilon\chi$:

$$(p, q) = \Phi^\varepsilon_\chi(p', q') = \Phi^1_{\varepsilon\chi}(p', q') \ .$$

This method *does not* produce all near the identity canonical transformations, but produces sufficiently many of them, and turns out to be very practical to transform functions. The transformed function could be expressed in series, as in (1.13). Here however we prefer to use expansions with remainders and write, for any function F, either

$$F \circ \Phi^1_{\varepsilon\chi} = F + \varepsilon \mathcal{R}_1[\chi, F] \ , \qquad \mathcal{R}_1[\chi, F] = \int_0^1 (L_\chi F) \circ \Phi^t_{\varepsilon\chi} \mathrm{d}t \ , \tag{2.19}$$

or

$$F \circ \Phi^1_{\varepsilon\chi} = F + \varepsilon L_\chi F + \varepsilon^2 \mathcal{R}_2[\chi, F] \ , \qquad \mathcal{R}_2[\chi, F] = \int_0^1 \mathrm{d}t \int_0^t (L^2_\chi F) \circ \Phi^{t'}_{\varepsilon\chi} \mathrm{d}t' \ ; \tag{2.20}$$

we recall that $L_\chi = \{\,\cdot\,, \chi\} = \frac{\mathrm{d}\,\cdot}{\mathrm{d}t}$. Both (2.19) and (2.20) are in fact trivial (but useful) identities.

C. The fundamental equation of perturbation theory. As is quite remarkable, both methods, when applied to a nearly integrable Hamiltonian system in order to to push the perturbation to order ε^2, lead exactly to the same differential equation for, respectively, $\mathcal{S}$ and χ. This is an important point and we shall look at it in detail.

Let $S(I',\varphi) = I'\cdot\varphi + \varepsilon\mathcal{S}(I',\varphi)$, so that

$$I = I' + \varepsilon\frac{\partial\mathcal{S}}{\partial\varphi}(I',\varphi)\ , \qquad \varphi' = \varphi + \varepsilon\frac{\partial\mathcal{S}}{\partial I'}(I',\varphi)\ .$$

A trivial inversion, rough but sufficient to our purposes, gives

$$\varphi = \varphi' + \mathcal{O}(\varepsilon)\ , \qquad I = I' + \varepsilon\frac{\partial\mathcal{S}}{\partial\varphi}(I',\varphi') + \mathcal{O}(\varepsilon^2)\ .$$

By substitution in $H(I,\varphi) = h(I) + \varepsilon f(I,\varphi)$ we then immediately find

$$H'(I',\varphi') = h(I') + \varepsilon\Big(\omega(I')\cdot\frac{\partial\mathcal{S}}{\partial\varphi}(I',\varphi') + f(I',\varphi')\Big) + \mathcal{O}(\varepsilon^2)\ .$$

So, H' gets the wanted form (2.14) if there exist $\mathcal{S}$ and g such that

$$\omega(I')\cdot\frac{\partial\mathcal{S}}{\partial\varphi}(I',\varphi') + f(I',\varphi') = g(I')\ .$$

The function g, in fact, is not free: since $\frac{\partial\mathcal{S}}{\partial\varphi}$ has zero average in the angles, it is necessarily $g = \langle f\rangle$, so the equation for $\mathcal{S}$, dropping the primes, reads

$$\omega(I)\cdot\frac{\partial\mathcal{S}}{\partial\varphi}(I,\varphi) = -f(I,\varphi) + \langle f\rangle\ . \tag{2.21}$$

Let us now come to the Lie method. Using (2.20) to transform h and (2.19) to transform f, we immediately find

$$H' = h + \varepsilon\left(\{h,\chi\} + f\right) + \mathcal{O}(\varepsilon^2)\ ,$$

and since $\{h,\chi\} = -\omega\cdot\frac{\partial\chi}{\partial\varphi}$, the equation for χ turns out to be

$$\omega(I)\cdot\frac{\partial\chi}{\partial\varphi}(I,\varphi) = f(I,\varphi) - \langle f\rangle\ . \tag{2.22}$$

This is the same equation as before, after the identification $\chi = -\mathcal{S}$. *We are in front of a substantial fact,* and equation (2.21), or (2.22), should be considered as the basic equation of Hamiltonian perturbation theory. The equation is nontrivial, and as we shall see in the next subsection, if we try to solve it we are immediately confronted with a basic difficulty, namely the presence of the "small divisors". The difficulty is sometimes called "Poincaré difficulty", since it is at the basis of a celebrated theorem by Poincaré on the generic non–existence of constants of motion in perturbed systems. (We however will not discuss Poincaré's non-existence theorem; see [87] and, for generalizations, [47] [20]).

2.6 The Poincaré difficulty

An easy proposition states that "generically" the fundamental equation cannot be solved.

PROPOSITION 2.6 (Poincaré) *Let*

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

and assume:

i) h is non degenerate, more precisely

$$\det\left(\frac{\partial^2 h}{\partial I \partial I}\right) \neq 0 \tag{2.23}$$

in an open subset B_0 of B.

ii) f has essentially full Fourier series, more precisely, denoting

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

for any $\nu \in \mathbb{Z}^n$ there exists ν' parallel to ν such that $f_{\nu'}(I) \neq 0$ in B_0.

Then a function χ [or $\mathcal{S}$] solving (2.22) [or (2.21)] in B_0 does not exist.

So, generically, the perturbation cannot be pushed to order ε^2, no matter how small B_0 is taken (recall that, according to Proposition 2.4, the existence of $\mathcal{S}$ would be necessary for this).

PROOF. Denote $\chi(I,\varphi) = \sum_\nu \chi_\nu(I) e^{i\nu\cdot\varphi}$, so that

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

In order for (2.22) to be satisfied in B_0, it would be necessary that everywhere in B_0

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

But in the above assumptions, not all of these equations can be solved by writing, as it would appear natural,

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

Indeed, for assumption i), in a dense subset $B_0' \subset B_0$ it is $i\nu\cdot\omega(I) = 0$ for at least one ν, and thus for all $\nu' \parallel \nu$, while by assumption ii) not all of the $f_{\nu'}$ vanish. $\square$

Remark. As is know, relations of the form

$$\nu \cdot \omega = 0 \ , \qquad \nu \in \mathbb{Z}^n \setminus \{0\} \ , \tag{2.24}$$

are called "resonances" of ω. If there are d, but not $d+1$, independent vectors $\nu^{(1)}, \ldots, \nu^{(d)} \in \mathbb{Z}^n$ such that (2.24) is satisfied, then ω is said to be d times resonant and d is called the *multiplicity* of the resonance. If (2.23) is satisfied, then the frequencies can be taken (locally) as independent variables; it is then evident that arbitrarily close to any point of the phase space there are points where some resonance condition (2.24) is realized. More precisely,

LEMMA 2.7 *Let Ω be any ball in $\mathbb{R}^n$ (thought of as the space of frequencies). For $0 \le d \le n-1$, the subset $\Omega^{(d)}$ of Ω where exactly d independent resonance relations of the form (2.24) are satisfied, is dense in Ω. If (2.23) is satisfied, the subset $B_0^{(d)}$ of B_0 where exactly d independent resonance relations are satisfied, is dense in B_0.*

The proof is an easy exercise. For any $\omega \in \mathbb{R}^n$, the set of integer vectors that resonate with ω,

$$\mathcal{L} = \{\nu \in \mathbb{Z}^n : \nu \cdot \omega = 0\} \ ,$$

is a subgroup of $\mathbb{Z}^n$ called the *resonant lattice* or also the *resonant modulus* of ω, and the multiplicity d of the resonance is the dimension of $\mathcal{L}$. For example, if

$$\omega = (0, \ldots, 0, \omega_{d+1}, \ldots, \omega_n) \tag{2.25}$$

then

$$\mathcal{L} = \{\nu \in \mathbb{Z}^n : \nu = (\nu_1, \ldots, \nu_d, 0, \ldots, 0)\} \ . \tag{2.26}$$

The example is in a sense generic: indeed,

LEMMA 2.8 *Let $\mathcal{L} \in \mathbb{Z}^n$ be a resonant lattice of dimension d. There exists a change of coordinates of the form (2.5), with A integer and* $\det A = 1$*, such that the new frequency $\omega' = A^{-T}\omega$ and the new resonant lattice $\mathcal{L}' = A^{-1}\mathcal{L}$ are respectively as in (2.25) and as in (2.26).*

The proof is easy, though a little tricky (see for example [26]). We shall occasionally profit of this lemma in the next sections.

Resonances and small denominators are the main trouble in perturbation theory, though not the only one. The meaning of the above proposition is that, because of resonances, there is no "easy way" to develop Hamiltonian perturbation theory, so as to prove the averaging principle, by naively solving the basic equation. The question however is subtle: as we shall see, in spite of the difficulty raised by the above proposition, it is nevertheless possible to develop in several ways perturbation theory, and work out a variety of results on the stability of the actions (also for exponentially long or infinite times), in such a way to accomplish the averaging principle. But some work, occasionally hard and tricky, needs to be done. In the next lectures we shall play a little with the statement of theorem 2.6, and show that, by suitably weakening its assumptions or its claim, it becomes possible to escape the difficulties, and in fact, any modification gives rise to a possible branch of perturbation theory. In this way we aim to present Hamiltonian perturbation theory in a unitary perspective.

3 Isochronous systems

Perturbed isochronous systems are systems of the general form (2.7), with $h(I) = \omega \cdot I$, so that

$$H(I, \varphi) = \omega \cdot I + \varepsilon f(I, \varphi) \ . \tag{3.1}$$

Thanks to the fact that the unperturbed frequencies are constant, the study of such systems is purely analytic (no "geometric part" is necessary). It is worthwhile to study isochronous systems in detail, besides their intrinsic interest, as the best introduction to the analytic part of perturbation theory. Isochronous systems clearly violate assumption i) of Poincaré theorem.

Remark. Isochronous systems are often introduced as models for systems of weakly coupled harmonic oscillators,[3]

$$H(p, q) = \frac{1}{2} \sum_{j=1}^{n} (p_j^2 + q_j^2) + \varepsilon V(q) \ . \tag{3.3}$$

The passage to the action–angle variables, via

$$p_j = \sqrt{2I_j \omega_j} \, \cos \varphi_j \ , \qquad q_j = \sqrt{2I_j / \omega_j} \, \sin \varphi_j \ ,$$

gives indeed the Hamiltonian (3.3) the form (3.1). But unless V is even, the perturbation f is not analytic in the actions when any of them vanishes, and such a lack of analyticity prevents the use of (I, φ) as coordinates near $I_j = 0$. The difficulty is not serious and since Birkhoff's work it is known how to proceed perturbatively in Cartesian coordinates, for isochronous systems. But we cannot enter the question, so we shall restrict the attention to the case of f analytic. For developments of perturbation theory in Cartesian coordinates, including rigorous estimates, see [52] [53]; for extensions to the non isochronous case, see [45] [81] [62] [86].

[3] A special case is the ε-neighborhood of a stable equilibrium position,

$$H(p, q) = \frac{1}{2} \sum_{j=1}^{n} (p_j^2 + q_j^2) + V_3(q) + V_4(q) + \cdots \ , \tag{3.2}$$

where V_k is a homogeneous polynomial of degree k. Indeed, the homogeneous rescaling $p_j = \varepsilon p'_j$, $q_j = \varepsilon q'_j$, blows a neighborhood of the origin of diameter ε up to a ball of diameter one, and gives the new Hamiltonian $H' = \varepsilon^{-2} H$ the form (3.2), with $V = V_3 + \varepsilon V_4 + \cdots$.

3.1 Perturbations with finitely many Fourier components

To begin with, we consider the most elementary situation, namely the case in which the perturbation f is a Fourier polynomial:

$$f(I,\varphi) = \sum_{\nu\in\mathcal{N}} f_\nu(I)\, e^{i\nu\cdot\varphi} \ ,$$

$\mathcal{N}$ being a finite subset of $\mathbb{Z}^n$ (so, the second assumption of Poincare's theorem is also violated). If there are no resonances, namely

$$\nu\cdot\omega \neq 0 \qquad \forall\nu\in\mathcal{N}\setminus\{0\} \ ,$$

then the basic equation (2.22) is solved by

$$\chi(I,\varphi) = \sum_{\nu\in\mathcal{N}\setminus\{0\}} \frac{f_\nu(I)}{i\nu\cdot\omega}\, e^{i\nu\cdot\varphi} \ ;$$

correspondingly the new Hamiltonian takes the form

$$H'(I',\varphi') = \omega\cdot I' + \varepsilon g(I') + \varepsilon^2 f'(I',\varphi') \ ,$$

with $g = \langle f\rangle$, $f' = \mathcal{R}_2[\chi,h] + \mathcal{R}_1[\chi,f]$.

Assume instead $\mathcal{N} = \mathcal{N}_0 \cup \mathcal{N}_1$, with

$$\begin{aligned} \nu\cdot\omega &= 0 \qquad \text{for } \nu\in\mathcal{N}_0 \\ &\neq 0 \qquad \text{for } \nu\in\mathcal{N}_1 \ . \end{aligned}$$

An useful example is $\omega = (0,\ldots,0,\omega_{d+1},\ldots,\omega_n)$, with correspondingly

$$\mathcal{N}_0 = \big\{\nu\in\mathcal{N} : \nu = (\nu_1,\ldots,\nu_d,0,\ldots,0)\big\} \ . \tag{3.4}$$

In this case it is not possible to eliminate the resonant harmonics $\nu\in\mathcal{N}_0$, and the best one can do is taking

$$\chi(I,\varphi) = \sum_{\nu\in\mathcal{N}_1} \frac{f_\nu(I)}{i\nu\cdot\omega}\, e^{i\nu\cdot\varphi} \ ;$$

the new Hamiltonian has then the form

$$H'(I',\varphi') = \omega\cdot I' + \varepsilon g(I',\varphi') + \mathcal{O}(\varepsilon^2) \ , \tag{3.5}$$

g being the projection of f onto $\mathcal{N}_0$:

$$g(I',\varphi') = \sum_{\nu\in\mathcal{N}_0} f_\nu(I')\, e^{i\nu\cdot\varphi'} \ .$$

In the example, g depends only on the first d angles, that is $g = g(I',\varphi'_1,\ldots,\varphi'_d)$; in such a case, it is easy to conclude that $I'_{d+1},\ldots,I'_n$, and consequently

$I_{d+1}, \ldots, I_n$ which are ε–close to them, evolve only on a time scale $t \sim \varepsilon^{-2}$, while $I'_1, \ldots, I'_d$, and thus $I_1, \ldots, I_d$, are out of control already at $t \sim \varepsilon^{-1}$. In the general case, there are some linear combinations of the actions which stay constant up to $t \sim \mathcal{O}(\varepsilon^{-2})$, while some other linear combinations possibly evolve already at $t \sim \varepsilon^{-1}$. This is obvious if we recall (Subsection 2.6) that the example (3.4) is generic, but is also evident from the Hamilton equation

$$\dot{I}' = -\varepsilon \frac{\partial g}{\partial \varphi'} = \varepsilon \sum_{\nu \in \mathcal{N}_0} c_\nu(I', \varphi')\, \nu \ , \qquad c_\nu(I', \varphi') = -i f_\nu(I')\, e^{i\nu\cdot\varphi'} \ ;$$

this expression shows indeed that the evolution of the actions takes place on the plain parallel to $\mathcal{N}_0$, passing through the initial datum $I'(0)$.

Remark. In the example, the original Hamiltonian has the form

$$H(I, \varphi) = h(I_{d+1}, \ldots, I_n) + \varepsilon f(I, \varphi) \ , \qquad h = \sum_{i>d} \omega_j I_j \ ,$$

that is, the actions $I_1, \ldots, I_d$ enter only the perturbation. It is then easy to find perturbations such that $I_1, \ldots, I_d$ do evolve on the time scale ε^{-1}. Indeed, let f depend only on the first d degrees of freedom,

$$H(I, \varphi) = h(I_{d+1}, \ldots, I_n) + \varepsilon f(I_1, \ldots, I_d, \varphi_1, \ldots, \varphi_d) \ ;$$

then the motion of $I_1, \ldots, I_d, \varphi_1, \ldots, \varphi_d$, on the time scale ε^{-1}, is determined only on the perturbation and can be anything. In the general case, it is easy to find perturbations such that some linear combinations of the actions do evolve on the time scale $t \sim \varepsilon^{-1}$.

3.2 The Siegel arithmetic condition

Next we consider the case of a generic perturbation f having a possibly full Fourier series. In such a case, solving the basic equation (2.22) is more difficult in an essential way. It is *not* sufficient to assume that ω is nonresonant, i.e.

$$\nu \cdot \omega \neq 0 \qquad \forall \nu \in \mathbb{Z}^n \setminus \{0\} \ . \tag{3.6}$$

One needs more, namely that the function χ that solves (2.22),

$$\chi(I, \varphi) = \sum_{\nu \in \mathbb{Z}^n} \chi_\nu(I)\, e^{i\nu\cdot\varphi} \ , \qquad \chi_\nu = \frac{f_\nu(I)}{i\nu \cdot \omega} \ , \tag{3.7}$$

does exist, i.e. that the series in (3.7) converges to a regular function. For this it is necessary that the Fourier coefficients χ_ν decrease sufficiently fast by increasing (the norm of) ν, while according to (3.6) the denominators $\nu \cdot \omega$, though never vanishing, can nevertheless be arbitrarily close to zero for large ν.

Having in mind the exponential estimates, we shall restrict ourselves to the analytic framework; for such a case, a good arithmetic condition on ω was introduced by Siegel [89].

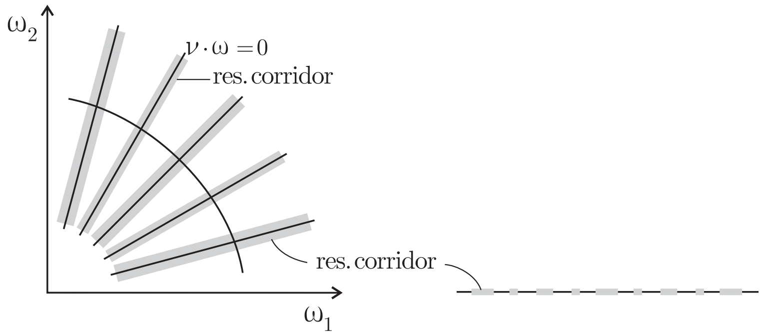

Figure 12: *Resonant lines and resonant corridors, for $n = 2$.*

Definition 3.1 *The frequency ω is said to be (γ, τ)–Diophantine, if there exist real costants $\gamma > 0$ and $\tau > n - 1$ such that*

$$|\nu \cdot \omega| \geq \frac{\gamma}{|\nu|^\tau} \qquad \forall \nu \in \mathbb{Z}^n \setminus \{0\} \ , \tag{3.8}$$

with $|\nu| = |\nu_1| + \cdots + |\nu_n|$.

Condition (3.8) is known as the *Siegel Diophantine condition* or simply *Diophantine condition.* The following lemma holds:

Lemma 3.2 *Let Ω be a ball of radius R in $\mathbb{R}^n$, and let $\Omega_{\gamma,\tau}$ be the subset of Ω where (3.8) holds; denote $\overline{\Omega}_{\gamma,\tau} = \Omega \setminus \Omega_{\gamma,\tau}$.*

i) *If $\tau < n - 1$, then $\Omega_{\gamma,\tau}$ is empty.*

ii) *If $\tau = n - 1$ and γ is sufficiently small, $\gamma < \gamma_0(n)$, then $\Omega_{\gamma,\tau}$ is non denumerable, but for no matter how small γ, it has zero Lebesgue measure.*

iii) *If $\tau > n - 1$ and γ is small, then $\Omega_{\gamma,\tau}$ has large measure, more precisely there exists $C_{n,\tau} > 0$ such that*

$$\mathrm{mes}\, \overline{\Omega}_{\gamma,\tau} < C_{n,\tau} \frac{\gamma}{R} \mathrm{mes}\, \Omega \ .$$

Points i) and ii) are classical results from number theory; point iii) is an easy exercise. [*Hint:* $\mathrm{mes}\, \overline{\Omega}_{\gamma,\tau} \leq \sum_\nu \mathrm{mes}\, \overline{\Omega}_{\gamma,\tau,\nu}$, where $\overline{\Omega}_{\gamma,\tau,\nu}$ is the set where $|\nu \cdot \omega| < \gamma / |\nu|^\tau$; each term of the sum is easily estimated and if assumption iii) is satsfied the series converges with sum proportional to γ].

Remark. For small γ the "bad" set $\overline{\Omega}_{\gamma,\tau}$ is small in measure, but is topologically important, namely is open dense. For $n = 2$, the set is represented in figure 12. The "corridors" form a dense set, namely there is one of them around each line with rational slope ω_2/ω_1, nevertheless their overall measure is small.

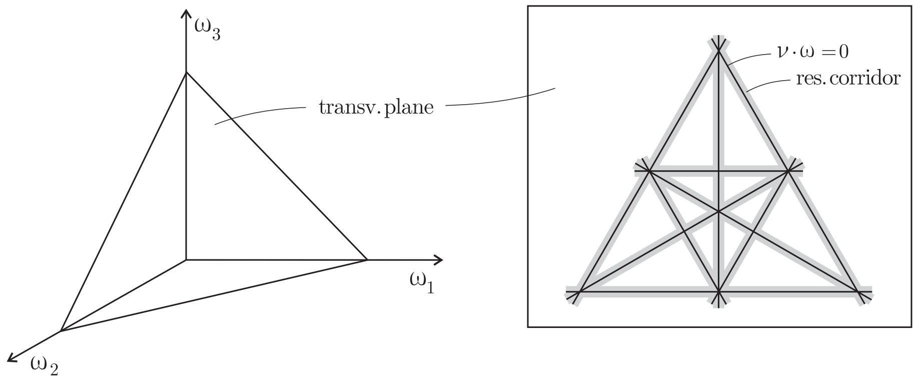

Figure 13: *The Arnol'd web, for $n = 3$.*

If cut by any transversal line, the resulting figure is a Cantor set, with a hole in correspondence with each corridor. For $n > 2$ it is not so easy to draw pictures. Figure 13 refers to $n = 3$: in place of resonant lines one has now resonant planes through the origin, and by sectioning with a transversal plane one obtains a "web" similar to the one represented in the left part of the figure. *For $n \geq 3$ such a web, called the "Arnol'd web", is not only open dense, but also connected.*

The Siegel condition (3.8) is adapted to work on analytic functions of the angles. Indeed, for $\rho > 0$ let $\mathcal{S}_\rho$ denote the strip

$$\mathcal{S}_\rho = \bigl\{\varphi \in \mathbb{C}^n : |\mathrm{Im}\,\varphi_j| < \rho\,,\ j = 1, \ldots, n\bigr\}$$

and consider any real analytic function F: $\mathcal{S}_\rho \to \mathbb{C}$, periodic of real period 2π in each argument (that is, the holomorphic extension of a periodic analytic function: $\mathbb{R}^n \to \mathbb{R}$). We shall denote by $\|F\|_\rho^\infty$ the supremum norm of F in the strip, and by $\|F\|_\rho \geq \|F\|_\rho^\infty$ its "Fourier norm", namely

$$\|F\|_\rho = \sum_{\nu \in \mathbb{Z}^n} |F_\nu|\, e^{\rho|\nu|}\ .$$

The next lemma collects some well known properties of the Fourier series of analytic functions:

Lemma 3.3 *The following properties hold:*

i) If F is analytic in $\mathcal{S}_\rho$, and $\|F\|_\rho^\infty < \infty$, then the amplitude of the Fourier components F_ν decreases exponentially with $|\nu|$, according to

$$|F_\nu| \leq \|F\|_\rho^\infty e^{-\rho|\nu|}\ ;$$

conversely, if there exists $\mathcal{F}$: $|F_\nu| \leq \mathcal{F} e^{-\rho|\nu|}$ for any ν, then F is analytic in $\mathcal{D}_\rho$.

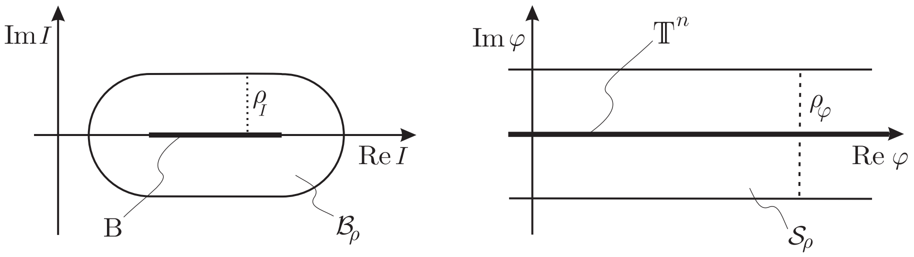

Figure 14: *The domains $\mathcal{B}_\rho$ and $\mathcal{S}_\rho$.*

ii) If F is analytic in any strip larger than $\mathcal{S}_\rho$, then $\|F\|_\rho < \infty$; conversely, if $\|F\|_\rho < \infty$, then F is analytic in $\mathcal{S}_\rho$.

The Fourier norm is especially suited to handle our basic equation (2.22) for χ. Let us complete the definition of domains and norms in the following way: for any $\rho = (\rho_I, \rho_\varphi)$ with positive entries, we shortly denote $\mathcal{S}_\rho = \mathcal{S}_{\rho_\varphi}$ and then, for any ball $B \in \mathbb{R}^n$,

$$\mathcal{B}_\rho = \bigcup_{I \in B} \Delta_\rho(I) \ , \qquad \Delta_\rho(I) = \left\{ I' \in \mathbb{C}^n : |I'_j - I_j| < \rho_I \ \forall j \right\} \ ;$$

finally,

$$\mathcal{D}_\rho = \mathcal{B}_\rho \times \mathcal{S}_\rho \ , \tag{3.9}$$

see figure 14. Concerning norms, for any analytic function $F : \mathcal{D}_\rho \to \mathbb{C}$ we extend the definition of the Fourier norm by adding the supremun on the actions:

$$\|F\|_\rho = \sup_{I \in \mathcal{B}_\rho} \|F(I, .)\|_\rho$$

(for functions depending only on the actions, like the Fourier components, this coincides with the supremum norm). For later use, we also introduce the Fourier norm for vector valued functions: $\mathcal{D}_\rho \to \mathbb{C}^n$. Precisely, for any $\sigma = (\sigma_\rho, \sigma_\varphi) \le \rho$ (the inequality is intended to work separately for the two entries), if

$$w = (w_{I_1}, \ldots, w_{I_n}, w_{\varphi_1}, \ldots, w_{\varphi_n})$$

is an analytic function: $\mathcal{D}_\sigma \to \mathbb{C}^n$, we denote

$$\|w\|_\sigma = \max \left\{ \frac{\|w_{I_1}\|_\sigma}{\rho_I}, \ldots, \frac{\|w_{I_n}\|_\sigma}{\rho_I}, \frac{\|w_{\varphi_1}\|_\sigma}{\rho_\varphi}, \ldots, \frac{\|w_{\varphi_n}\|_\sigma}{\rho_\varphi} \right\} \ .$$

Pay attention that ρ, and not σ, is used to normalize the different components of w. (A more precise notation would be $\|w\|_{\rho,\sigma}$, but for simplicity the dependence of the norm on ρ, which in the concrete use is fixed, is omitted).

LEMMA 3.4 *Let ω be (γ,τ)–Diophantine. If $\|f\|_\rho < \infty$, then χ solving (2.22) is analytic in $\mathcal{D}_\rho$ and in any reduced domain $\mathcal{D}_{(1-\delta)\rho}$, $0<\delta<1$, the following estimate holds:*

$$\|\chi\|_{(1-\delta)\rho} < c_\tau\, \delta^{-\tau} \rho_\varphi^{-\tau} \frac{\|f\|_\rho}{\gamma} \ ,$$

c_τ being a positive constant depending only on τ.

PROOF. Quite obviously, it is

$$\|\chi\|_{(1-\delta)\rho} \le \sum_{\nu\in\mathbb{Z}^n} \|f_\nu\|_\rho e^{(1-\delta)\rho_\varphi|\nu|} |\nu|^\tau/\gamma \ .$$

From the elementary inequality $x^\tau e^{-ax} \le (\tau/ae)^\tau$, used with $a=\delta\rho$ and $x=|\nu|$, the conclusion is immediate. □

The reduction of the actions domain is unnecessary and was made only for notational simplicity; still for simplicity, from now on we shall restrict the attention to $\tau = n$, and denote by c any convenient constant depending only on n, in general different in the different inequalities. (The meaning of the inequalities containing c will be: there exists c depending only on n, such that the inequality holds.)

3.3 One perturbation step for isochronous systems

We are now ready to introduce a proposition which states the possibility of performing one perturbation step for the isochronous system (3.1), in the assumption that the perturbation f is analytic in the above introduced domain $\mathcal{D}_\rho$.

PROPOSITION 3.5 *Let H be as in (3.1), with (γ, n)–Diophantine frequency ω. Assume f is analytic in $\mathcal{D}_\rho$ for some $\rho=(\rho_I,\rho_\varphi)$, and $\|f\|_\rho<\infty$. Denote*

$$E = \gamma \rho_I \rho_\varphi^{n+1} \ . \tag{3.10}$$

For any $0<\eta<1$, if

$$\varepsilon < \varepsilon_0 = c\eta^{n+2} E/\|f\|_\rho \ , \tag{3.11}$$

there exists an analytic canonical transformation $w : \mathcal{D}_{(1-\eta)\rho} \to \mathcal{D}_\rho$ such that the new Hamiltonian $H' = H\circ w : \mathcal{D}_{(1-\eta)\rho} \to \mathbb{C}$ has the form

$$H'(I',\varphi') = \omega\cdot I' + \varepsilon g(I') + \varepsilon^2 f'(I',\varphi') \ , \tag{3.12}$$

with $g=\langle f\rangle$, the remainder f' satisfying the estimate

$$\|f'\|_{(1-\eta)\rho} < c\,\eta^{-n-2}\, \|f\|_\rho^2/E \ . \tag{3.13}$$

The canonical transformation is ε–close to the identity, more precisely

$$\|w - \mathbb{I}\|_{(1-\eta)\rho} < \frac{\varepsilon}{\varepsilon_0}\eta \ . \tag{3.14}$$

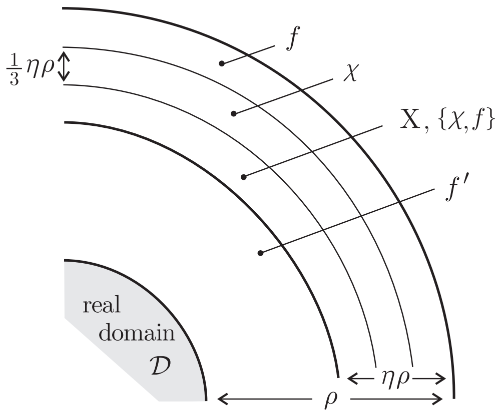

Figure 15: *The domains entering proposition 3.5.*

PROOF. We already know from Lemma 3.4 that χ solving the basic equation (2.22) does exist and correspondingly the new Hamiltonian has the form (3.12), with

$$g = \langle f \rangle \ , \qquad f' = \mathcal{R}_2[\chi, \omega \cdot I] + \mathcal{R}_1[\chi, f] \ .$$

We must then provide the estimates for f' and for the canonical transformation $w = \Phi^1_{\varepsilon\chi}$, as well as show that if condition (3.11) on ε is satisfied, then w is, as required, a map $\mathcal{D}_{(1-\eta)\rho} \to \mathcal{D}_\rho$.

All of this is easily aceived by a sequence of elementary estimates in progressively reduced domains, see figure 15, from $\mathcal{D}_\rho$ to $\mathcal{D}_{(1-\eta)\rho}$.

a) Let $\delta = \frac{1}{3}\eta$. From Lemma 3.4 we know that

$$\|\chi\|_{(1-\delta)\rho} < c\rho_\varphi^{-n}\delta^{-n}\frac{\|f\|_\rho}{\gamma} \ ;$$

then a Cauchy estimate[4] immediately leads to an estimate of the vector field $X_\chi = (-\frac{\partial\chi}{\partial\varphi}, \frac{\partial\chi}{\partial I})$ in the reduced domain $\mathcal{D}_{(1-2\delta)\rho}$, precisely

$$\|X_\chi\|_{(1-2\delta)\rho} < c\rho_I^{-1}\rho_\varphi^{-n-1}\delta^{-n-1}\frac{\|f\|_\rho}{\gamma} \ . \qquad (3.15)$$

[4]For $f : \mathbb{C} \to \mathbb{C}$ analytic in the complex disc $\Delta_r(z)$ of radius r centered in z, the basic Cauchy estimate, trivially deduced from Cauchy integral formula, is

$$|f'(z)| \le \frac{1}{r} \sup_{z \in \Delta_r(z)} |f(z')| \ .$$

From this expression, used for the partial derivatives of χ, it is easy to deduce (3.15). Cauchy estimates are usually introduced with reference to the supremum norm, but work with the Fourier norm, too.

b) A sufficient condition to ensure that $\Phi^1_{\varepsilon\chi}$ sends $\mathcal{D}_{(1-\eta)\rho}$ into $\mathcal{D}_{(1-2\delta)\rho} \subset \mathcal{D}_\rho$ is obviously

$$\varepsilon \|X_\chi\|_{(1-2\delta)\rho} < \delta \ ;$$

in turn, this is ensured by the smallness condition (3.11) on ε.

c) From the very definition of Poisson bracket, it follows

$$\{f,\chi\} = \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{F}(s)\Big|_{s=0} \ , \qquad \mathcal{F}(s) = f\Big(I - s\frac{\partial\chi}{\partial\varphi}(I,\varphi), \varphi + s\frac{\partial\chi}{\partial I}(I,\varphi)\Big) \ .$$

Now, from (3.15) it is guaranteed that the point $(I,\varphi) + sX_\chi(I,\varphi)$ does not escape $\mathcal{D}_\rho$, for $(I,\varphi) \in \mathcal{D}_{(1-2\delta)\rho}$, up to

$$|s| \le s_* = \frac{2\delta}{\|X_\chi\|_{(1-2\delta)\rho}} \ ;$$

a Cauchy estimate on $\mathcal{F}$ then gives

$$\|\{f,\chi\}\|_{(1-2\delta)\rho} \le \frac{\|f\|_\rho}{s_*} < c\frac{\|f\|_\rho}{\varepsilon_0} \ .$$

d) The estimate on the Poisson bracket in turn allows to estimate the remainders. Concerning $\mathcal{R}_1[\chi, f]$, from its expression (2.19), taking into account that $\Phi^t_{\varepsilon\chi}(I,\varphi) \in \mathcal{D}_{(1-2\delta)\rho}$ for $(I,\varphi) \in \mathcal{D}_{(1-\eta)\rho}$ and $|t| \le 1$, it trivially follows

$$\|\mathcal{R}_1\|_{(1-\eta)\rho} < \|\{f,\chi\}\|_{(1-2\delta)\rho} \ ;$$

in the same way from the expression (2.20) of $\mathcal{R}_2$, also recalling $\{\omega\cdot I, \chi\} = f - \langle f\rangle$, it follows

$$\|\mathcal{R}_2\|_{(1-\eta)\rho} < \frac{1}{2}\|\{f,\chi\}\|_{(1-2\delta)\rho} \ ,$$

and (3.13) is immediate. □

Remark. The chain of estimates entering the above proposition was

$$f \quad\longrightarrow\quad \chi \quad\longrightarrow\quad X_\chi \quad\longrightarrow\quad f' \ .$$

A more technical (but also more deep) procedure [41] allows to avoid the estimates of the functions f and g, and to look only at (the Fourier norm of) their vector fields X_f and X_χ. The chain of estimates gets shorter,

$$X_f \quad\longrightarrow\quad X_\chi \quad\longrightarrow\quad X_{f'} \ ;$$

the advantage — besides the conceptual point that only the derivatives of χ and f are involved, while the functions themselves are not— is that the factor η^{n+2} in (3.11) and (3.13) is replaced by η^{n+1} (one Cauchy estimate is saved). *We shall later take into account this small improvement.* The value $n+1$ of the exponent is believed to be optimal.

3.4 Finitely many steps

Once we know how to perform one step, thus passing from ε to ε^2, it is easy to proceed recursively and perform any finite number r of steps, so as to push the remainder to order ε^{r+1}. The input of a typical step is a Hamiltonian H_s in "normal form" up to order $s < r$,

$$H_s(I,\varphi) = \omega \cdot I + \varepsilon g_s(I) + \varepsilon^{s+1} f_s(I,\varphi) \ , \tag{3.16}$$

defined in a domain $\mathcal{D}_{(1-s\eta)\rho}$; the output is a new Hamiltonian H_{s+1} of exactly the same form, but with s replaced everywhere (including the domain) by $s+1$. Initially it is of course $s = 0$, with $f_0 = f$, $g_0 = 0$. The final Hamiltonian H_r is defined in $\mathcal{D}_{(1-r\eta)\rho}$, and *if we wish that this domain is not empty, we must take* $\eta < r^{-1}$, for example $\eta = (2r)^{-1}$ so as the final domain is $\mathcal{D}_{\rho/2}$.

The step from s to $s+1$ can be performed by the Lie method, with $w_s = \Phi^1_{\varepsilon^{s+1}\chi_s}$. The new Hamiltonian is immediately found to be

$$\begin{aligned} H_{s+1} \;&=\; \omega \cdot I + \varepsilon g_s + \varepsilon^{s+1}(\{\omega \cdot I, \chi_s\} + f_s) \\ &\qquad + \varepsilon^{s+2}\mathcal{R}_1[\chi_s, g_s + \varepsilon^s f_s] + \varepsilon^{2s+2}\mathcal{R}_2[\chi_s, \omega \cdot I] \ , \end{aligned}$$

so the goal is aceived if the third term is independent of the angles: that is *if* χ_s *satisfies the basic equation*

$$\omega \cdot \frac{\partial \chi_s}{\partial \varphi}(I,\varphi) = f_s(I,\varphi) - \langle f_s \rangle \ .$$

It is of crucial importance that the equation to be solved at each step is the same. As we know, the Diophantine condition on ω allows to solve the equation; the new Hamiltonian has then the wanted form, with

$$g_{s+1} = g_s + \varepsilon^{s+1}\langle f_s \rangle \ , \qquad f_{s+1} = \varepsilon^{s+2}\mathcal{R}_1[\chi_s, g_s + \varepsilon^s f_s] + \varepsilon^{2s+2}\mathcal{R}_2[\chi_s, \omega \cdot I] \ .$$

Trivial changes in the proof of the above proposition lead to the following iterative lemma:

LEMMA 3.6 *Let H_s be as in (3.16), with (γ, n)–Diophantine ω, and assume H_s is analytic in $\mathcal{D}_{(1-s\eta)\rho}$, with $\eta \le \frac{1}{2(s+1)}$. If ε is sufficiently small, namely*

$$\varepsilon^{s+1} < c\eta^{n+2}\, E/\|f_s\|_\rho$$

with E as in (3.10), then there exists an analytic canonical transformation $w_s : \mathcal{D}_{(1-(s+1)\eta)\rho} \to \mathcal{D}_{(1-s\eta)\rho}$, such that the new Hamiltonian $H_{s+1} = H_s \circ w_s$ has the form (3.16), with $s+1$ in place of s, and the following estimates hold:

$$\begin{aligned} \|g_{s+1}\|_{(1-(s+1)\eta)\rho} \;&\le\; \|g_s\|_{(1-s\eta)\rho} + \varepsilon^s \|f_s\|_{(1-s\eta)\rho} \\ \|f_{s+1}\|_{(1-(s+1)\eta)\rho} \;&\le\; c\eta^{-n-2}\Big(\|g_s\|_{(1-s\eta)\rho} + \varepsilon^s\|f_s\|_{(1-s\eta)\rho}\Big)\, \|f_s\|_{(1-s\eta)\rho}/E \ . \end{aligned}$$

In the further reduced domain $\mathcal{D}_{(1-s\eta-x)\rho}$, with any x such that $\eta \le x < \frac{1}{2}$, the canonical transformation is ε^{s+1}–near the identity, namely

$$\|w_s\|_{(1-s\eta-x)\rho} < c\,\varepsilon^{s+1}\,\frac{\|f_s\|_{(1-s\eta)\rho}}{xE} \ .$$

Remark. The only (minor) novelty in the proof is the presence g_s in the estimate of f_{s+1}. The estimate remains elementary, but it is worthwhile to observe that the leading contribution to the norm of f_{s+1} comes precisely from g_s. The domain reduction $x\rho$ in the estimate of w_s is free; the estimate will be later used with $x = \frac{1}{4}$ and not with x as small as η (for large r, η gets small, and the estimate is much worse).

The lemma is easily iterated r times, leading to a final Hamiltonian H_r with a remainder of order ε^{r+1}. Using $\eta = (2r)^{-1}$, the final Hamiltonian H_r turns out to be defined in $\mathcal{D}_{\rho/2}$, and it is easy to check iteratively that for $s = 0, \ldots, r$ the following estimates hold:

$$\|g_s\|_{(1-s\eta)\rho} \le \|f\|_\rho \sum_{j=0}^{s-1} 2^{-j} < 2\|f\|_\rho \ , \qquad \|f_s\|_{(1-s\eta)\rho} \le \Lambda^s \|f\|_\rho \ ,$$

with

$$\Lambda = c\eta^{-n-2}\|f\|_\rho/E = cr^{n+2}\|f\|_\rho/E \ .$$

A sufficient condition of smallness of ε in order for the lemma can be iterated r times is easily found to be

$$\varepsilon < \frac{1}{2\Lambda} \ . \tag{3.17}$$

Using $x = \frac{1}{4}$, the overall canonical transformation $w = w_{r-1} \circ \ldots \circ w_0$ turns out to be bounded by

$$\|w - \mathbb{I}\|_{\frac{1}{4}\rho} < c\varepsilon\,\|f\|_\rho/E \ .$$

Remark. Thanks to the already quoted improvement of ref. [41], see the remark at the end of the previous subsection, the value of Λ slightly improves to

$$\Lambda = cr^{n+1}\frac{\|f\|_\rho}{E} \ .$$

This is the value that we shall use in the future.

3.5 Exponential estimates for isochronous systems

A quite elementary trick allows to pass without any further work from estimates of finite order ε^r to an exponential estimate, which in principle implies a limit $r \to \infty$. It is clearly impossible to take the limit $r \to \infty$ at finite ε: indeed on one hand this is prevented by the smallness condition (3.17) on ε which cannot be any more fulfilled, on the other hand the result would be useless, since the estimate on the remainder

$$\varepsilon^{r+1}\|f_r\|_{\frac{1}{2}\rho} < \varepsilon^{r+1}(Cr^{n+1})^r\ , \qquad C = c\|f\|_\rho/E\ , \tag{3.18}$$

for $r \to \infty$ would explosively diverge. *It is however possible to take a simultaneous limit $\varepsilon \to 0$, $r \to \infty$, in such a way that the condition on the smallness of ε is satisfied and the remainder converges to zero exponentially fast.* Indeed it is enough to take $r = r(\varepsilon)$ such that

$$Cr^{n+1}\varepsilon = e^{-1} \qquad \text{i.e.} \qquad r = (Ce\varepsilon)^{-\frac{1}{n+1}}$$

(forget here the inessential fact that r must be integer), to work out the exponential estimate

$$\varepsilon^{r+1}\|f_r\|_{\frac{1}{2}\rho} < c\,\varepsilon\, e^{-(\varepsilon^*/\varepsilon)^{\frac{1}{n+1}}}\ , \quad \varepsilon^* = (Ce)^{-1}\ .$$

As pointed out in [54], this choice of r as function of ε is optimal (up to a minor improvement in ε^*), of course if the estimate (3.18) on the final remainder is considered optimal. The exponential estimates for isochronous systems with Diophantine ω first appeared in [49] and then were developed in [54] [26]. The contribution was inspired by Nekhoroshev's theorem.

The result we proved (up to minor details which are left as exercises) is summarized in the following proposition:

PROPOSITION 3.7 (Nekhoroshev–Gallavotti) *Let H be as in (3.1); assume:*

i) *f is analytic in $\mathcal{D}_\rho$, $\rho = (\rho_I, \rho_\varphi)$.*

ii) *ω is (γ, n)–Diophantine.*

iii) *$\varepsilon < \varepsilon^* = cE/\|f\|_\rho$, with $E = \gamma\rho_I\rho_\varphi^n$.*

There exists an analytic canonical transformation $(I, \varphi) = w(I', \varphi')$, $w : \mathcal{D}_{\frac{1}{2}\rho} \to \mathcal{D}_\rho$, such that the new Hamiltonian $H' = H \circ w$ has the form

$$H'(I', \varphi') = \omega \cdot I' + \varepsilon g(I') + \varepsilon e^{-(\varepsilon^*/\varepsilon)^a} f'(I', \varphi')\ , \qquad a = \frac{1}{n+1}\ , \tag{3.19}$$

g and f' satisfying the estimates

$$\|g\|_{\frac{1}{2}\rho} < 2\|f\|_\rho\ , \qquad \|f'\|_{\frac{1}{2}\rho} < \|f\|_\rho\ .$$

The canonical transformation is small with ε, more precisely for $(I', \varphi') \in \mathcal{D}_{\frac{1}{4}\rho}$ it is

$$|I'_j - I_j| < \frac{\varepsilon}{\varepsilon^*}\rho_I\ , \qquad |\varphi'_j - \varphi_j| < \frac{\varepsilon}{\varepsilon^*}\rho_\varphi\ , \qquad j = 1, \ldots, n\ . \tag{3.20}$$

From this proposition it immediately follows that

COROLLARY 3.8 *For real initial data it is*

$$\|I(t) - I^o\| \le 3\frac{\varepsilon}{\varepsilon^*}\rho_I$$

up to

$$|t| < \mathcal{T}\, e^{(\varepsilon^*/\varepsilon)^a} \ , \qquad \mathcal{T} = c\gamma^{-1} \ . \tag{3.21}$$

PROOF. From (3.19) it is

$$\|\dot{I}'\| < \varepsilon e^{-(\varepsilon^*/\varepsilon)^a} \|f\|_\rho / \rho_\varphi \ ,$$

so that

$$\|I'(t) - I'(0)\| < \frac{\varepsilon}{\varepsilon^*}\rho_I$$

up to the time

$$|t| < c\frac{\rho_I \rho_\varphi}{\varepsilon^* \|f\|_\rho} e^{(\varepsilon^*/\varepsilon)^a} \le c\gamma^{-1}\, e^{(\varepsilon^*/\varepsilon)^a} \ .$$

Taking into account (3.20), the conclusion is immediate. □

So the averaging principle is satisfied up to exponentially long times and the situation is similar to the one depicted in figure 11, with however the *great* improvement that $e^{(\varepsilon^*/\varepsilon)^a}$ replaces $1/\varepsilon$ on the horizontal axis.

4 Nondegenerate anisochronous systems

In this section we shall consider perturbed nondegenerate anisochronous systems, namely Hamiltonian systems of the form

$$H(I,\varphi) = h(I) + \varepsilon f(I,\varphi) \ , \qquad (I,\varphi) \in B \times \mathbb{T}^n, \tag{4.1}$$

such that the nondegeneracy condition

$$\det\Big(\frac{\partial^2 h}{\partial I_i \partial I_j}\Big) \ne 0 \tag{4.2}$$

is satisfied in B. In such a case the frequency $\omega = \frac{\partial h}{\partial I}$ does depend on I, more precisely the map $I \mapsto \omega(I)$ is a local diffeomorphism. These systems satisfy the assumption i) of Poincare's theorem.

The study of anisochronous systems includes both an arithmetic–analytic part and a geometric part. The former is very similar to the isochronous case (including the difficulties related to small divisors and divergence of series), and we do not really need to learn anything new. The latter is instead new, so we shall devote to it most of our attention. The idea is to understand which assumptions on h are necessary to prove that the system obeys the averaging principle. The nondegeneracy condition (4.2) alone is not sufficient: indeed in section 2.3 we already met an elementary anisochronous system for which the averaging principle is not satisfied; see (2.11).

In essence, the problem with anisochronous systems is that any resonance relation
$$\nu \cdot \omega(I) = 0 \tag{4.3}$$
defines a manifold in the action space B, and the presence of the resonance strongly influences the nearby motions. So, B needs to be decomposed into different regions, characterized by different resonance properties. The geometric part consists precisely in finding the appropriate decomposition and understanding how to work into the the different regions, in such a way that the different local results patch consistently into an unique picture.

As for isochronous systems, we shall first study in detail how to perform a single perturbation step, and only later (section 5.3) we shall discuss the case of many steps. The study of the single step, in turn, starts from a specific example, namely a chain of rotators with nearest–neighbors interaction [22] [23], and then proceeds by progressive generalizations.

4.1 The model of rotators: analytic results

Let us consider a periodic chain of n rotators with positional nearest–neighbors interaction, namely
$$H(I,\varphi) = h(I) + \varepsilon f(\varphi)$$
with
$$h(I) = \frac{1}{2}\sum_{j=1}^{n} I_j^2 \ , \qquad f(\varphi) = -\sum_{j=1}^{n}\cos(\varphi_j - \varphi_{j-1}) \ , \quad \varphi_0 \equiv \varphi_n \ ;$$
f is minimal when all phases are equal. The perturbation has finitely many Fourier components, so that assumption ii) of Poincaré's theorem is violated. The Fourier components entering the perturbation are of the special type $\pm\nu^{(j)}$, with
$$\nu^{(j)} = (0,\ldots,0,-1,1,0,\ldots,0) \ , \tag{4.4}$$
the nonvanishing entries occupying positions $j-1, j$. Such a local character of the interaction helps to understand the results, but it isn't of any real importance, as well as the fact that f is independent of the actions and $\langle f\rangle = 0$. The form of h is instead important, and will be used; understanding how to generalize it is not trivial. The associated frequency is
$$\omega(I) = I \ .$$

A. Nonresonant motions. Let us perform one perturbation step using the Lie method, with $w = \Phi^1_{\varepsilon\chi}$. As we know, $H' = H \circ w$ is of the form
$$H' = h + \varepsilon(\{h,\chi\} + f) + \varepsilon^2(\mathcal{R}_2[\chi,h] + \mathcal{R}_1[\chi,f]) \ ,$$

so the term of order ε is independent of the angles iff χ satisfies the basic equation (2.22). In this example the equation takes the form

$$\omega(I) \cdot \frac{\partial \chi}{\partial \varphi}(I, \varphi) = -\sum_{j=1}^{n} \cos(\varphi_j - \varphi_{j-1}) \ ;$$

its solution is

$$\chi = -\sum_{j=1}^{n} \frac{\sin(\varphi_j - \varphi_{j-1})}{I_j - I_{j-1}} \ ,$$

and the new perturbation f' is immediately found to be of the form

$$f' = \{f, \chi\} + \frac{1}{2}\{\{h, \chi\}, \chi\} + \mathcal{O}(\varepsilon) = \frac{1}{2}\{f, \chi\} + \mathcal{O}(\varepsilon) \ ,$$

which gives

$$\begin{aligned} f'(I, \varphi) \quad &= \quad \frac{1}{2}\sum_{i=1}^{n} \frac{1 - \cos 2(\varphi_i - \varphi_{i-1})}{(I_i - I_{i-1})^2} + \frac{1}{4}\sum_{i=1}^{n} \Big[\cos(\varphi_{i+1} - \varphi_{i-1}) \\ &- \quad \cos(\varphi_{i+1} - 2\varphi_i + \varphi_{i-1}) \Big] \Big[\frac{1}{(I_{i+1} - I_i)^2} - \frac{1}{(I_i - I_{i-1})^2} \Big] + \mathcal{O}(\varepsilon) \ . \end{aligned}$$

A few comments are in order.

i) The expression of χ contains the small divisors

$$\nu^{(j)} \cdot \omega(I) = I_j - I_{j-1}$$

and the expression of f' contains the squares of the same divisors. Both expressions are thus meaningless on the resonant manifolds (4.3), which in the case at hand are the n planes

$$I_j - I_{j-1} = 0 \ .$$

ii) The new perturbation contains new Fourier components, but at the lowest order in ε there are only finitely many of them, namely those with

$$\nu = \pm\nu^{(j)} \pm \nu^{(j-1)} \ .$$

The expression of f' shows that, in enforcement to the above point i), the perturbation step we made is useless not only on the resonant planes, but also in neighborhoods of them of size $\mathcal{O}(\sqrt{\varepsilon})$. Indeed far of the resonances, when $|I_j - I_{j-1}| = \mathcal{O}(1)$ for all j, the new perturbation is of order ε^2, that is, one order in ε is gained. But if we approach the resonant planes, the gain gets progressively lost; in particular, *if any of the denominators $|I_j - I_{j-1}|$ is as small as $\sqrt{\varepsilon}$, the new perturbation has norm $\varepsilon^2\|f'\| \sim \varepsilon$, i.e. is as large as the original one.* In intermediate situations something is gained, though less than a factor ε: for example, at distance $\mathcal{O}(\varepsilon^{1/4})$ from the resonances it is $\varepsilon^2\|f'\| \sim \varepsilon^{3/2}$.

B. Motions inside a simple resonance. Let $|I_2 - I_1| = \mathcal{O}(\sqrt{\varepsilon})$ and $|I_j - I_{j-1}| \gg \sqrt{\varepsilon}$, say for simplicity $|I_j - I_{j-1}| = \mathcal{O}(1)$, for $j \neq 2$. The best normal form we can produce is obtained by using

$$\chi = -\sum_{j\neq 2} \frac{\sin(\varphi_j - \varphi_{j-1})}{I_j - I_{j-1}} \ ;$$

correspondingly the new Hamiltonian has the form

$$H'(I', \varphi') = \frac{1}{2}({I'_1}^2 + {I'_2}^2) - \varepsilon \cos(\varphi'_2 - \varphi'_1) + \frac{1}{2}\sum_{j>2} {I'_j}^2 + \mathcal{O}(\varepsilon^2) \ .$$

As far as the remainder of order ε^2 can be neglected, the motion of the first two rotators is governed by the reduced Hamiltonian

$$\mathcal{H}(I_1, I_2, \varphi_1, \varphi_2) = \frac{1}{2}(I_1^2 + I_2^2) - \varepsilon \cos(\varphi_2 - \varphi_1)$$

(primes have been dropped). The overall momentum $I_1 + I_2$ is constant, so the mean motion decouples and the system is essentially one dimensional, actually a pendulum; the motion clearly consists in periodic oscillations of the actions, of amplitude of order $\sqrt{\varepsilon}$ and period of order $1/\sqrt{\varepsilon}$, around the constant average $J = \frac{1}{2}(I_1 + I_2)$. A convenient description of the resonant motion (useful for generalizations) is obtained as follows: assume, with no loss of generality,[5] $I_1 + I_2 = 0$. Then $I_1 \simeq I_2 \simeq \sqrt{\varepsilon}$, and it is convenient to perform a canonical blow–up, see section 1.4,

$$I_j = \sqrt{\varepsilon}\, \hat{I}_j \ , \qquad \varphi_j = \hat{\varphi}_j \ , \qquad t = \frac{1}{\sqrt{\varepsilon}}\, \hat{t} \ . \tag{4.5}$$

The domain of the new variables now is of order one and *the new Hamiltonian* $\hat{\mathcal{H}} = \varepsilon^{-1}\mathcal{H}$ *becomes independent of* ε*,*

$$\hat{\mathcal{H}} = \frac{1}{2}(\hat{I}_1^2 + \hat{I}_2^2) - \cos(\hat{\varphi}_2 - \hat{\varphi}_2) \ .$$

The motions of the variables I_1, I_2 (as far as the $\mathcal{O}(\varepsilon^2)$ remainder is neglected) coincide with the ε–independent bounded motions of this Hamiltonian, simply rescaled in space and time by factors dependent on ε. Such motions of course are trivial, since the system we are considering has only one effective degree of freedom; this in turn follows from the fact that we are studying a single resonance.

[5] It is always possible to subtract the mean motion of the pair of rotators, by setting

$$J = \frac{1}{2}(I_1^o + I_2^o) \ , \qquad I_j = J + \tilde{I}_j \ , \qquad \varphi_j = Jt + \tilde{\varphi}_j \ ;$$

the new variables follow the same Hamiltonian $\mathcal{H}'$, and $\tilde{I}_1 + \tilde{I}_2 = 0$.

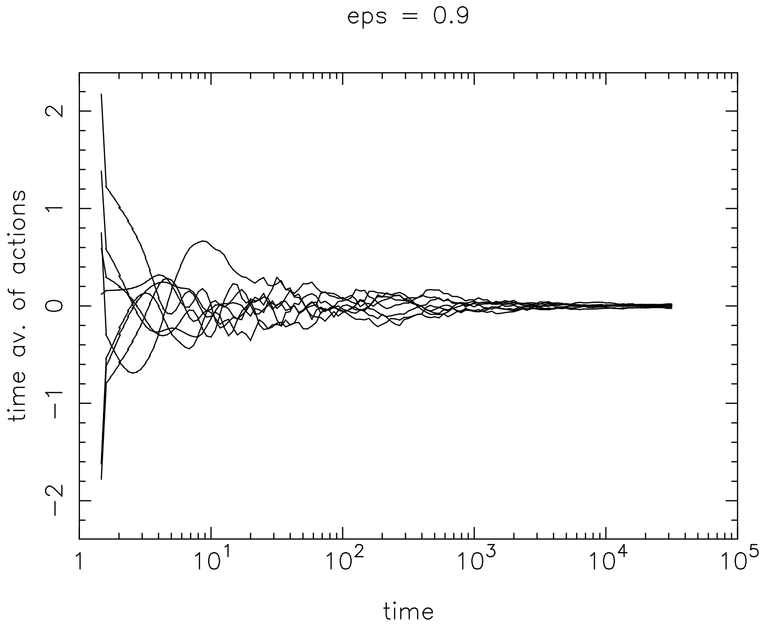

Figure 16: $\overline{I_j}(T)$ *as a function of* T, *for* $\varepsilon = 0.9$.

C. Motions inside multiple resonances. It is sufficient to understand the case of two nearby resonances, say

$$|I_2 - I_1|\,,\ |I_3 - I_2| \;=\; \mathcal{O}(\sqrt{\varepsilon})\ ; \tag{4.6}$$

the generalization to the case of multiple resonances is then trivial. Proceeding as above, we easily produce a normal form adapted to the resonance, i.e.

$$H'(I',\varphi') = \mathcal{H}(I',\varphi') + \frac{1}{2}\sum_{j>3} {I'_j}^2 + \mathcal{O}(\varepsilon^2)\ , \tag{4.7}$$

this time with (dropping the primes)

$$\mathcal{H}(I_1,I_2,I_3,\varphi_1,\varphi_2,\varphi_3) = \frac{1}{2}\sum_{j=1}^{3} I_j^2 - \varepsilon\left[\cos(\varphi_2-\varphi_1)+\cos(\varphi_3-\varphi_2)\right]\ . \tag{4.8}$$

As before we ignore the $\mathcal{O}(\varepsilon^2)$ coupling and analyze the motions of $\mathcal{H}$. We suppose as above that the overall angular momentum $I_1 + I_2 + I_3$ vanishes; it is then $I_1 \simeq I_2 \simeq I_3 \simeq \sqrt{\varepsilon}$, and the blow up (4.5) gives

$$\hat{\mathcal{H}} = \frac{1}{2}\sum_{j=1}^{3} \hat{I}_j^2 \;-\; \left[\cos(\hat\varphi_2-\hat\varphi_1)+\cos(\hat\varphi_3-\hat\varphi_2)\right]\ .$$

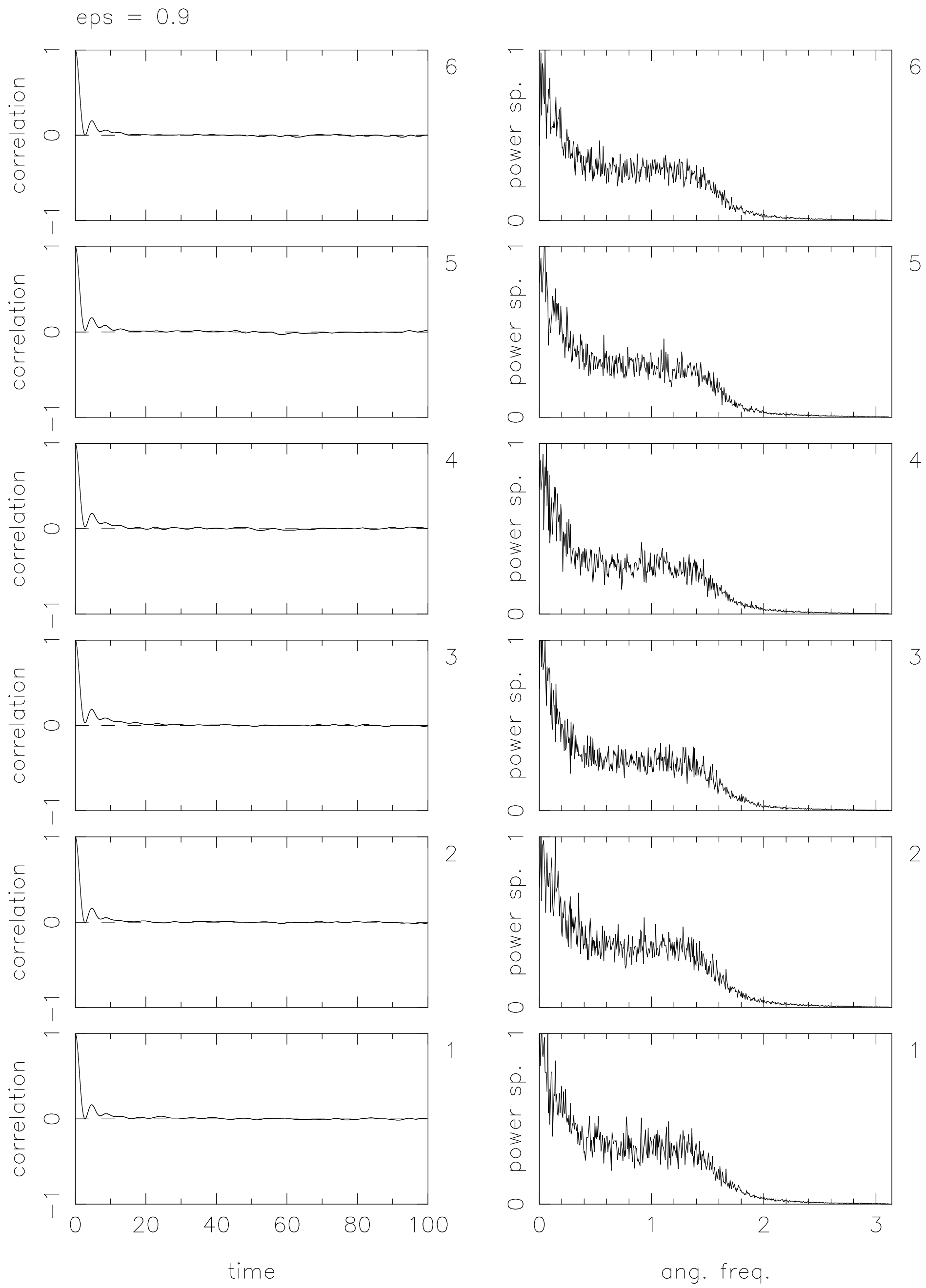

Figure 17: γ_j *(left) and* W_j *(right),* $j = 1, \ldots, 6$*, bottom to top, for* $\varepsilon = 0.9$.

This is a strongly coupled, ε–independent, Hamiltonian systems with two effective degrees of freedom, which is not integrable and possibly includes chaotic motions.[6] Chaos evidently develops on a length scale and a time scale which are independent of ε, and thanks to the energy conservation, all motions are bounded. Correspondingly, *no matter how small is ε, $\mathcal{H}$ has bounded possibly chaotic motions, which develop on a length scale $\sqrt{\varepsilon}$ and on a time scale $1/\sqrt{\varepsilon}$.* As far as the $\mathcal{O}(\varepsilon^2)$ term is neglected,[7] chaotic motions do exist in the original Hamiltonian, too, and appear as chaotic motions localized inside the double resonance, actually bounded by $\|I(t) - I^o\| \sim \sqrt{\varepsilon}$, and developing on times of order $1/\sqrt{\varepsilon}$ (the Lyapunov exponent would scale as $\sqrt{\varepsilon}$). In this special model, chaos is localized in the chain, namely it involves only three nearby rotators (up to a minor effect on the other ones, due to the fact that I'_j and I_j do not exactly coincide).

It is worthwhile to stress that this situation, in which some variables move regularly and some other perform chaotic motions, the oscillations of all actions being however bounded by $\sqrt{\varepsilon}$ up to $|t| \le \varepsilon^{-1}$, does accomplish the averaging principle, which by no means requires the motions to be regular (or to be bounded by ε rather than by $\sqrt{\varepsilon}$).

Remark. We arrived to this picture on the basis of heuristic considerations. It is not difficult, however, to deduce rigorously from Hamiltonian (4.7) that inside the resonance (4.6) and far from other resonances ($I_{j+1} - I_j = \mathcal{O}(1)$ for $j > 3$) the actions $I_4, \dots, I_n$ oscillate at most by quantities of order ε and $I_1, \dots, I_3$ oscillate at most by quantities of order $\sqrt{\varepsilon}$, for $|t| \le \varepsilon^{-1}$. (It is indeed sufficient to observe that $I_4, \dots, I_n$, as well as $J = (I_1 + I_2 + I_3)$, have a drift bounded by (const)ε for $|t| \le \varepsilon^{-1}$, and then to use the energy conservation.)

4.2 The model of rotators: numerical results

Before generalizing the above results to a wider class of models, it is convenient to take more familiarity with the model of rotators by means of convenient numerical experiments. We follow here literally [23].

The model we consider is a chain of $n = 8$ rotators, with Hamiltonian as in the previous subsection. The computed quantities include:

(a) The time averages of the actions up to time T, namely

$$\overline{I_j}(T) = \frac{1}{T}\int_0^T I_j(t)\,\mathrm{d}t \ .$$

[6] The presence of chaotic motions, expected theoretically and found numerically in [23], is hard to prove rigorously. However, if one modifies the model by introducing a small parameter μ in front of one of the cosines, then it gets easy to prove, by means of the so–called Melnikov method, that an homoclinic intersection does exist for small μ [26].

[7] At variance with integrability, chaos in general is persistent under small perturbations. In some cases this is a theorem, for example the presence of a homoclinic transversal intersection (see the previous footnote) is not destroyed by a small smooth perturbation.

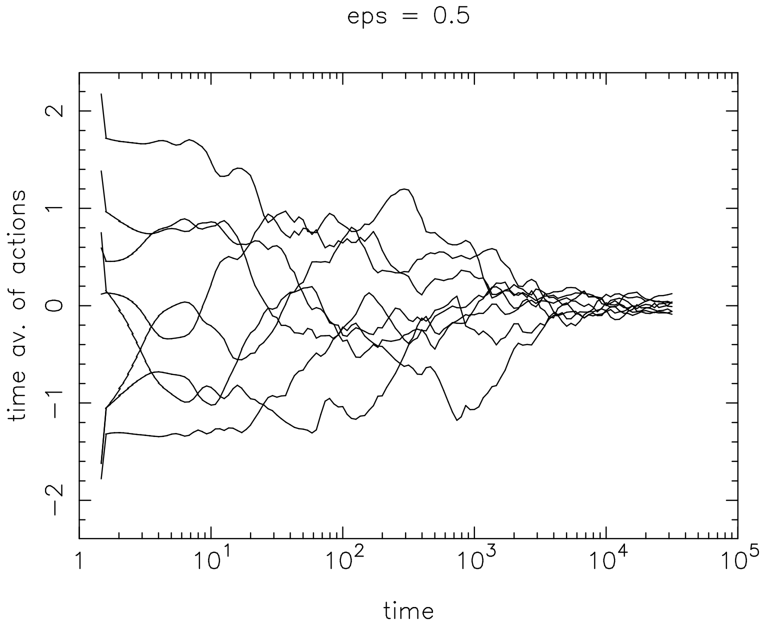

Figure 18: *Same as in figure 16, for $\varepsilon = 0.5$.*

Practically, if s is the time–step and $T = Ns$, the computed quantity is

$$\overline{I_j}(sN) = \frac{1}{N} \sum_{k=0}^{N-1} I_j(ks) \ .$$

In principle one would like to compute the infinite average $\overline{I_j}(\infty) = \lim_{T\to\infty} \overline{I_j}(T)$; in practice of course the computation is truncated at a large $T_{\max} = sN_{\max}$.

(b) The (normalized) autocorrelation function $\gamma_j(\tau)$, defined as follows:

$$\gamma_j(\tau) = \frac{\Gamma_j(\tau) - \overline{I_j}^2(\infty)}{\sigma^2} \ , \qquad \Gamma_j(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T I_j(t+\tau) I_j(t)\, \mathrm{d}t \ ,$$

where $\sigma^2 = \Gamma_j(0) - \overline{I_j}^2(\infty)$. Practically, the computation is discretized as above and truncated to $T_{\max}$.

(c) The power spectrum W_j of I_j, whose practical definition, including discretization and truncation, is

$$W_j\Big(\frac{2\pi k}{N_{\max}}\Big) = \Big|\mathcal{I}_j\Big(\frac{2\pi k}{N_{\max}}\Big)\Big|^2 \ , \qquad k = 0, \ldots, N_{\max} - 1 \ ,$$

with

$$\mathcal{I}_j\Big(\frac{2\pi k}{N_{\max}}\Big) = \frac{1}{N_{\max}} \sum_{s=0}^{N_{\max}-1} I_j(ks)\, e^{-2\pi iks/N_{\max}} \ .$$

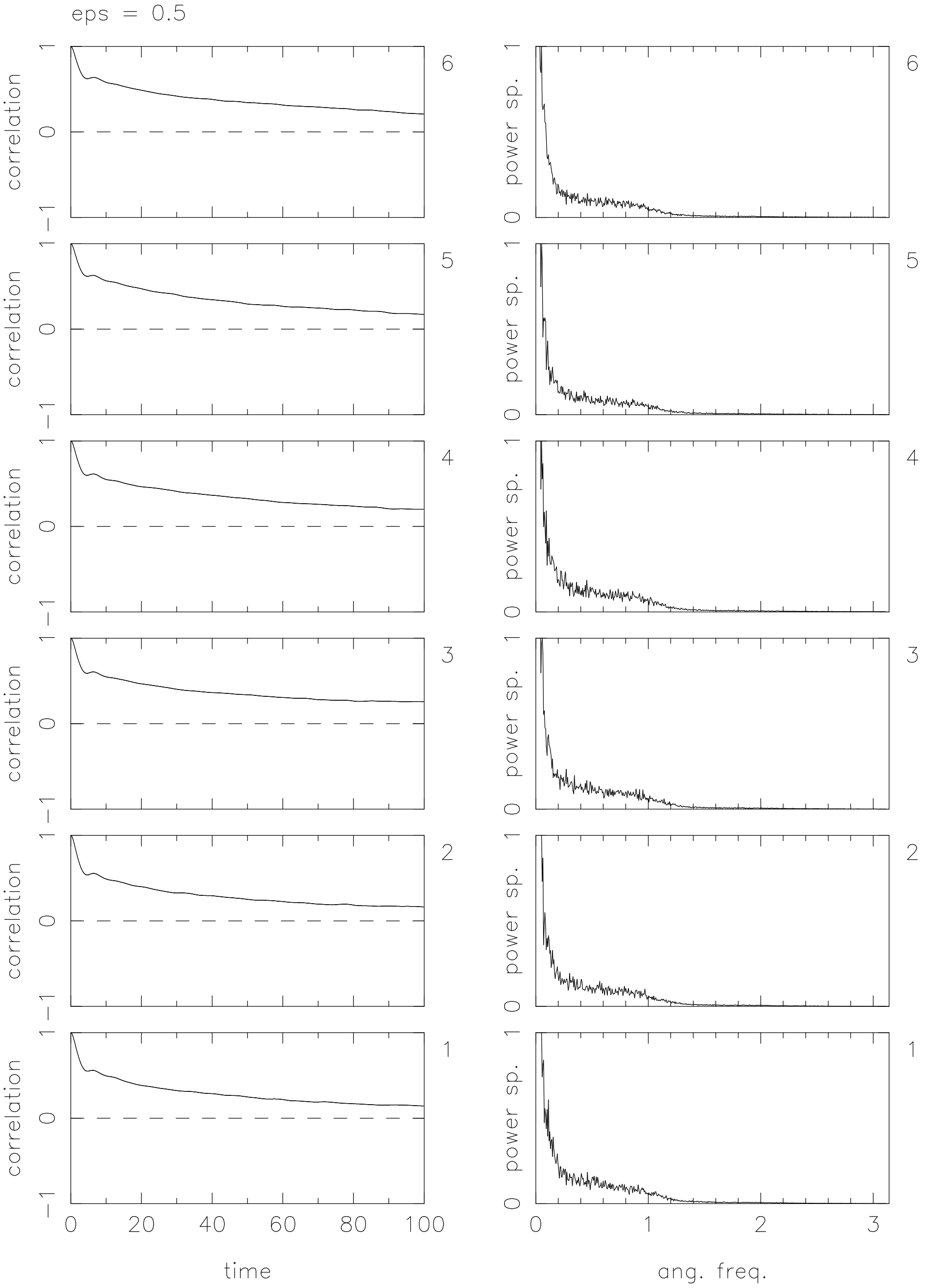

Figure 19: *Same as in figure 17, for* $\varepsilon = 0.5$.

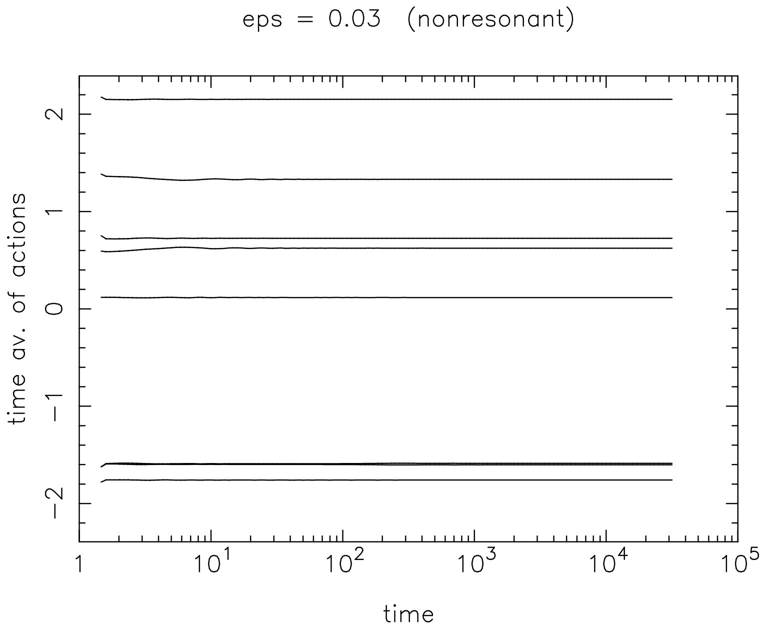

Figure 20: *Same as in figures 16 and 18, for* $\varepsilon = 0.03$ *and nonresonant initial data.*

> The power spectrum turns out to be the Fourier transform of the correlation function, so it does not provide a really independent observation, nevertheless it is useful to look at it, too.[8]

All results refer to initial data with total momentum $I_1 + \cdots + I_n = 0$. Figure 16 shows the behavior of the averages $\overline{I_j}$ for large $\varepsilon = 0.9$. The result is consistent with the conjecture that in such conditions the system is ergodic, so that $\overline{I_j}(\infty) = 0$ for all j. Figure 17 shows in the same conditions the correlation functions (left) and the spectra (right); for graphical reasons, the figure is limited to $j = 1, \ldots, 6$. Correlations decay quite rapidly, in a few time units (the initial angular velocities are of order one, so a few time units means a few turns); the irregular character of the motion gives W_j the form of an essentially continuous band, extending to a maximal frequency $\omega_{\max} \simeq \frac{\pi}{2}$. All rotators behave exactly in the same way.

[8]Such an elementary Fourier analysis, going back to 1980, might appear rough if compared to modern frequency analysis (see the lectures by M. Guzzo and J. Laskar in this volume). As we shall see, however, it is fairly sufficient to provide a qualitative insight into the essential features of resonant and nonresonant motions. Practically, W_j is obtained as an intermediate product of the computation of the correlation function, according to the following procedure: (1) The movement $I_j(t)$ is computed in the interval $(0, T_{\max})$; the average $\overline{I_j}$ is subtracted, and the result is normalized by the variance. (2) The resulting function is extended on a double interval by adding zeros, so as to avoid spurious correlations between $t \simeq 0$ and $t \simeq T_{\max}$. (3) The (fast) Fourier transform is performed; its square gives W_j. (4) The antitransform gives γ_j.

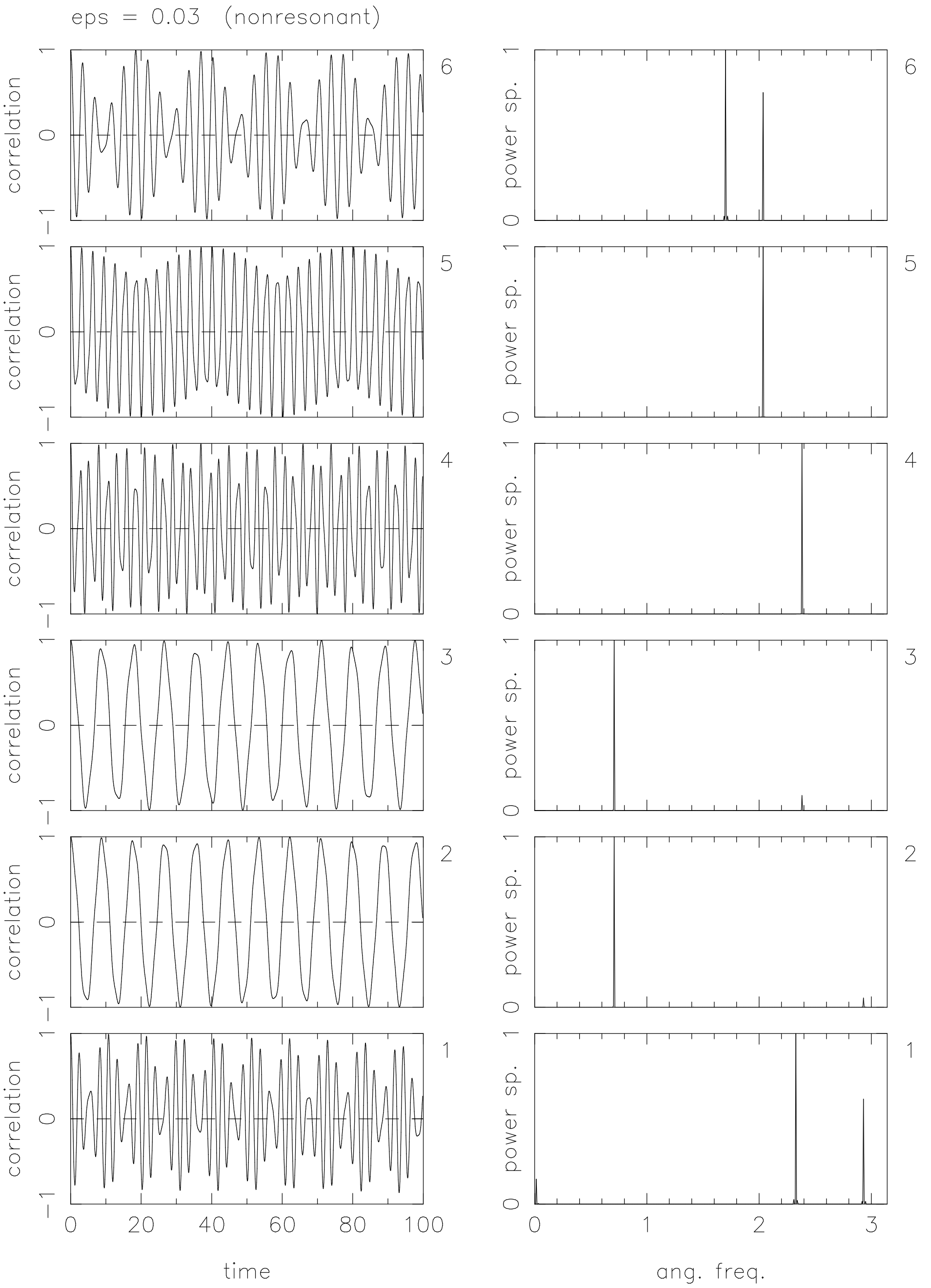

Figure 21: *Same as in figures 17 and 19, for* $\varepsilon = 0.03$ *and nonresonant initial data.*

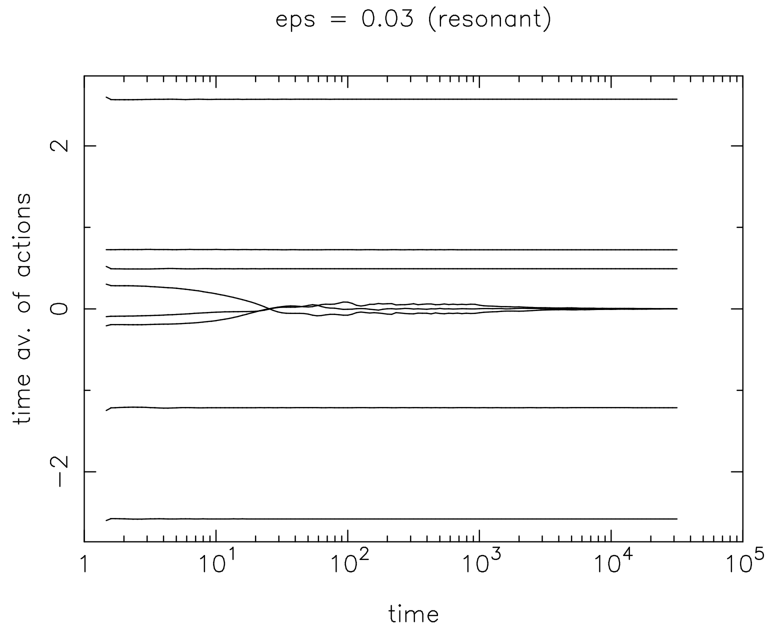

Figure 22: *Same as in figure 20, for the same $\varepsilon = 0.03$, but inside the resonance $I_1 \simeq I_2 \simeq I_3 \simeq \sqrt{\varepsilon}$.*

Figure 18 and 19 report the same quantities as, respectively, figures 16 and 17, but for lower $\varepsilon = 0.5$. There are no essential differences, but the fact that the same phenomena occur on a longer time scale. The next figures 20 and 21 refer to a considerably smaller $\varepsilon = 0.03$ and initial data carefully chosen so as to avoid resonances. The situation now is quite different: actions look stable, correlations do not decay, and the spectrum looks discrete (the lines in W_j are placed exactly at the frequencies $I_{j+1} - I_j$ and $I_j - I_{j-1}$).

The next figures 22, 23 refer to motions inside the double resonance $I_1 \simeq I_2 \simeq I_3 \simeq \sqrt{\varepsilon}$, at $\varepsilon = 0.03$, with initial data such that $I_1^o + I_2^o + I_3^o = 0$. Quite clearly, the first three rotators do interact: the time averages of their actions converge to zero, the correlations decay, the spectrum has the shape of a narrow but continuous band. On the same time the other rotators advance regularly, essentially as in the previous expariments. The decorrelation time provides a time scale of the chaotic motion (alternatively, one could refer to the inverse of the Lyapunov exponent, but it is less easy); according to the above theoretical analysis, this typical time is expected to scale as $1/\sqrt{\varepsilon}$, while correspondingly the width of the continuous bands is expected to scale as $\sqrt{\varepsilon}$. Such scaling laws are easily checked by reducing ε, for example, by a factor four, and looking if the decorrelation time and the width of the bands get respectively multiplied and divided by a factor two. Figure 24 allows to compare γ_j and W_j, $j = 1, 2, 3$, at $\varepsilon = 0.03$ and $\varepsilon = 0.0075$; the time and frequency axes are there

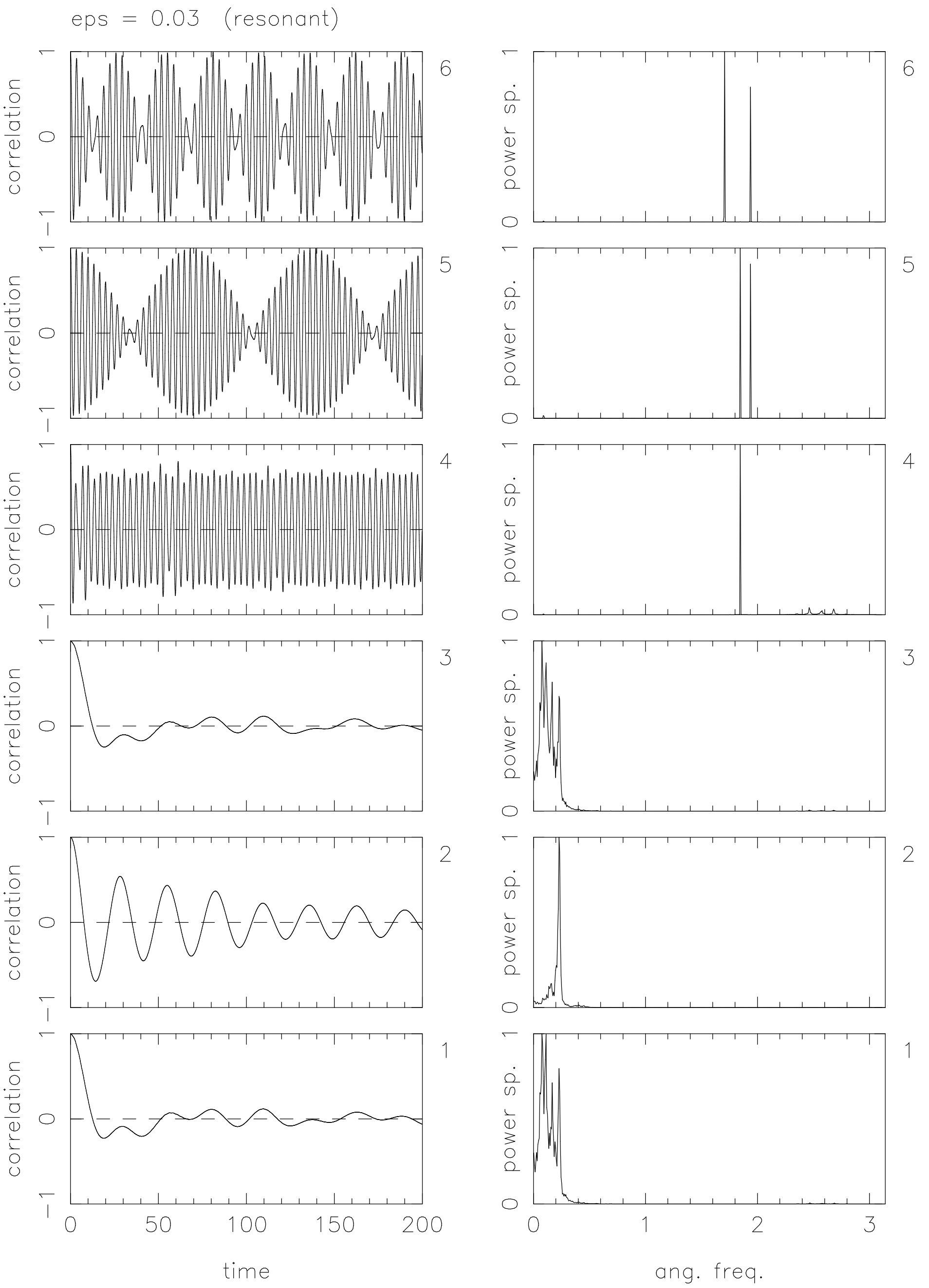

Figure 23: *Same as in figure 21, for the same* $\varepsilon = 0.03$*, but inside the resonance* $I_1 \simeq I_2 \simeq I_3 \simeq \sqrt{\varepsilon}$*. Initial data such that* $I_1^o + I_2^o + I_3^o = 0$*.*

appropriately scaled (moreover, for both values of ε the frequency axis has an extra enlargement of a factor five). The correspondence of the rescaled figures is remarkable.

Remark. Figures 16 – 24 are useful to understand the theory explained in the previous subsection 4.1. But the time scale of the figures at small ε is too large, in order for the first order analysis of section 4.1 to explain completely the numerical results. Higher order perturbation theory would be necessary for a better understanding.

4.3 First order perturbation theory for general convex systems

We now proceed by progressively generalizing, as far as we can, the results we worked out in subsection (4.1) in connection with the model of rotators.

A. Perturbations with finitely many Fourier components: nonresonant motions. Here we mantain provisionally the restriction that f is a Fourier polynomial, though we eliminate any other special property of the perturbation like the local character of the interaction in the chain, and deal with

$$f(I,\varphi) = \sum_{|\nu|\le N} f_\nu(I)\, e^{i\nu\cdot\varphi} \ ,$$

N being some positive constant. Concerning h, for the moment we let it be free, but our main goal will be understanding which are the appropriate assumptions on h which, in agreement with the averaging principle, guarantee the confinement of the actions at least for $|t| \le 1/\varepsilon$.

Let us begin with the nonresonant case. For each Fourier component $\nu \ne 0$ entering the perturbation, we must avoid a "resonant manifold" M_ν of codimension one, defined by

$$\nu\cdot\omega(I) = 0 \ . \tag{4.9}$$

If the initial datum is taken at distance much larger than $\sqrt{\varepsilon}$ from any of such manifolds,

$$|\nu\cdot\omega(I)| \gg \sqrt{\varepsilon} \qquad \text{for} \quad 0<|\nu|\le N \ ,$$

then it is possible to eliminate all of the harmonics of f at order ε, so producing a nonresonant first order normal form. More precisely, the basic equation for χ is solved by

$$\chi(I,\varphi) = \sum_{0<|\nu|\le N} \frac{f_\nu(I)}{i\nu\cdot\omega(I)}\, e^{i\nu\cdot\varphi} \ , \tag{4.10}$$

and the new Hamiltonian is

$$H'(I',\varphi') = h(I') + \varepsilon\langle f\rangle + \varepsilon^2 f'(I',\varphi')$$

with

$$f' = \mathcal{R}_1[\chi,f] + \mathcal{R}_2[\chi,h] = \{f,\chi\} - \frac{1}{2}\{(f-\langle f\rangle),\chi\} + \mathcal{O}(\varepsilon) \ .$$

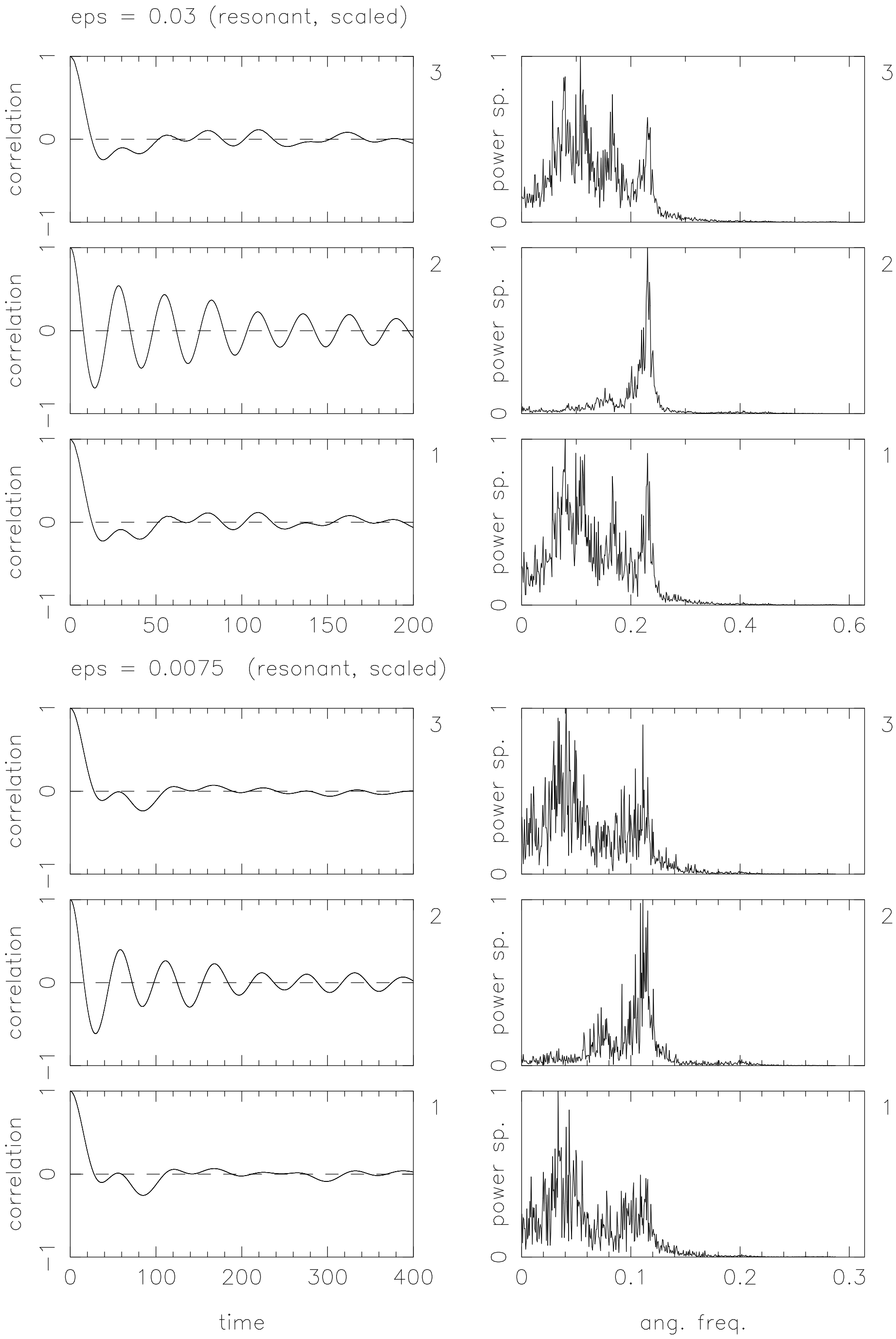

Figure 24: *Correlation functions and power spectra of I_1, I_2, I_3, in the resonance $I_1 \simeq I_2 \simeq I_3 \simeq \sqrt{\varepsilon}$, at $\varepsilon = 0.03$ (top) and $\varepsilon = 0.0075$ (bottom).*

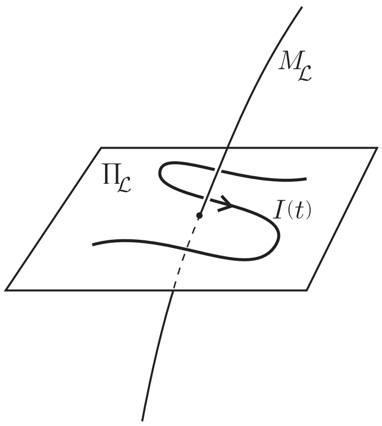

Figure 25: *The motion of the actions is "flattened" on the plane of fast drift* $\Pi_{\mathfrak{L}}$.

Quite clearly, f' contains the derivatives of χ with respect to I, so it contains denominators as small as the squares of the denominators entering χ. The situation is indeed the same as in the case of rotators, and it would be meaningless to use the nonresonant normal form in a $\sqrt{\varepsilon}$–neighborhood of a resonant manifold $\nu \cdot \omega(I) = 0$. No assumptions are instead necessary, for the moment, on h.

B. The resonant construction and the role of convexity. Let $\mathfrak{L}$ be any d–dimensional lattice (i.e. subgroup of $\mathbb{Z}^n$) which contains an integer basis $\nu^{(1)}, \dots, \nu^{(d)}$ such that $|\nu^{(k)}| \leq N$, $1 \leq k \leq d$. The associated *"resonant manifold"* $M_{\mathfrak{L}}$, of dimension d, is defined in the obvious way as

$$M_{\mathfrak{L}} = \left\{ I \in B : \nu \cdot \omega(I) = 0 \ \forall \nu \in \mathfrak{L} \right\} .$$

In a $\sqrt{\varepsilon}$–neighborhood of $M_{\mathfrak{L}}$, but far from other resonances,[9] the best one can do is to eliminate the resonant terms from the expression (4.10) of χ, thus giving the new Hamiltonian H' the resonant normal form

$$H'(I', \varphi') = h(I') + g(I', \varphi') + \mathcal{O}(\varepsilon^2) \ , \tag{4.11}$$

g being the projection of f onto $\mathfrak{L}$, i.e.

$$g(I', \varphi') = \sum_{\nu \in \mathfrak{L}} f_\nu(I')\, e^{i\nu \cdot \varphi'} \ .$$

Let us forget about the $\mathcal{O}(\varepsilon^2)$ remainder and try to understand how the normal part of H' moves the actions. For notational simplicity, we drop the primes (by the way, we recall that the distance between I and I' remains always small with ε). From the Hamilton equations, it is clear that $\dot{I}$ is a linear combination of vectors $\nu \in \mathfrak{L}$, precisely

$$\dot{I} = \varepsilon \sum_{\nu \in \mathfrak{L}} C_\nu(I, \varphi)\, \nu \ , \qquad C_\nu(I, \varphi) = i f_\nu(I)\, e^{i\nu \cdot \varphi} \ .$$

[9] Let us remain a little vague on this notion of being "far from". We shall be more precise in section 5.3, where we shall introduce the geometric construction entering the Nekhoroshev theorem.

So, the dynamics produced by the normal form is flattened on the plane $\Pi_{\mathcal{L}}(I^o)$ parallel to $\mathcal{L}$ through the initial point I^o, while motions transversal to $\Pi_{\mathcal{L}}(I^o)$ are produced only by the remainder; see figure 25. This plane is called the *"plane of fast drift"*.

To provide the wanted confinement inside $\Pi_{\mathcal{L}}$, an assumption of geometric nature on the unperturbed Hamiltonian h turns out to be necessary, and the simplest one is *"quasi convexity"*. Let us denote $h'' = \frac{\partial^2 h}{\partial I_i \partial I_j}$; h is said to be quasi–convex in the set B, if for any $I \in B$

$$\begin{cases} \omega(I) \cdot v = 0 \\ h''(I) v \cdot v = 0 \end{cases} \qquad \Rightarrow \qquad v = 0 \ .$$

Quasi–convexity is an evident generalization of convexity (or concavity), that is $h''(I)v \cdot v > 0$ (< 0) for $v \neq 0$; an example of quasi–convex but not convex function is $h = I_1^2 + I_2^2 + I_3$. Quasi–convexity of h has two relevant consequences:

i) The plane of fast drift $\Pi_{\mathcal{L}}$ and the resonant manifold $M_{\mathcal{L}}$ are transversal (as in the figure). Indeed, a lack of transversality would imply that some vector $v \in \Pi_{\mathcal{L}}(I^o)$,

$$v = \sum_{\nu \in \mathcal{L}} c_\nu \, \nu \ , \tag{4.12}$$

is tangent to $M_{\mathcal{L}}$ in some point $I^* \in M_{\mathcal{L}}$. But since $\omega(I) \cdot v = 0$ on $M_{\mathcal{L}}$, the gradient $u(I) = h''(I)v$ of $\omega(I) \cdot v$ is orthogonal to $M_{\mathcal{L}}$ in $I \in M_{\mathcal{L}}$, and so $v \cdot u(I^*) = h''(I^*)v \cdot v = 0$, while simultaneously, according to (4.12), it is $\omega(I^*) \cdot v = 0$; this is against quasi–convexity.

ii) The unperturbed Hamiltonian h, restricted to $\Pi_{\mathcal{L}}(I^o)$, has a quadratic extremal in the (unique) intersection point I^* of $\Pi_{\mathcal{L}}(I^o)$ with $M_{\mathcal{L}}$. Indeed,

$$h(I^* + v) = h(I^*) + \omega(I^*) \cdot v + \frac{1}{2} h''(I^*) v \cdot v + \cdots \ ,$$

and for $v \in \Pi_{\mathcal{L}}$ the second term vanishes, while the second for quasi–convexity has a definite sign.

As a consequence, energy conservation provides confinement. Indeed, the "kinetic part" h in the normal form Hamiltonian (4.11), being extremal, has level surfaces in the form of concentric ellipsoids around I^*. Since the "potential part" $\varepsilon g + \varepsilon^2 f'$ has oscillation bounded by ε, the motion of the actions is consistently confined between two nearby level surfaces, and extends over a length scale of order $\sqrt{\varepsilon}$; see figure 26. The picture holds as far as the $\mathcal{O}(\varepsilon^2)$ remainder does not appreciably move the actions along the resonant manifold. In our case, due to the poor first order analysis we are performing, we can only guarantee that the transversal motion is bounded (say) by ε for $t \sim 1/\varepsilon$, but a higher order analytic work, leading to a smaller remainder, would produce a longer confinement.

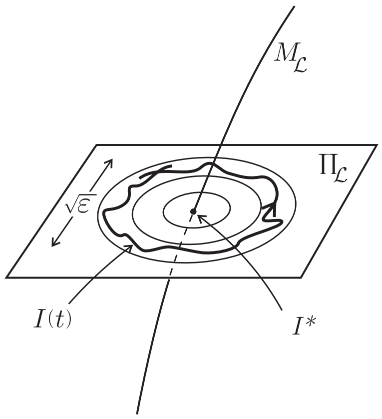

Figure 26: *The confinement of actions in the plane of fast drift for quasi–convex systems.*

Let us observe that the counterexample (2.11) to the averaging principle, introduced in section 2.3, has $h = \frac{1}{2}(I_1^2 - I_2^2)$, so that

$$\omega(I) = (I_1, -I_2) \ , \qquad h'' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ .$$

A lack of quasi–convexity then occurs on the lines $I_1 = \pm I_2$; indeed such lines are on the same time the resonant manifold and the "plane" of fast drift, which are no more transversal. Unbounded motions occur precisely along these lines.

Remark. Quasi–convexity is the simplest property leading to the confinement of the actions in the plane of fast drift. As widely discussed by Nekhoroshev [79], quasi–convexity can be generalized to a weaker property called "steepness". The notion of steepness is a little complicated, and the mechanism of confinement based on steepness is complicated as well (it requires a more detailed and global wiew of the phase space as decomposed into resonant regions). We shall spend very few words on the subject in section 5.3.

C. Slow and fast variables; scaling. Let us now proceed by considering a special resonance of multiplicity d, namely

$$\mathcal{L} = \left\{\nu \in \mathbb{Z}^n : \nu = (\nu_1, \ldots, \nu_d, 0, \ldots, 0)\right\} \ ; \qquad (4.13)$$

$M_{\mathcal{L}}$ is correspondingly defined by

$$\omega_j(I) = 0 \ , \qquad j = 1, \ldots, d \ .$$

The normal form Hamiltonian adapted to such a resonance is

$$H'(I', \varphi') = h(I') + \varepsilon g(I', \varphi'_1, \ldots, \varphi'_d) + \mathcal{O}(\varepsilon^2) \ ,$$

and the plane of fast drift $\Pi_{\mathcal{L}}(I^o)$ is the plane of constant $I_{d+1}, \ldots, I_n$. Following [26], we introduce a basic distinction of the variables into "slow" ones (S, σ) and "fast" ones (F, ψ), by posing

$$(S, \sigma) = (I'_1, \ldots, I'_d, \varphi'_1, \ldots, \varphi'_d) \ , \qquad (F, \psi) = (I'_{d+1}, \ldots, I'_n, \varphi'_{d+1}, \ldots, \varphi'_n) \ ;$$

the Hamiltonian is then

$$H'(S,F,\sigma,\psi) = h(S,F) + \varepsilon g(S,F,\sigma) + \mathcal{O}(\varepsilon^2) \ .$$

As far as the small remainder can be neglected, the fast actions become constant parameters, so we are left with a Hamiltonian with only d degrees of freedom. For each F, let $S^*(F)$ be implicitly defined by $I^* = (S^*(F),F)$, and move the origin of the slow actions to $S^*(F)$, by setting

$$S = S^*(F) + s \ . \tag{4.14}$$

As far as the fast actions are regarded as parameters, the translation is canonic. The domain of the s variables we are interested in, namely the domain where we wish the motion to be confined, is of size $\sqrt{\varepsilon}$. It is then worthwhile to bring this domain to order one by means of the canonical blow–up

$$s = \sqrt{\varepsilon}\,\hat{s} \ , \qquad t = \frac{\hat{t}}{\sqrt{\varepsilon}} \ , \qquad \hat{\mathcal{H}} = \varepsilon^{-1}\mathcal{H} \ ,$$

and then to expand $\hat{\mathcal{H}}$ in powers of $\sqrt{\varepsilon}$; the result is

$$\hat{\mathcal{H}}(\hat{s},\sigma) = \frac{1}{\varepsilon}h(S^*,F) + \frac{1}{\sqrt{\varepsilon}}\omega(S^*,F)\cdot\hat{s} + \frac{1}{2}h''(S^*,F)\hat{s}\cdot\hat{s} + g(S^*,F,\sigma) + \mathcal{O}(\sqrt{\varepsilon}) \ .$$

Now, $h(S^*,F)$ is a constant, while $\omega(S^*,F)\cdot\hat{s}$ exactly vanishes. The effective Hamiltonian governing the slow variables is then

$$\hat{\mathcal{H}}(\hat{s},\sigma) = \frac{1}{2}A_F\hat{s}\cdot\hat{s} + g_F(\sigma) + \mathcal{O}(\sqrt{\varepsilon}) \ ,$$

with

$$A_F = h''(S^*(F),F) \ , \qquad g_F(\sigma) = g(S^*(F),F,\sigma) \ .$$

This is a strongly coupled system, and for a generic choice of g_F (a generic choice of the initial perturbation f) the presence of chaotic motions is expected. (For example, it is easy to produce systems of this form containing homoclinic intersections [26]). Such possibly chaotic motions take place on a length scale and a time scale of order one, i.e. independent of ε; correspondingly, $(S,\sigma) = (I'_1,\ldots,I'_d,\varphi'_1,\ldots,\varphi'_d)$ perform possibly nontrivial motions on a length and a time scale respectively $\sqrt{\varepsilon}$ and $1/\sqrt{\varepsilon}$, while simultaneously the actions $I'_{d+1},\ldots,I'_n$ are stable and their conjugated angles $(\varphi'_{d+1},\ldots,\varphi'_n)$ regularly turn. The picture holds as long as the fast actions do not appreciably move, say for $|t| \ll \varepsilon^{-2}$. The original variables (I,φ) differ from (I',φ') only for the $\mathcal{O}(\varepsilon)$ deformation due to the canonical transformation, so the picture transports to them.

All of these considerations were developed in connection with a special resonance, characterized by the particular resonant lattice (4.13). But as we know from section 2.6, the example is generic, namely it is always possible to introduce a linear canonical transformation adapted to the resonance, such that the resonant lattice takes the form (4.13). In conclusion, inside *any* resonance of multiplicity d:

(i) The motion of the actions is flattened on the plane of fast drift and is there bounded in a domain of order $\sqrt{\varepsilon}$.

(ii) d suitably chosen degrees of freedom perform possibly nontrivial localized motions, on a time scale of order $1/\sqrt{\varepsilon}$, the other degrees of freedom remaining regular.

(iii) The picture holds (at least) for $|t| \ll 1/\varepsilon^2$, and only on longer times the system on the whole has a chance to evolve.

A better analytic part would produce longer confinement times, but not a better confinement, due to the natural width $\mathcal{O}(\varepsilon^{1/2})$ of the resonant regions. Such a picture is the natural generalization of the phenomenology we encountered when studying the model of rotators. The distinction of the variables into slow and fast ones, with different phenomena for the different variables on different time scales, is in a sense the heart of perturbation theory for nondegenerate anisochronous systems.

D. Infinitely many Fourier components. The last generalization we need is the extension of the results to a generic analytic f, having a possibly full Fourier series. *As is remarkable, both assumptions of Poincaré's theorem 2.6 are now satisfied.* The central idea here is to introduce, following Arnol'd [5], an "ultraviolet cut–off" N, and decompose f as the sum of an "infrared" and an "ultraviolet" part,

$$f = f^{\leq N} + f^{>N} \ , \tag{4.15}$$

with

$$f^{\leq N}(I,\varphi) = \sum_{|\nu|\leq N} f_\nu(I)\, e^{i\nu\cdot\varphi} \ , \qquad f^{>N} = f - f^{\leq N} \ . \tag{4.16}$$

The following easy lemma holds:

LEMMA 4.1 *If f is analytic in the domain $\mathcal{D}_\rho$ defined in (3.9), then*

$$\|f^{>N}\|_{\frac{1}{2}\rho} \leq e^{-\frac{1}{2}\rho_\varphi N} \|f\|_\rho \ .$$

The proof is trivial with the Fourier norm, but the inequality holds, up to a constant depending on n and a little worsening of the exponent, for the supremum norm, too. The use of the lemma is immediate: if we take N such that $e^{-\frac{1}{2}\rho_\varphi N} = \varepsilon$, i.e. $N(\varepsilon) \sim \log \varepsilon^{-1}$, then the ultraviolet part of the perturbation εf is already as small as ε^2, and there is no need to make any work on it. We can then proceed exactly as before and use, inside a generic resonance of multiplicity d,

$$\chi(I,\varphi) = \sum_{\nu\notin\mathcal{L};|\nu|\leq N} \frac{f_\nu(I,\varphi)}{i\nu\cdot\omega(I)}\, e^{i\nu\cdot\varphi} \ ;$$

the resulting normal form Hamiltonian is

$$H' = h + \varepsilon g + \varepsilon^2 f' + \varepsilon \tilde{f}^{>N} \ ,$$

where g is the projection of $f^{\leq N}$ onto $\mathcal{L}$, while

$$f' = \mathcal{R}_1[\chi, f^{\leq N}] + \mathcal{R}_2[\chi, h]$$

and $\tilde{f}^{>N} = f^{>N} \circ w$. The additional term $\varepsilon\tilde{f}^{>N}$ in the remainder is inessential, and the mechanism of confinement in resonant regions of size $\sqrt{\varepsilon}$ still works.[10] *The averaging principe is satisfied, and is not in conflict with Poincaré's theorem,* although both assumptions of the theorem are now satisfied. The result indeed is weaker: on the one hand, for $\varepsilon \to 0$ it is $N \to \infty$, so it is not possible to produce any resonant or nonresonant normal form, which remains valid in any given open set. On the other hand the remainder, though as small as ε^2, is not trivially proportional to ε^2 (in the sense of series expansions), as there pretended.[11]

Remark. This treatment is heuristic, but it is not worthwhile to make it rigorous, and produce now a precise mathematical statement on the confinement of the actions, since the time scale $t \ll \varepsilon^{-2}$ of the confinement is very poor compared to the possibilities of perturbation theory. The only purpose here was to introduce the geometric "elements" entering the theory — the different resonant regions, the planes of fast drift, the role of convexity, the slow and fast variables and their scales, the ultraviolet cut–off — in the simplest situation, almost free of analytic work. A precise mathematical setting and the top results of perturbation theory (KAM theorem, Nekhoroshev theorem) are deferred to the next section.

5 Infinitely many steps

There are two main results, in Hamiltonian perturbation theory, which are based on the iteration of infinitely many (or unboundedly many) perturbation steps, namely the Kolmogorov–Arnol'd–Moser (KAM) theorem [66] [5] [75] and the Nekhoroshev theorem [79]. KAM theorem focuses entirely the attention on nonresonant motions, and states the perpetual stability of the actions in a nontrivial large subset of the phase space, of Lebesgue measure close to one though with empty interior. The theorem does not require any geometric assumption like convexity, and concerns general analytic[12] Hamiltonians of the form

$$H(I,\varphi) = h(I) + \varepsilon f(I,\varphi) \ , \qquad (I,\varphi) \in D = B \times \mathbb{T}^n \ , \tag{5.1}$$

[10] There is in fact an additional difficulty, that we are disregarding in this heuristic treatment: namely for $\varepsilon \to 0$ the cut-off N goes (slowly) to infinity, and correspondingly the resonant manifolds get dense; in such conditions some care must be paid, to be sure that motions starting inside a resonant region, but close to a nearby one, in the course of time do not enter the latter. We do not discuss the question here, but we shall take it into consideration in the (sketch of) proof of the Nekhoroshev theorem that we shall produce in the next section.

[11] One might think there is a third difference, namely inside resonances the normal form is not as strong as the nonresonant normal to which Poincaré's theorem refers (the term of order ε is not completely independent of the angles). This point however is not essential, and it would be easy to generalize Poincaré's theorem, so as to forbid even a partial dependence of the term of order ε on the angles.

[12] There exist differentiable versions of the theorem, but we shall not consider them.

with $\det h''(I) \neq 0$ in B. Nekhoroshev theorem provides a result in a sense complementary: perpetual stability is lost and replaced by long time stability, but the result holds uniformly in the phase space, including resonances. The necessity of working out stability results also inside resonances produces some relevant differences with respect to KAM theorem. A geometric assumption on h, like convexity, becomes mandatory (this we already know from the first order work of section 4.3); moreover, while KAM theorem is essentially local in the action space, Nekhoroshev theorem instead requires a more global geometric view. In fact, Nekhoroshev theorem in its essence should be considered as a kind of geometric theory of resonant motions. The role of geometry gets particularly evident in some extensions of the theory to "degenerate" Hamiltonians, as we shall discuss in section 6.

The next two subsections are devoted to KAM theorem, that we shall introduce both in the Kolmogorov and in the Arnol'd version. Both versions are indeed interesting, from our point of view, for the different stategy they adopt to escape the Poincaré difficulty. (Let us recall that Hamiltonian (5.1) satisfies both assumptions of proposition 2.6, and so the aimed result, namely a canonical transformation defined in some open subset of the phase space, such that the new Hamiltonian loses the dependence on the angles, must be weakened.) We shall limit ourselves to the statements, without entering the not easy proofs. The third subsection is instead devoted to the Nekhoroshev theorem, for which, thanks to the preparatory work of the previous sections, we are able to provide, though not a complete proof, at least an account of it.

5.1 KAM theorem (according to Kolmogorov)

Let $I^* \in B$ and denote

$$\omega^* = \omega(I^*) \ , \qquad \Gamma = h''(I^*) \ , \qquad H^*(J,\varphi) = H(I^* + J, \varphi) \ .$$

A Taylor development gives

$$H^*(J,\varphi) = \mathrm{const} + \omega^* \cdot J + \frac{1}{2}\Gamma\, J \cdot J + \varepsilon\big[A(\varphi) + B(\varphi)\cdot J + \frac{1}{2}C(\varphi)J\cdot J\big] + \mathcal{O}(\|J\|^3) \ , \tag{5.2}$$

where of course

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

If A and B in (5.2) vanish, the Hamiltonian H^* is said to be in *"Kolmogorov weak normal form"*. That is: the dependence on the angles is not completely eliminated—the system is not integrable—but is eliminated from terms of order zero and one in J. In such a case the Hamilton equations assume the form

$$\dot J = \mathcal{O}(\|J\|^2) \ , \qquad \dot\varphi = \omega^* + \mathcal{O}(\|J\|) \ ,$$

and so the Hamiltonian H^* has the family of solutions (parametrized by φ^o)

$$J(t) = 0 \ , \qquad \varphi(t) = \varphi^o + \omega^* t \ ;$$

correspondingly, H admits quasi–periodic motions, with frequency ω^*, on the invariant torus $I = I^*$. The idea of Kolmogorov is that A and B can be reduced to zero by means of a convenient canonical transformation, whenever ω^* satisfies Siegel's arithmetic condition (see section 3.2) and the perturbation is sufficiently small. Precisely:

PROPOSITION 5.1 *Let H be as in (5.1) and suppose it be analytic in a complex neighborhood $\mathcal{D}_\rho = \mathcal{B}_\rho \times \mathcal{S}_\rho$ of $D = B \times \mathbb{T}^n$ (see section 3), with some $\rho = (\rho_I, \rho_\varphi)$. Let $I^* \in B$ be such that $\omega^* = \omega(I^*)$ is (γ, n)–Diophantine and $\Gamma = h''(I^*)$ is invertible, and define $H^*(J, \varphi)$ as above.*

There exists a positive constant $\mathcal{F}$ depending on n, ρ, γ, $\|\omega^*\|$, $\|\Gamma^{-1}\|$, actually proportional to $\gamma^4\|\Gamma^{-1}\|$, such that, if $\varepsilon\|f\|_\rho^\infty < \mathcal{F}$, then there exists a neighborhood B_0 of $J = 0$ and a near the identity canonical transformation $(J, \varphi) = w(J', \varphi')$, defined in $B_0 \times \mathbb{T}^n$, such that the new Hamiltonian $H' = H^* \circ w$ is in Kolmogorov weak normal form:

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

Γ' being a constant matrix close to Γ.

It follows that

COROLLARY 5.2 *For each Diophantine ω^*, if the norm of the perturbation $\varepsilon\|f\|_\rho^\infty$ is sufficiently small, the original Hamiltonian (5.1) has an invariant torus, close to the unperturbed torus $I^* \times \mathbb{T}^n$, supporting quasi–periodic motions with frequency ω^*. For given small ε, the set of invariant tori has Lebesgue measure differing from one, at most, by quantities of order $\varepsilon^{1/4}$.*

Kolmogorov's original paper is rather short and omits many details, but the scheme of the proof, including all difficulties and how overcoming them, is complete; working out a detailed proof on the basis of Kolmogorov's indications is basically an exercise [9] [25]. More recently, Kolmogorov's proof has been revisited with an essentially new idea, and quite interesting improvements have been obtained [55] [56] [57].

5.2 KAM theorem (according to Arnol'd)

Here we skip the original formulation of the theorem as produced by Arnol'd, and refer instead to the later formulation due to Lazutkin [68], Chierchia and Gallavotti [34] and Pöschel [83]. The aim is the usual one, namely constructing a canonical transformation w which conjugates $H(I, \varphi)$ to a normal form Hamiltonian $h'(I)$, so as the dependence on the angles is completely removed. The idea is to escape the difficulties arising from Poincaré's theorem by making weaker not the normal form, as in Kolmogorov's, but the notion of conjugacy. A possible formulation of the theorem is the following:

Proposition 5.3 *Let H be as in (5.1); assume:*

i) *H is analytic in a complex neighborhood $\mathcal{D}_\rho = \mathcal{B}_\rho \times \mathcal{S}_\rho$ of $D = B \times \mathbb{T}^n$ (see section 3), with some $\rho = (\rho_I, \rho_\varphi)$.*

ii) *h'' is non degenerate, namely $\det h''(I) \neq 0$ in B.*

iii) *ε is small, namely smaller than a suitable threshold ε_0 depending on n, ρ, $\|h\|_\rho^\infty$, $\|f\|_\rho^\infty$, $\|(h'')^{-1}\|_\rho^\infty$.*

Then there exist:

(a) *A near the identity canonical transformation $(I, \varphi) = w(I', \varphi')$, of class C^∞, defined in the real domain D.*

(b) *A C^∞ integrable Hamiltonian h', defined in B.*

(c) *A subset $B_\varepsilon \subset B$ of large measure,*

$$\mathrm{mes}\, B \setminus B_\varepsilon \sim \varepsilon^{1/2} \ ,$$

such that the new Hamiltonian $H'(I', \varphi') = H(w(I', \varphi')) : D \to \mathbb{R}$ coincides with h' for $I' \in B_\varepsilon$:

$$H'(I', \varphi') \stackrel{B_\varepsilon}{=} h'(I') \ ;$$

the symbol $\stackrel{B_\varepsilon}{=}$ means indeed that the equality holds for $I' \in B_\varepsilon$, there extending also to all derivatives. Moreover, the derivative $\omega = \frac{\partial h'}{\partial I'}$ maps the set B_ε into the set of the (γ, n)–Diophantine frequencies, with $\gamma \sim \varepsilon^2$.

Quite clearly, tori $\{I'\} \times \mathbb{T}^n$, for $I' \in B_\varepsilon$, are left invariant by the Hamiltonian flow $\Phi^t_{H'}$, and motions on them are quasi–periodic with Diophantine frequency. Altogether, these invariant tori form a set $D_\varepsilon = B_\varepsilon \times \mathbb{T}^n$ of large Lebesgue measure. In the original variables one has a similar set of invariant tori — namely the image $w(D_\varepsilon)$ of D_ε — of the same measure, each torus being close to the unperturbed one supporting motions with the same frequency. This is usually reported by saying that "Diophantine tori" are not destroyed by a sufficiently small perturbation, but only slightly deformed.

The set B_ε, as constructed in the theorem, has large measure, but is closed with empty interior (it does not contain any open set), so the notion of conjugacy used in the theorem is essentially weaker than the usual one. Correspondingly, the complement $B \setminus B_\varepsilon$ is open and dense, and so are $D \setminus D_\varepsilon$ as well as its image $w(D \setminus D_\varepsilon)$. For $n = 2$, two–dimensional invariant tori separate the three–dimensional surfaces of constant energy (where of course the motion is confined) into thin strips:[13] motions starting inside a strip cannot escape. Correspondingly, the actions are perpetually stable for *all* initial data, including resonant

[13] Properly speaking, in order for the set of invariant tori has large measure in *each* of the energy surfaces, and not only in most of them, the following "isoenergetic nondegeneracy

ones. The result applies for example to the stability of the Lagrange top, if the perturbing potential is symmetric for proper rotations: indeed the symmetry allows to reduce from three to two the number of degrees of freedom, and then the theorem applies. Another application is the planar circular restricted three body problem, see [3] [6].

If instead $n \geq 3$, then the "bad" set $B \setminus B_\varepsilon$ (as constructed in the theorem) is not only open and dense, but also connected, and so the stability of the actions is no more guaranteed. Arnol'd constructed a very special model [4] for which KAM theorem applies, and nevertheless, *given any two open balls on any energy surface, there exists a motion connecting them.* Such a phenomenon is called *Arnol'd diffusion.* Understanding it, and proving it does exist for some interesting class of Hamiltonian systems, is among the most interesting but hardest challenges in Hamiltonian dynamics.

Remark. Propositions 5.1 and 5.3 differ in an interesting point, namely the relation between the Diophantine constant γ and ε is $\gamma \sim \varepsilon^{1/4}$ in the former case, $\gamma \sim \varepsilon^{1/2}$ in the latter. Correspondingly, the measure of the invariant tori is one up to quantities proportional to, respectively, $\varepsilon^{1/4}$ and $\varepsilon^{1/2}$. Proposition 5.1 is, from this respect, more efficient.

5.3 Nekhoroshev theorem

A possible statement of the theorem, adapted to the quasi–convex case, is the following:

PROPOSITION 5.4 *Consider a nearly–integrable Hamiltonian system with* n *degrees of freedom*

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

and assume:

i) H *is analytic in a complex neighborhood* $\mathcal{D}_\rho = \mathcal{B}_\rho \times \mathcal{S}_\rho$ *of* D, $\rho = (\rho_I, \rho_\varphi)$.

ii) h *is* (l,m)*–quasi–convex in* B*, namely the system*

$$|\omega(I) \cdot v| < l\|v\| \ , \qquad |h''(I)v \cdot v| < m\|v\|^2$$

has no solution for $I \in B$.

Denote $\omega_0 = \inf_{I \in \mathcal{B}_\rho} \|\omega(I)\|$. *There exist positive constants* ε^*, $\mathcal{I}$, $\mathcal{T}$ *depending on* ρ, l, m, ω_0, $\|h\|_\rho$, $\|f\|_\rho$, *and positive constants* a, b *depending only on* n,

condition":

$$\det \begin{pmatrix} & & & \omega_1 \\ & h''_{ij} & & \cdots \\ & & & \omega_n \\ \omega_1 & \cdots & \omega_n & 0 \end{pmatrix} \neq 0$$

needs to be satisfied.

such that any initial motion $(I(t),\varphi(t))$ *with initial datum* $(I^o,\varphi^o)\in D$ *satisfies the exponential estimate*

$$\|I(t)-I^o\| < \mathcal{I}\Big(\frac{\varepsilon}{\varepsilon^*}\Big)^b \qquad \text{for} \quad |t| < \mathcal{T} e^{(\varepsilon^*/\varepsilon)^a} \ . \tag{5.3}$$

Possible values of a *and* b *are*

$$a=b=\frac{1}{2n} \qquad \textit{as well as} \qquad a=\frac{1}{4n}\ ,\ b=\frac{1}{4}\ . \tag{5.4}$$

The theorem was stated and proved in [79], in the assumption that h is "steep" (this is weaker than quasi–convexity). A proof focused on the quasi–convex case was later produced in [24] and [26]. A revolutionary proof adapted to the quasi–convex case, leading to the above improved (likely optimal) values $a=b=1/n$ of the exponents, appeared in [70] [72]; the result, with some more freedom in the exponents as in (5.4), was then recovered within the traditional way of proof in [84]. The steep case was recently revisited in [82], with likely optimal values of the exponent a and b. A recent review paper, focused on arithmetic questions, is [71].

A complete proof of the theorem would be very long and annoying, for the numerous details and estimates entering it. But thanks to the work done in the previous sections, we have all the necessary elements to provide, at least, a reasonable sketch of it.

Proof (sketch). The proof puts together a geometric part, not so different from the analysis we made in section 4.3 to study the motion inside resonances at first order in ε, and an analytic part, not so different from the work we made in section 3 to work out exponential estimates in the isochronous case.

(a) *The geometric part (also called the "geography of resonances").* We introduce a cut–off

$$N \sim \varepsilon^a \tag{5.5}$$

and decompose the perturbation into the infrared and ultraviolet parts, $f = f^{\le N} + f^{>N}$, as in section 4.3D. According to Lemma 4.1, $f^{>N}$ is already exponentially small, namely as small as $\exp -\frac{1}{2}N\rho_\varphi$, so we can disregard it and take care only of $f^{\le N}$, that is only of a finite number of Fourier components. To this end, we consider all the d–dimensional sublattices $\mathcal{L}$ of $\mathbb{Z}^n$, $1\le d\le n$, which contain a basis $\{\nu^{(1)},\ldots,\nu^{(d)}\}$ such that $|\nu^{(j)}|\le N\ \forall j$. (We can restrict the attention to the "maximal" lattices, namely those lattices which are not properly contained into lattices of the same dimensionality.) The cells of any $\mathcal{L}$ have a minimal (Euclidean) volume $\|\mathcal{L}\|\ge 1$.

We then introduce a convenient covering of the action space B by "resonant zones", in which the frequency $\omega(I)$ has well defined resonance properties. Precisely, for each lattice $\mathcal{L}$ of dimension $d\ge 1$ we introduce (see figure 27):

- The "resonant manifold"

$$M_{\mathcal{L}} = \big\{ I\in B : \nu\cdot\omega(I)=0 \quad \forall \nu\in\mathcal{L}\big\}\ ,$$

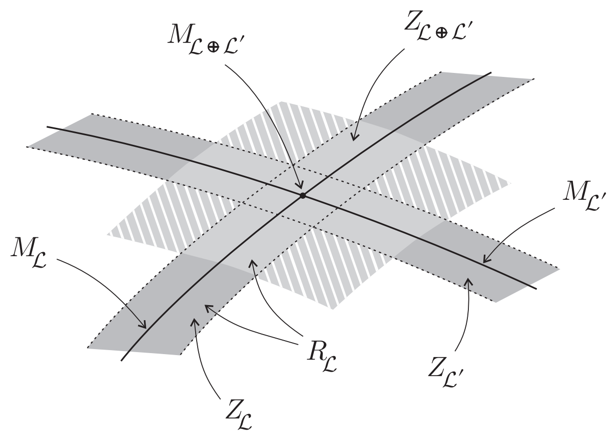

Figure 27: *Resonant manifolds, resonant regions and resonant zones entering the "geography of resonances". The four dashed corners belong both to* $\mathcal{Z}_{\mathcal{L}\oplus\mathcal{L}'}$ *and to* $\mathcal{Z}_0$.

actually a manifold[14] of codimension d.

- The "resonant region" $\mathcal{R}_\mathcal{L} \supset M_\mathcal{L}$, where the orthogonal projection $\omega_\mathcal{L}$ of ω onto $\mathcal{L}$ is small:

$$\mathcal{R}_\mathcal{L} = \left\{ I \in B : \|\omega_\mathcal{L}(I)\| \le \lambda_d / \|\mathcal{L}\| \right\} , \qquad \lambda_d = C/(cN)^{n-d+1} ,$$

  C, c being convenient positive constants. In $\mathcal{R}_\mathcal{L}$ the divisors $\omega \cdot \nu$, with $\nu \in \mathcal{L}$, are small.

- The "resonant zone" $\mathcal{Z}_\mathcal{L} \subset \mathcal{R}_\mathcal{L}$, namely

$$\begin{aligned} \mathcal{Z}_\mathcal{L} &= \mathcal{R}_\mathcal{L} && \text{for} \quad \mathcal{L} = \mathbb{Z}^n \\ \mathcal{Z}_\mathcal{L} &= \mathcal{R}_\mathcal{L} \setminus \bigcup_{\mathcal{L}' : \dim \mathcal{L}' = d+1} \mathcal{R}_{\mathcal{L}'} && \text{for } d = \dim \mathcal{L} < n \end{aligned}$$

  In $\mathcal{Z}_\mathcal{L}$ all divisors $\omega(I) \cdot \nu$, for $\nu \notin \mathcal{L}$ and within the cut–off, are bounded away from zero. We also introduce the "nonresonant zone"

$$\mathcal{Z}_0 = B \setminus \bigcup_{\mathcal{L} : \dim \mathcal{L} = 1} \mathcal{R}_\mathcal{L} .$$

By lowering ε the cut–off N raises, and resonant manifolds get dense, but *any two resonant regions* $\mathcal{R}_\mathcal{L}$ *and* $\mathcal{R}_{\mathcal{L}'}$*, relative to different* $\mathcal{L}, \mathcal{L}'$ *of the same dimensionality, do not overlap and stay well separated;* it is essential for this that $\lambda_{d+1} \gg \lambda_d$. (Zones of different dimensionality can instead overlap, see figure 27, but this will be irrelevant.)

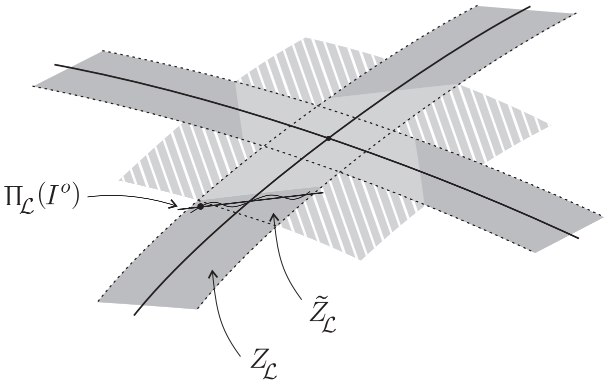

Figure 28: *The plane of fast drift* $\Pi_{\mathfrak{L}}(I^o)$*; the extended resonant zone* $\tilde{\mathfrak{Z}}_{\mathfrak{L}}$.

The idea is that motions should be considered as resonant with $\mathfrak{L}$, if the initial datum I^o belongs to the zone $\mathfrak{Z}_{\mathfrak{L}}$ (if I^o belongs to more than one zone, just take the one of minimal dimensionality). Motions resonant with $\mathfrak{L}$ will obey a normal form adapted to $\mathfrak{L}$, in which all Fourier components of the perturbation but those of $\mathfrak{L}$ are eliminated up to an exponentially small remainder (see below the analytic part). As we already know from the first order analysis of section 4.3, such motions are flattened on the plane of constant drift $\Pi_{\mathfrak{L}}(I^o)$ and, thanks to convexity, are there bounded, namely the distance from the intersection I^* of $\Pi_{\mathfrak{L}}(I^o)$ and $M_{\mathfrak{L}}$ cannot substantially grow. For consistency, the normal form adapted to $\mathfrak{L}$ needs to be valid not only inside the zone $\mathfrak{Z}_{\mathfrak{L}}$, but also in the "extended zone" $\tilde{\mathfrak{Z}}_{\mathfrak{L}}$, namely the zone that can be reached by starting in $\mathfrak{Z}_{\mathfrak{L}}$, moving along $\Pi_{\mathfrak{L}}(I^o)$ as long as the normal form permits it; see figure 28. If the constant c entering the construction of the zones is sufficiently large, then it can be proven [84] that for $I \in \tilde{\mathfrak{Z}}_{\mathfrak{L}}$ it is

$$|\omega(I)\cdot\nu| \geq \gamma_{\mathfrak{L}} = \frac{\lambda_d}{2\|\mathfrak{L}\|} \tag{5.6}$$

for any $\nu \in \mathbb{Z}\setminus\mathfrak{L}$, $|\nu| \leq N$. The inequality is not trivial, since for large N (small ε) the angle between different $\mathfrak{L}$ and $\mathfrak{L}'$ can get small; think for example to the one dimensional lattices generated by $\nu = (N,1,0,\ldots,0)$ and $\nu' = (N-1,1,0,\ldots,0)$. To take care of this possibility, it is necessary to take $\lambda_{d+1} \gg \lambda_d$ and moreover to take thinner regions for larger $\|\mathfrak{L}\|$.

(b) *The analytic part (normal forms with exponentially small remainders).* For each resonant lattice $\mathfrak{L}$ we must produce a canonical transformation $(I,\varphi) = w_{\mathfrak{L}}(I',\varphi')$, such that the new Hamiltonian $H_{\mathfrak{L}} = H\circ w_{\mathfrak{L}}$ has the form

$$H_{\mathfrak{L}}(I',\varphi') \;=\; h(I') + \varepsilon g_{\mathfrak{L}}(I',\varphi') + \text{remainder}$$

[14]From quasi–convexity it follows that the determinant of h'' restricted to the plane orthogonal to ω is different from zero. In turn, this implies that the r equations defining $M_{\mathfrak{L}}$ are independent.

$$g_{\mathcal{L}}(I',\varphi') \;=\; \sum_{\nu\in\mathcal{L}} g_{\mathcal{L},\nu}(I')\, e^{i\nu\cdot\varphi'} \ .$$

The domain of the transformation needs to be (a small neighbourhood of) $D_{\mathcal{L}} = \tilde{\mathcal{Z}}_{\mathcal{L}}\times\mathbb{T}^n$. The canonical transformation is constructed as in the isochronous case (section 3) by a sequence of r elementary steps, and a convenient choice, see below, is $r\sim N$. The construction of $w_{\mathcal{L}}$ starts, as in the isochronous case, by extending in the complex the domain $D_{\mathcal{L}}$; then at each step the domain is reduced, as in section 3, by a thin strip, in such a way that after r steps the original extension gets reduced, say, by a factor two. A crucial point, new with respect to the isochronous case, is that the extension of the actions now needs to be small, otherwise the control of the small divisors in the extended domain is lost. But it is easy to check that if c is sufficiently large, namely $c\geq M=\sup_{I\in B}\|h''(I)\|$, and the extension does not exceed λ_d, then in the extended set it is

$$|\omega(I)\cdot\nu| \geq \gamma_{\mathcal{L}} - MN\lambda_d \geq \frac{1}{2}\gamma_{\mathcal{L}} \ ,$$

that is, denominators worsen at most by an irrelevant factor two. In a typical step, the input is a Hamiltonian of the form

$$H^{(s)}_{\mathcal{L}} = h(I) + \varepsilon g^{(s)}_{\mathcal{L}} + \varepsilon^{s+1} f^{(s)}(I,\varphi) \ , \qquad g^{(s)}_{\mathcal{L}}(I,\varphi) = \sum_{\nu\in\mathcal{L}} g^{(s)}_{\mathcal{L},\nu}(I)\, e^{i\nu\cdot\varphi} \ ;$$

the output is a Hamiltonian of the same form, with $s+1$ in place of s. The game is over if $f^{(s)}$ does not grow, passing to $s+1$, by a factor larger than $\Lambda=(e\varepsilon)^{-1}$: indeed in such a case (precisely as for the isochronous case) the final remainder is as small as $(\varepsilon\Lambda)^r = e^{-r}$, that is, thanks to $r\sim N$, exponentially small as the the ultraviolet part $f^{>N}$ of the perturbation. This explains the choice $r\sim N$. A rough estimate of Λ is not difficult: forgetting constants of order one, Λ contains essentially a large factor $\gamma_{\mathcal{L}}^{-1}$, due to the small divisors, and a large factor $(\lambda_d/N)^{-1}$, due to a Cauchy estimate with domain reduction $\lambda_d/r\sim\lambda_d/N$. (A clever work needs only one Cauchy estimate per step, to estimate just one Poisson bracket.) The condition $\varepsilon\Lambda = \mathcal{O}(1)$, which needs to be valid for all $\mathcal{L}$, fixes then N as a function of ε, and as a result, the exponent a entering (5.5) turns out to be $a=1/(2n)$. The final remainder is then as small as $\exp -1/\varepsilon^a$.

(c) *Conclusion.* The proof is essentially over: as long as the remainder does not move appreciably the actions, that is up to $t\sim\exp N\sim\exp 1/\varepsilon^a$, the normal form forces the actions to be flattened on the plane of constant drift $\Pi_{\mathcal{L}}(I^o)$. Energy conservation assures that the motion on such a plane is bounded, namely the distance of $I(t)$ from $I^* = \Pi_{\mathcal{L}}(I^o)\cup M_{\mathcal{L}}$ cannot grow too much, and as is crucial, the normal form remains valid during this motion (no new small divisors are added). The maximal distance covered by $I(t)$ is bounded (up to factors of order one) by the size of the resonant region $\mathcal{R}_{\mathcal{L}}$, measured along the plane of fast drift, and in the worst case ($d=n$) this is of the order of $\lambda_n\sim N^{-1}$. This determines the exponent b entering the exponential law (5.3), namely $b=a=1/(2n)$. (The alternative choice of a and b, which appears in

(5.4), is obtained by letting some constants, that we considered of order one and disregarded, to depend instead on ε.) □

Remark. As in section 4.3B, we invoked the energy conservation as the mechanism confining the motion on the plane of fast drift. An alternative idea, equivalent in the quasi–convex case but more general, is the following: look again at figure 28. If the geography of resonances is well arranged as above, than any motion with $I^o \in \mathcal{Z}_{\mathcal{L}}$, proceeding along the plane of constant drift, can leave $\tilde{\mathcal{Z}}_{\mathcal{L}}$, but only through the "lateral walls", and in this way it necessarily enters a less resonant zone $\mathcal{Z}_{\mathcal{L}'}$, with $\mathcal{L}' \subset \mathcal{L}$, $\dim \mathcal{L}' < \dim \mathcal{L}$ (the nonresonant zone, in the figure). But here a stronger normal form is valid, and the game repeats, till, in the worst case, the trajectory enters $\mathcal{Z}_0$, where the nonresonant normal form holds and the actions cannot further evolve.[15] Such a mechanism, actually exploited by Nekhoroshev in his original proof of the theorem, looks more complicated than the elementary one based on the energy conservation. But it is more deep, and moreover it generalizes to the nonconvex "steep" cases, for which instead the energy conservation does not help any more. The notion of "steepness", unfortunately, is complicated, and since we shall not use it in these lectures, we shall not introduce it. After the Nekhoroshev paper [79], nonconvex steep cases have been considered in [63] and [18], in connection with applications to Celestial Mechanics. The general steep case has been recently revisited, with substantial improvements, in [82]. For a quick account of the notion of steepness, see [11].

6 Nekhoroshev theory and the Euler perturbed rigid body

6.1 Degeneracy and presence of singularities in physical systems

Nekhoroshev theorem, in its general form 5.4, is often unsuited to be applied to interesting physical systems. Two pathologies are frequently met:

(a) *The unperturbed system is "superintegrable", or "properly degenerate".* For such systems the number n_1 of constants of motion exceeds the number n of degrees of freedom, so that the unperturbed motion does not really take place on n–dimensional tori but is restricted to manifolds of dimension $n_0 = 2n - n_1$ which, in typical assumptions [80] [73] [43], turn out to be tori $\mathbb{T}^{n_0}$. Correspondingly, the number of the actions effectively entering

[15] Such a "trapping mechanism" is apparently in conflict with the temporal symmetry of Hamiltonian systems. A more careful reasoning, in fact, avoids the asymmetry: indeed, even the exit from $\tilde{\mathcal{Z}}_{\mathcal{L}}$ to $\mathcal{Z}_{\mathcal{L}'}$ turns out to be forbidden, otherwise the Hamiltonian $-H$ (different from H for all motions are reversed) would have a motion acquiring, instead than losing, a resonance, and this is impossible since $-H$ also satisfies all of the assumptions of Nekhoroshev theorem.

the unperturbed Hamiltonian h is $n_0 < n$. Using the notation (I,φ) for the n_0 "nondegenerate" degrees of freedom entering h, and (p,q) for the remaining $n - n_0$ "degenerate" ones, the perturbed Hamiltonian has the form

$$H(I,\varphi,p,q) = h(I) + \varepsilon f(I,\varphi,p,q) \ . \tag{6.1}$$

Important examples are: (i) The Euler-Poinsot rigid body (the rigid body with a fixed point, in absence of external torques). Such a system has $n = 3$ degrees of freedom and $n_1 = 4$ constants of motion; only $n_0 = 2$ actions enter h, and motions take place on two–dimensional tori. The degenerate variables represent the orientation of the angular momentum in the space (due to the absence of extrenal torques, h cannot depend on them). This is the example we shall discuss in detail below. (ii) The Kepler problem, which is doubly degenerate, with $n = 3$ and $m = 5$. Only one action (actually a function of the semi–major axis) enters h and all motions take place on one-dimensional manifolds, i.e. are periodic (here we are limiting ourselves to bounded motions). The degenerate variables represent the eccentricity and the inclination of the orbit and the related angles, namely the *anomaly of the pericenter* and the *anomaly of the ascendent node.*

The methods of perturbation theory easily adapt to Hamiltonians of the form (6.1), basically as if the degenerate variables (p,q) would not exist. More precisely, inside a resonant zone it is not difficult to work out a normal form Hamiltonian

$$\begin{aligned} H_{\mathcal{L}}(I',\varphi',p',q') &= h(I') + \varepsilon g_{\mathcal{L}}(I',\varphi',p',q') + \text{small remainder} \\ g_{\mathcal{L}}(I',\varphi',p',q') &= \sum_{\nu\in\mathcal{L}} g_\nu(I',p',q')\, e^{i\nu\cdot\varphi'} \ , \end{aligned}$$

and if h is convex, then the mechanism of confinement of the actions does apply. But the result risks to be useless, since nothing is known, already on the time–scale $1/\varepsilon$, on the degenerate variables (p,q). Indeed, on the one hand these variables are in general very significant, so it is rather disappointing to have no informations on them. On the other hand, precisely because such variables are soon out of control, there is a chance they escape their domain of definition, for example because they meet a singularity; from that moment on, all informations on the actions I, too, are lost. The problem of the control of the degenerate variables is not technical but substantial: for example, if the perturbation is separated, that is

$$H(I,\varphi,p,q) = h(I) + \varepsilon[f_1(I,\varphi) + f_2(p,q)] \ ,$$

then it is obvious that the behavior of p, q depends entirely on f_2 and, in lack of further informations on the perturbation, anything could happen in a time $1/\varepsilon$.

(b) *The action–angle variables have singularities.* When we introduced the Liouville–Arnol'd theorem, we already observed that the foliation of the phase space into invariant tori in general has singularities. This occurs, for example, for a set of n harmonic oscillators, if any of them is at rest: a family of invariant tori $\mathbb{T}^n$ there collapses onto a lower dimensional torus. A similar situation occurs for the central motion, in correspondence with the circular orbits, and for the Euler–Poinsot rigid body, which has similar singularities on the proper rotations. These examples show that singularities frequently occur in connection with important motions, which cannot be ignored. Whenever a singularity of the foliation is met, the Hamiltonian is no more an analytic function of the actions, but only of their square roots, and so the assumptions of analyticity of the Hamiltonian, that we used throughout our treatment, are no longer satisfied. Fortunately, the difficulty is only technical and not really serious: as we shall see in the example of the rigid body, it is possible to introduce, in place of the polar–like action–angle coordinates, convenient Cartesian–like coordinates which remove the singularity; this allows to proceed, with minor changes, by the usual tools of perturbation theory.

We shall not discuss at a general level these questions, rather we shall discuss in detail one example, namely the perturbed Euler–Poinsot rigid body, for which both difficulties arise.

6.2 The Euler–Poinsot rigid body

A. The Poinsot description ot the motion. Here we recall some important facts concerning the unperturbed system, namely the Euler–Poinsot rigid body. Let (O, e_x, e_y, e_z) be any arbitraryly chosen "fixed" or "spatial" frame, with origin in the fixed point of the body, and (O, e_1, e_2, e_3) be a "proper" frame, with the same origin and axes parallel to the principal axes of inertia of the body. The configuration of the system is specified by a matrix $R \in SO(3)$, such that $e_1 = Re_x$ and so on. The dynamical state is then fully specified by assigning R and the angular momentum $m = (m_x, m_y, m_z) \in \mathbb{R}^3$. This means that the phase space of the system is $M = \mathbb{R}^3 \times SO(3)$. Such a situation is exceptional: in general, if the configuration space is a manifold Q of dimension n, its phase space is the cotangent bundle T^*Q, which in general is not trivial (is not diffeomorphic to $\mathbb{R}^n \times Q$). But whenever Q is a group, like $SO(3)$ in our case, then both the tangent and the cotangent bundles are trivial,[16] and in particular $T^*SO(3)$ is diffeomorphic to $\mathbb{R}^3 \times SO(3)$.

[16] Compare with the case of a point constrained on a sphere S^2. For any $q \in S^2$, the velocity v stays in the tangent plane $T_q S^2$, and there is no way to identify all of these two dimensional planes with a unique Euclidean space $\mathbb{R}^2$. That is: the fiber bundle TS^2 is not trivial (the same occurs to the cotangent bundle T^*S^2). If instead Q is a group, and $q_0 \in Q$ is any of its elements (for example the identity), then for each $q \in Q$ the element $\tilde{q} = q_0 q^{-1} \in Q$ is a map $Q \to Q$ which sends q into q_0; correspondingly the tangent map $D\tilde{q}$ sends the tangent space $T_q Q$ into $T_{q_0} Q$. All tangent spaces then identify diffeomorphically with $T_{q_0} Q$ and so the tangent bundle TQ is diffeomorphic to $\mathbb{R}^n \times Q$. The same happens for T^*Q. The well known elementary results that the angular velocity Ω and the angular momentum m of a rigid body do exist as vectors of $\mathbb{R}^3$, are exactly equivalent to say, in geometric language, that $TSO(3)$

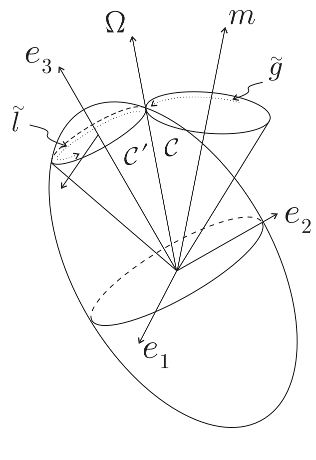

Figure 29: *The motion of the Euler–Poinsot rigid body.*

The system has three degrees of freedom and four constants of motion, namely the kinetic energy K and the three components of the angular momentum. To review the dynamics and see that motions are quasi periodic with two frequencies, we shall restrict ourselves to the simpler case of a symmetric body, more precisely a body with moments of inertia $A_1 = A_2 \neq A_3$ (but the resulting picture, up to minor points, is general). In such a case, denoting by m_1, m_3, m_3 the components of m in the proper frame, the kinetic energy is given by

$$K = \frac{(m_1^2 + m_2^2)}{2A_1} + \frac{m_3^2}{2A_3} = \frac{\|m\|^2}{2A_1} + \Big(\frac{1}{2A_3} - \frac{1}{2A_1}\Big) m_3^2 \ , \tag{6.2}$$

so the conservation of K is equivalent to the conservation of m_3, and the latter constant of motion is easier to use. Moreover, the motion is as follows (Poinsot description): the angular momentum m, the angular velocity Ω and the symmetry axis e_3 are coplanar, have fixed length and form constant angles (figure 29); the vector m stays fixed in space, while Ω and e_3 preceed with constant speed around it. More precisely, assuming that e_3 and m are not parallel, one can draw two circular cones $\mathcal{C}$ and $\mathcal{C}'$, with vertex in O and axes respectively m and e_3, tangent in Ω. Then during any motion of the system the cone $\mathcal{C}$ stays fixed, while $\mathcal{C}'$ rolls uniformly without sliding around it, as in a gear. So, during the motion, two angles advance uniformly on circles, namely (see the figure) an angle $\tilde{g}$ around $\mathcal{C}$ from some fixed origin to the plane containing Ω and e_3, determining the orientation of e_3 in space, and one angle $\tilde{l}$ around $\mathcal{C}'$ from Ω to the axis e_1, determining the proper rotation of the body around e_3. All other quantities stay constant. *Any motion with m and e_3 non parallel runs on a torus $\mathbb{T}^2$ and is quasi periodic.* The frequencies of the motion depend on $\|m\|$ and m_3, but of course not on the irrelevant orientation of m in the space. If m and e_3 are parallel, then the cones degenerate into lines, and only one angle,

---

and $T^* SO(3)$ are trivial bundles.

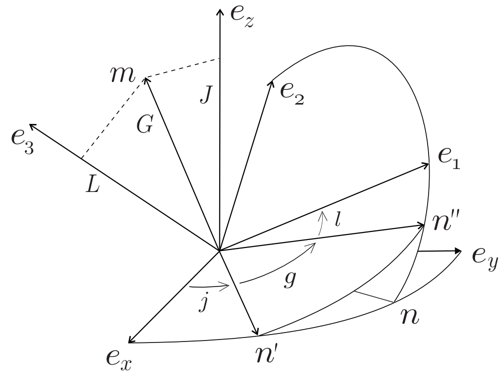

Figure 30: *The action–angle coordinates G, L, J, g, l, j for the symmetric rigid body.*

giving the proper rotation of the body, is well defined and advances uniformly; a family of tori $\mathbb{T}^2$ there degenerate into a circle. These are the proper rotations around e_3, also called gyroscopic rotations.

B. The action–angle coordinates. As a matter of fact, it is possible to introduce canonical action–angle coordinates adapted to the above description, for the case in which m and e_3 are not parallel. Precisely (see figure 30) take any fixed frame (e_x, e_y, e_z) with e_z not parallel to m; let $G = \|m\|$, $J = m_z = m \cdot e_z$, and denote by j the angle in the xy plane from the axis e_x to the nodal line $n' = e_z \times m$; quite clearly, G, J, j fix m in the space. Two coordinates then allow to fix e_3, and a good choice is $L = m_3 = m \cdot e_3$ and $g =$ angle from n' to $n'' = m \times e_3$ in the plane orthogonal to m. A last angle l from n'' to e_1, in the plane ortogonal to m_3, allows to determine completely the orientation of the body in the space. The angles g and l coincide with the previously introduced angles $\tilde{g}$ and $\tilde{l}$ on the Poinsot cones, up to the choice of the origin.

From the very construction it follows that G, L, J, g, l, j are local coordinates in $M = T^*SO(3)$, good whenever $e_z \times m \neq 0$ and $m \times e_3 \neq 0$, and that two charts of such coordinates, relative to different spatial frames (with different axis e_z) bypass the singularity $e_z \times m = 0$ and provide an atlas for

$$M_0 = \left\{(m, R) \in M : m \times e_3 \neq 0\right\} .$$

G, L, J, g, l, j are called Andoyer–Deprit coordinates [2] [37]. The relation between such coordinates and the familiar Euler coordinates is reported in detail in [14]. By working on that expressions, with some attention to handle the chart singularities $e_z \times m = 0$ and $e_z \times e_3 = 0$ of, respectively, the Andoyer–Deprit coordinates and the Euler coordinates, it is not difficult to prove that:

PROPOSITION 6.1 *The above Andoyer–Deprit coordinates are canonical, with* *canonical 2-form*

$$\mathrm{d}G \wedge \mathrm{d}g + \mathrm{d}L \wedge \mathrm{d}l + \mathrm{d}J \wedge \mathrm{d}j \ .$$

Two charts of Andoyer–Deprit coordinates, relative to two different spatial frames (e_x, e_y, e_z) *and* $(e_{x'}, e_{y'}, e_{z'})$ *with* $e_z \times e_{z'} \neq 0$*, provide an analytic symplectic atlas of* M_0*. In such cooordinates the kinetic energy of the rigid body depends only on* G *and* L*, and writes*

$$h(G,L) = \frac{1}{2A_1}(G^2 + \eta L^2) \ , \tag{6.3}$$

with

$$\eta = \frac{A_1 - A_3}{A_3} \ , \qquad -\frac{1}{2} \le \eta < \infty \ . \tag{6.4}$$

(Quite clearly, (6.3) is identical to (6.2).) In absence of external torques, h is the Hamiltonian. So, in agreement with the above Poinsot description, the actions G, L, J and also the angle j are costants of motion (m and m_3 stay constant), while the angles g, l advance uniformly with frequency

$$\omega = \frac{1}{A_1}(G, \eta L)$$

on a two dimensional torus, and for most initial data densely fill the torus.

It is important to observe that the coordinates G, L, l are intrinsic, i.e. are independent of the chart, while instead J, g, j do depend on it (on the choice of the spatial frame). A good notation for the set of coordinates is then

$$(G, L, J^{(i)}, g^{(i)}, l, j^{(i)}) \ , \qquad i = 1, 2 \ ,$$

the index i refering to the chart. But as is evident from the construction, $g^{(1)}$ and $g^{(2)}$ differ only for the origin, more precisely the transition function of this coordinate is of the form

$$g^{(2)} = g^{(1)} + g_{12}(J/G, j) \ . \tag{6.5}$$

Correspondingly, the tori on which $g^{(i)}$ and l run, the other coordinates staying fixed, are intrinsic; this is in fact obvious since, as remarked above, such tori are directly defined by the dynamics.

C. The double fibration of the phase space. It is easy now to understand the global structure of the phase space M [42][43]. Indeed (see figure 31) to identify a dynamical state of the system, we can first specify $G = \|m\|$ and $L = m \cdot e_3$, which are both intrinsic and belong to the "action space"

$$A = \big\{(G, L) \in \mathbb{R}^2 : |L| \le G\big\} \ .$$

Consider for a moment only the interior A_0 of A and the corresponding subset M_0 of M. For any given $(G, L) \in A_0$, the angular momentum m is still free

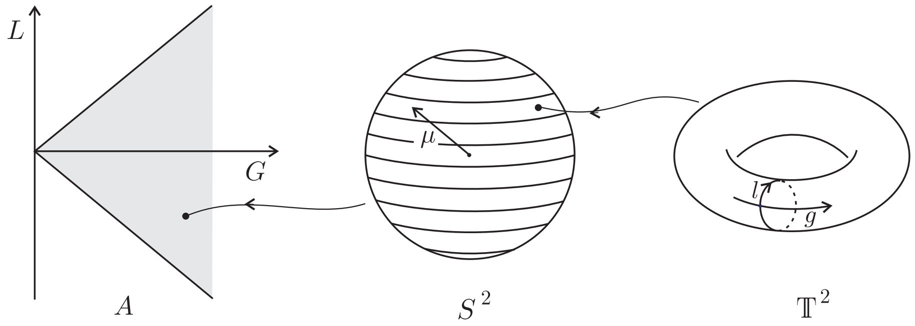

Figure 31: *Illustrating the double fibration of the phase space $T^*SO(3)$ of the Euler–Poinsot rigid body: the action space A, the sphere S^2 where $\mu = m/\|m\|$ runs, the torus $\mathbb{T}^2$.*

to move on a sphere, more precisely $\mu = m/G$ runs on the unit sphere S^2. (The sphere itself is intrinsic, and $(J^{(i)}/G, j^{(i)})$ provide a chart of cylindric–like coordinates on it.) Finally, for fixed G, L and μ, the angles $g^{(i)}$ and l stay on a torus $\mathbb{T}^2$.

The difference with respect to a nondegenerate integrable system is remarkable: in the nondegenerate case the phase space has the natural structure of a fiber bundle, with fiber $\mathbb{T}^n$ and, as basis, some manifold B of dimension n. Here the structure is more rich, since we have three manifolds, namely A_0, $\mathbb{T}^2$ and S^2, to play with. In fact, M_0 has a double structure of fiber bundle:

(a) It is a fiber bundle with two–dimensional fiber $\mathbb{T}^2$ and four dimensional basis $B = A_0 \times S^2$; this is indeed the finest fibration compatible with the dynamics (recall that trajectories generically are dense on tori).

(b) It is also a fiber bundle with two dimensional basis A_0, and four dimensional fiber $\mathfrak{F}$ obtained by grouping together all tori supporting motions with the same frequencies. Being characterized by the frequencies, $\mathfrak{F}$ too is a natural object entering the perturbation construction. (By the way, the fiber $\mathfrak{F}$ is in turn a fiber bundle with basis S^2 and fiber $\mathbb{T}^2$.)

The novelty is the presence of the "intermediate manifold" S^2, actually the manifold where the degenerate variables run, which can be attached either to A_0 or to $\mathbb{T}^2$ to produce the former or the latter fibration. What is instead meaningless, is trying to follow the Liouville–Arnol'd theorem and describing M_0 as a single fiber bundle with three–dimensional basis and fiber $\mathbb{T}^3$. This can be done by removing the poles from the sphere, thus getting a cylinder i.e. the product of an interval (with coordinate J) and a circle (with coordinate j), then attaching the interval to A_0 so as to form the three dimensional basis, and the circle to $\mathbb{T}^2$ so as to form $\mathbb{T}^3$. This is a totally unnatural procedure, depending in an essential way on the arbitrary choice of the external frame, and

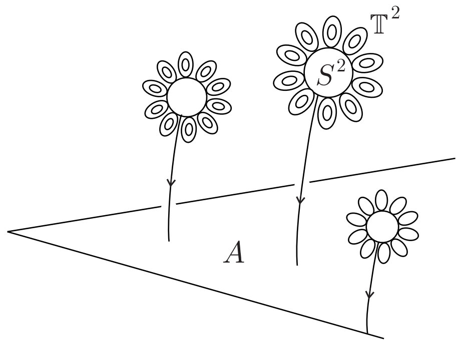

Figure 32: *A possible view $T^*SO(3)$ as a double fiber bundle.*

the resulting fibration, at variance with the two previous ones, contains spurious singularities. That is: *the Liouville–Arnol'd theorem in principle holds, but is misleading and does not provide a good starting point for the perturbation study.*

To complete the description of the phase space, we must consider the submanifold $M \setminus M_0$ of M, formed by the states with $|L| = G$. For $G > 0$ such states are the proper rotations: for them μ still runs on $\mathcal{S}^2$, but the tori $\mathbb{T}^2$ shrink to circles. So, proper rotations are singularities of both foliations, with lower dimensional fibers. We are of course very interested in proper rotations, but it is clear that such motions need a special treatment. We are not interest instead in the origin $L = G = 0$, anyhow it corresponds to the set of states with the body at rest, which together form $SO(3)$.

A pictorial view of the phase space is provided by figure 32: if we put a sphere in each point of A, and a torus in each point of the sphere, the result is more or less a field of daisies (spherical ones) [43]. The fiber $\mathcal{F}$ of fibration (b) is the flower. During the motion, the stem remains fixed, the petal also is constant, and the dynamics is quasi–periodic on it. The daisies on the border of the field have thinner petals S^1, and the motion on them is periodic.

Remark. The geometrical structure we have described, with three manifolds combining into two fibrations, is not at all peculiar of the rigid body, rather is typical of superintegrable systems. In the general case, the three manifolds are an action space of dimension $n_0 < n$, a torus of the same dimension n_0, and some intermediate manifold of dimension $2(n - n_0)$, characteristic of the system, representing the degenerate degrees of freedom. The resulting structure of double fibration is the outcome of a generalization of the Liouville–Arnol'd theorem to the superintegrable case [80] [73] [43]. The key–word for the subject is "noncommutative integrability".

6.3 The perturbed Euler–Poinsot system: non gyroscopic motions

A. Preliminaries. The Hamiltonian of the perturbed Euler–Poinsot system is, in any chart of action–angle coordinates,

$$H^{(i)}(G, L, J^{(i)}, g^{(i)}, l, j^{(i)}) = h(G, L) + \varepsilon f^{(i)}(G, L, J^{(i)}, g^{(i)}, l, j^{(i)}) \; ; \qquad (6.6)$$

two charts of action–angle coordinates, as we know, cover M_0. As is shown in [42] [15], it is possible to proceed with the perturbation construction in a way completely transparent to the change of chart, producing intrinsic canonical transformations and intrinsic normal forms. This is possible because of several facts: (i) The unperturbed Hamiltonian and the unperturbed frequency are intrinsic; consequently (after a choice of the cut–off N independent of the chart) the geography of the resonances is intrinsic. (ii) The Fourier decomposition of the perturbation is intrinsic: indeed, since l is intrinsic, and the transition map (6.5) for g is nearly trivial, each Fourier component is mapped by the change of chart into the same Fourier component; in particular the decomposition of the perturbation into the infrared and the ultraviolet parts, as well as the operation of extracting the average on the angles of any function, are intrinsic. (iii) Consequently, the solution χ of the basic equation of perturbation theory, which determines at each step the canonical transformation, is intrinsic. The conclusion is that both the canonical transformation and the trasformed Hamiltonian are intrinsic. From now on, to avoid a too heavy notation, we shall omit the chart superscript $i = 1, 2$ on all variables and functions, admitting implicitly that all expressions and statements involving coordinates, and all objects defined through coordinates, are local representatives, in the chart at hand, of intrinsic ones. In particular, in place of (6.6) we shall write

$$H(G, L, J, g, l, j) = h(G, L) + \varepsilon f(G, L, J, g, l, j) \; . \qquad (6.7)$$

Remark. A specially interesting case is the one in which f represents a purely positional potential (say a potential which, in any chart of Euler coordinates, depends only on the angles). In such a case, it is immediately checked that f is a homogeneous function of degree zero of the actions, and the study of the perturbed system, i.e. of (6.7) with small ε, is equivalent up to a trivial rescaling to the case of $\varepsilon = 1$ and large $G \sim 1/\sqrt{\varepsilon}$, i.e. to the case of a fast rotating body with angular velocity

$$\|\omega\| \sim 1/\sqrt{\varepsilon} \; .$$

Indeed, the homogeneous rescaling of actions and time

$$G = \sqrt{\varepsilon}\hat{G} \; , \qquad L = \sqrt{\varepsilon}\hat{L} \; , \qquad J = \sqrt{\varepsilon}\hat{L} \; , \qquad t = \hat{t}/\sqrt{\varepsilon}$$

conjugates H with the new Hamiltonian

$$\hat{H}(\hat{G}, \hat{L}, \hat{J}, g, l, j) = h(\hat{G}, \hat{L}) + f(\hat{G}, \hat{L}, \hat{J}, g, l, j)$$

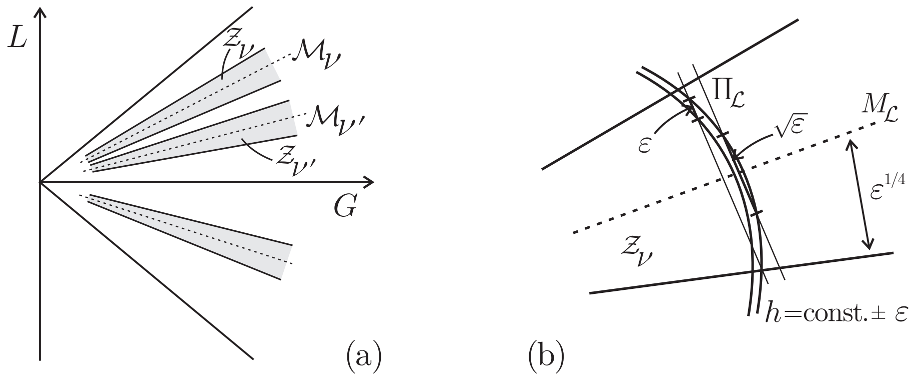

Figure 33: *(a) The geometry of resonanced for the Euler–Poinsot rigid body. (b) The special mechanism of confinement of the resonant motion.*

(thanks to the homogeneity, f is left unchanged), and any state with given G is mapped into a state with large $\hat{G}$ of order $1/\sqrt{\varepsilon}$. So, by studying (6.7) we are actually studying the fast rotations of the rigid body in a given ε–independent potential. The latter is indeed the language used in [15] [17].

B. The perturbation procedure. In spite of the fact that η in (6.3) can be negative, the function h is quasi–convex in A: inded, a lack of quasi–convexity of h occurs for $L = \pm G/\sqrt{-\eta}$, but for $\eta > -\frac{1}{2}$ (see (6.4)) such states are outside A; this is of course important to prove the confinement of the actions. Which kind of results are we interested in, and do we expect from the application of perturbation methods? It looks sensible to expect that the actions (G, L) and their conjugated angles (g, l) behave as is typical of Nekhoroshev theory, namely that for a long time, and up to a small noise, the actions stay constant ($\|m\|$ and $m \cdot e_3$ stay nearly constant), while the angles advance on the torus with almost constant frequency. What is instead far from obvious, and is the main question to be concerned with, is the behaviour of J and j, i.e. the motion of $\mu = m/G$ in space, which could be either regular or chaotic. (Proving that μ stays constant is out of question, as is easily seen by the counterexample of the Lagrange top in a small gravity $\mathcal{O}(\varepsilon)$: in that case μ advances with speed ε along the parallels of the sphere.)

The work proceeds as follows:

i) To keep away from the singularities of the foliation $m \times e_3 = 0$ ($L = \pm G$), we restrict the attention to the subset M_δ of the phase space M such that (G, L) stay inside

$$A_\delta = \left\{ (G, L) \in A : G - |L| > \varepsilon^{1/2}\,\delta \right\} .$$

In this way we exclude proper rotations and nearby motions, which need to be treated separately (see Subsection 6.5). We shall prove consistently that motions starting in, say, $M_{2\delta}$ will not escape M_δ. M_δ is covered by two charts of action–angle coordinates (for example relative to (e_x, e_y, e_z) and $(e_{x'}, e_{y'}, e_{z'})$ with $x' = y$, $y' = z$, $z' = x$).

ii) We choose a cut–off $N \sim \varepsilon^{-\frac{1}{4}}$ (the same N in the two charts), and proceed with the geography of resonances inside A_δ. Thanks to the low dimensionality and the simple form of h, the construction is very simple, namely resonant manifolds are lines $\mathcal{M}_\nu$ defined by

$$\nu \cdot \omega(G, L) = \nu_1 G + \nu_2 \eta L = 0 \ ,$$

while resonant regions and zones coincide and are corridors around resonant lines; the convenient definition of the zone $\mathcal{Z}_\nu$ is

$$\mathcal{Z}_\nu = \left\{ (G, L) \in A_\delta : |\nu \cdot \omega(G, L)| < \frac{G}{2N|\nu|} \right\} \ , \tag{6.8}$$

see figure 33. Different zones turn out to be well separated. The nonresonant zone $\mathcal{Z}_0$ is defined as the complement of the resonant zones.

iii) Inside each resonant or nonresonant zone we construct resonant and respectively nonresonant normal forms, with a number of perturbation steps $r \sim N \sim \varepsilon^{-1/4}$, leading to an exponentially small remainder of order $\exp -N$: namely

$$H_0'(G', L', J', g', l', j') = h(G', L') + \varepsilon u_0(G', L', J', j') + \mathcal{O}(\exp -N)$$

inside $\mathcal{Z}_0$ and

$$H_\nu'(G', L', J', g', l', j') = h(G', L') + \varepsilon u_\nu(G', L', J', \sigma, j') + \mathcal{O}(\exp -N) \ , \tag{6.9}$$

with $\sigma = \nu_1 g' + \nu_2 l'$, inside $\mathcal{Z}_\nu$; σ is indeed the slow combination of angles adapted to $\mathcal{Z}_\nu$. The canonical transformation is ε–near the identity. We stress once more that all of this is transparent to the change of chart, and all the functions we are constructing, in particular the normal forms, are local representatives, in the chart at hand, of functions on M_δ.

iv) For initial data inside $\mathcal{Z}_0$ the actions G', L', and thus G, L, stay obviously nearly constant for an exponentially long time. For initial data inside $\mathcal{Z}_\nu$, the motion $(G(t), L(t))$ is flattened on the line of fast drift Π_ν parallel to ν through (G^o, L^o), which is transversal to $\mathcal{M}_\nu$. Here the low dimensionality helps, with respect to the general case: as is sketched in figure 33b, a motion along Π_ν, being confined between two curves of constant h at distance ε, can never cross the whole resonant zone and advance of its diameter $\sim \varepsilon^{1/4}$: in the worst case, for initial data near the center of the resonance, the motion stays confined in a subregion of size $\varepsilon^{1/2}$, while near the border the confinement is even better, namely $\mathcal{O}(\varepsilon)$. This would be no longer true in higher dimension.

v) In the nonresonant case, thanks to the absence of the motion along the line of fast drift, $G'(t)$ and $L'(t)$ stay almost constant; together with the energy conservation, this implies that $u_0(G'(0), L'(0), J'(t), j'(t))$ also stays almost constant. The consequence is very important: *for nonresonant initial data, the point $(J'(t), j'(t))$ is confined to stay almost exactly on a level line of u_0*, and is not allowed to wander elsewere in the sphere. Such a mechanism of confinement, instead, does not work for resonant motions: both because $G'(t)$, $L'(t)$ are not as constant as in the nonresonant case, so that u_ν does not stay enough constant, and because the function u_ν entering the resonant normal form also depends on the "slow angle" σ, so even at fixed G and L, u_ν it is not even a function on the sphere.

Passing to the old variables, which are close to the new primed ones, and recalling that u_0 coincides, at the lowest order, with the average $\overline{f}(G, L, J, j)$ of the perturbing function on the angles g and l, the result of the above sketched work is the following proposition:

PROPOSITION 6.2 *Let H be as above; assume the perturbation f is analytic, the initial datum G^o is bounded away from zero, say $G^* < G^o < 2G^*$, and ε is smaller than a suitable constant ε^* (depending on the analyticity and norm of the perturbation, on G^* and on δ). Then:*

i) Any motion with initial datum $(G^o, L^o) \in M_{2\delta}$ does not escape M_δ up to

$$|t| = \mathrm{const}\, e^{(\varepsilon^*/\varepsilon)^{1/4}} \ .$$

ii) During this time, for all initial data it is

$$\begin{aligned} |G(t) - G^o|\,,\ |L(t) - L^o| &< \mathrm{const}(\varepsilon/\varepsilon^*)^b && (6.10)\\ |\dot g(t) - \dot g(0)|\,,\ |\dot l(t) - \dot l(0)| &< \mathrm{const}(\varepsilon/\varepsilon^*)^b\ , && (6.11)\end{aligned}$$

with $b = 3/4$ for nonresonant initial data, $b = 1/2$ for resonant initial data.

iii) With reference to the above introduced geography of resonances, if the initial datum G^o, L^o belongs to the nonresonant region $\mathcal{Z}_0$, then there exists a regular function $u(G, L, J, j)$ close to the average $\overline{f}$ of the perturbation on the angles g, l,

$$u(G, L, J, j) = \overline{f}(G, L, J, j) + \mathcal{O}(\varepsilon^{1/4})\ , \qquad (6.12)$$

such that during the motion it is

$$|u(G^o, L^o, J(t), j(t)) - u(G^o, L^o, J(0), j(0))| < \mathrm{const}\ \varepsilon/\varepsilon^*\ . \qquad (6.13)$$

For more detailes, see [15].

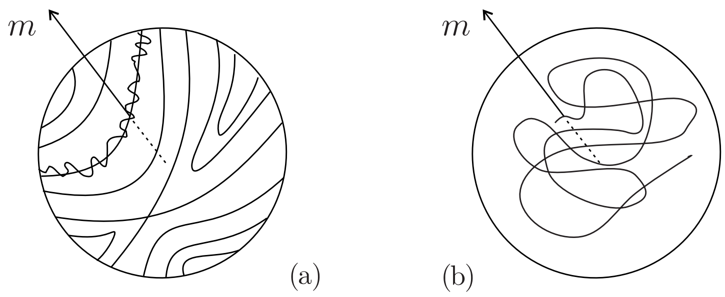

Figure 34: *(a) Nonresonant initial data: m follows, up to small errors, the lines of constant $\overline{f}$. (b) Resonant initial data: m moves unpredictably, possibly filling a wide portion of the sphere.*

So for any initial datum, according to (6.10), the system performs an approximate Euler–Poinsot motion around the instantaneous direction of the angular momentum m, with Poinsot cones of amplitude close to the initial one, rolling with frequencies close to the initial ones; in turn m moves slowly on the sphere, with speed $\mathcal{O}(\varepsilon)$. The motion, however, can be very different, depending on whether the initial datum (G^o, L^o) is resonant or not. In the nonresonant case, as an immediate consequence of (6.13), it turns out that the point $(J(t), j(t))$ is confined, with accuracy $\mathcal{O}(\varepsilon)$, to a level line of a smooth function on the sphere, which in turn, according to (6.12), is close to a level line of $\overline{f}$.[17] In the resonant case a similar mechanism does not exist, so m could wander unpredictably on the sphere, possibly filling densely a wide portion of it. The two possibilities are sketched in figure 34. *In principle, the irregular behavior could occur for no matter how small ε.* Since the speed of J, i is $\mathcal{O}(\varepsilon)$, the natural time unit for the irregular behavior, if any, is ε^{-1}; an appropriate name for it, taken from [39] [40], is "slow chaos". In the next subsection we shall further discuss the conditions for the existence of slow chaotic motions and provide numerical evidence of their effective presence in connection with elementary potentials.

6.4 Numerical results for non gyroscopic motions

The possibility of chaotic motions for resonant initial data is very concrete: indeed, let us consider the resonant normal form (6.9), and forget the small reminder. Proceeding as in section 4.3C, we can introduce in place of G', L', g', l' the fast and slow variables F, S, ψ, σ adapted to ν; treating F as a parameter, and passing from S to s as in (4.14), it is an easy exercise to obtain for the

[17] Such a behaviour generalizes, in a sense, the case of the Lagrange top in small gravity: indeed in that case m follows (for small ε) the parallels of the sphere, which are precisely the lines of constant gravitational potential (averaging here is unnecessary). For a generic perturbation the level lines are different, but the behavior is similar.

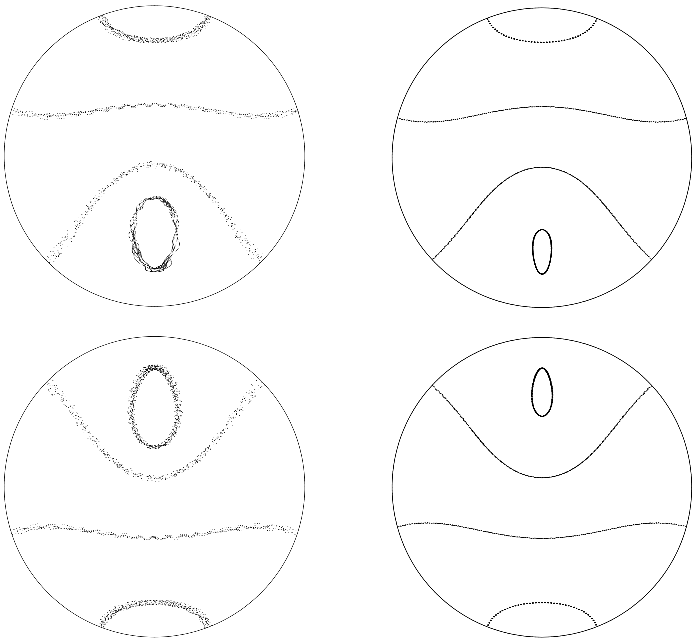

Figure 35: *Nonresonant motions for the potential* V_1, *for* $\varepsilon = 2 \times 10^{-2}$ *(left) and* $\varepsilon = 2 \times 10^{-3}$ *(right); integration times respectively* 10^6 *and* 10^7. *Polar stereographic projection, "north" emisphere up.*

resulting Hamiltonian the form

$$\mathcal{H}(s, \sigma, J, j) = \frac{s^2}{2A} + \varepsilon u(s, \sigma, J, j) \ , \tag{6.14}$$

with some $A > 0$ (F has been omitted). This is a system with two degrees of freedom, and thus it is reasonable to expect it has chaotic motions, too.

The question, however, is delicate, since the system (6.14) is not typical. An evident peculiarity is that J does not enter the unperturbed part (this reflects the degeneracy of the original system); moreover, J belongs to a domain of size independent of ε, while s (in the situation we are interested in, i.e. for motions around the center of the resonance) moves in a domain as small as $\sqrt{\varepsilon}$. *This difference in the domains makes useless the scaling arguments we invoked in section 4.3, to eliminate* ε *and show that chaotic motions are generically present.* More specifically, the absence of J in the unperturbed part has an immediate consequence, namely the degree of freedom (J, j) evolves on a time scale $\mathcal{O}(1/\varepsilon)$, while (s, σ), for s of order $\sqrt{\varepsilon}$, evolves on the faster time scale $\mathcal{O}(1/\sqrt{\varepsilon})$. For small ε the two time scales are well separated, and this might appear as an obstacle for the effective interaction of the two degrees of freedom. The situation however is different if $\mathcal{H}$, regarded at fixed (J, j), contains unstable

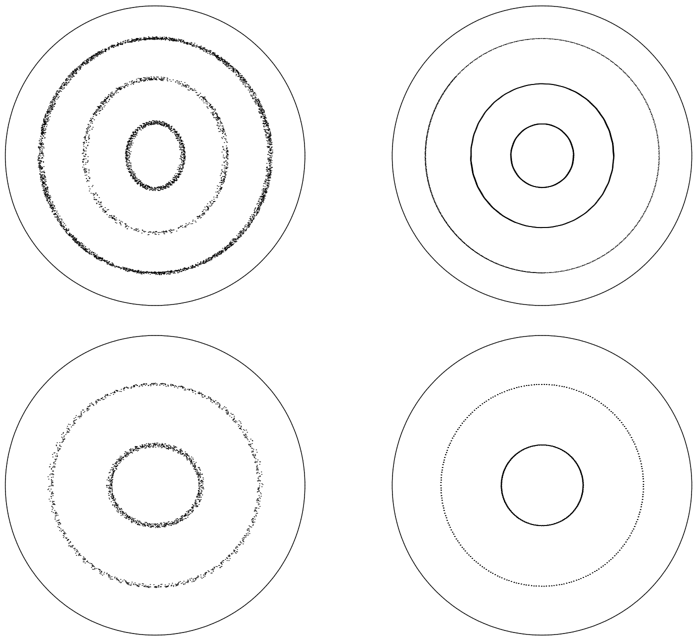

Figure 36: *Nonresonant motions for the potential V_2, for $\varepsilon = 2 \times 10^{-2}$ (left) and $\varepsilon = 2 \times 10^{-3}$ (right); integration times respectively 10^6 and 10^7.*

equilibrium points, with separatrices, for the pair (s, σ): for example

$$\mathcal{H} = \frac{s^2}{2} + \varepsilon(\tilde{u}(J, j) - J^2 \cos\sigma) \ ,$$

i.e. a pendulum with unstable equilibrium in $\pm\pi$ and separatrices passing through $(\pm 2J\sqrt{\varepsilon}, 0)$. Indeed around such states the variables (s, σ) are slowered, and in these conditions the interaction with (J, j) can get effective. *For each fixed (J, j) the interested portion of the phase space is thin, but it becomes larger if, due to the evolution of (J, j), the separatrices move and span a wide zone of the phase plane of s, σ.* As a matter of fact, this mechanism has been theoretically studied in [39] [40] [76] [77] [78], and it was even proven that under suitable assumptions, and in a sense to be made precise, slow chaotic motions do exist. The behaviour however, see in particular [78], is rather complex and not easily described.

As a final theoretical remark before passing to numerical results, let us say that the search for chaotic motions must be restricted to low resonances (small $|\nu|$): indeed it was proven in [60] that if $|\nu| > (\mathrm{const}) \log \varepsilon^{-1}$, then chaotic motions cannot be present. The question however is a little technical and we shall not enter it.

We now pass to the numerical part, following [13] [14]. Numerical integration

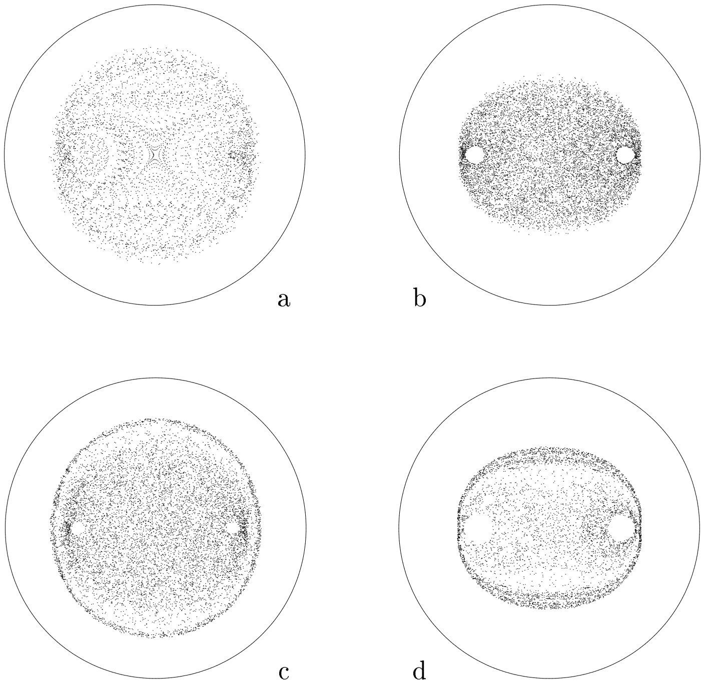

Figure 37: *Motions near the resonance* $(1, -1)$, *for the potential* V_2, *with* $\eta = 2$, *initial values* $(G, L) = (1, 0.5)$, $\varepsilon = 10^{-3}$ *(a),* 10^{-4} *(b),* 10^{-5} *(c),* 10^{-6} *(d). Only the upper emisphere is represented. Integration times, in the order:* 5×10^7, 5×10^8, 5×10^9, 2.9×10^{10}.

of the rigid body is not a completely trivial matter; the reason is that the phase space $\mathbb{R}^3 \times SO(3)$ of the system is not an Euclidean space but a manifold, which cannot be covered by a single chart of local coordinates. Since the motions are not localized, and in particular the angular momentum m is expected to possibly wander throughout S^2, it is hopeless to work within a single set of local coordinates. Several ways out have been proposed in the literature. The algorithm we found convenient for our purposes is a second order "splitting algorithm"

$$\Psi_\tau = \Phi_{\varepsilon f}^{\tau/2} \circ \Phi_h^\tau \circ \Phi_{\varepsilon f}^{\tau/2} \ , \tag{6.15}$$

where Φ_h^t and $\Phi_{\varepsilon f}^t$ denote the Hamiltonian flow of, respectively, the kinetic part h and the potential part εf of the Hamiltonian, while τ is the time–step. It should be stressed that: (i) Ψ_τ is a well defined map on the manifold, independent of the coordinates in use, which in principle can be implemented in any set of local coordinates (it is also possible to use as variables $(m, R) \in \mathbb{R}^3 \times SO(3)$, as was done in [91] to avoid singularities). (ii) The algorithm is symplectic, with the well known advantages (see point D of the Conclusions).

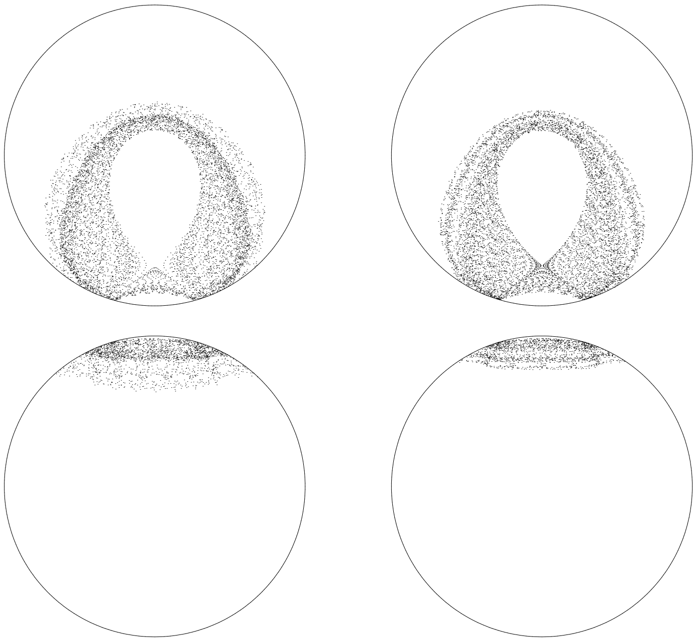

Figure 38: *One chaotic orbit for the potential* V_3*, inside the resonance* $(1, -2)$*, with* $\eta = 2$*,* $(G, L) = (1, 0.25)$*;* $\varepsilon = 10^{-3}$ *(left) and* 10^{-4} *(right); integration time respectively* 10^7 *and* 10^8*.*

In [13] [14] the intrinsic character of the algorithm (6.15) was exploited by implementing it in two charts of Euler coordinates, chosen in such a way to cover the whole phase space $\mathbb{R}^3 \times SO(3)$; whenever approaching a chart singularity, one switches from one chart to the other. The performance of this method was quite satisfactory; for details, see [13]. Conceptually, however, the way in which the above algorithm is implemented is irrelevant, so we shall not further comment this point. We considered three perturbations, chosen in such a way that they are sufficiently simple both in Euler coordinates, for the numerical integration, and in action–angle coordinates, to keep a good theoretical control of the results and to be able to produce at hand, whenever necessary, the (lowest order) normal form. After a few attempts, we restricted the attention on three potentials, which in (a chart of) Euler coordinates have the form

$$\begin{aligned} V_1(\vartheta, \varphi, \psi) &= \sin\vartheta \cos\varphi + \sin^2\vartheta \cos^2\psi \\ V_2(\vartheta, \varphi, \psi) &= \sin^2\vartheta \sin\varphi \sin\psi + \cos\vartheta \\ V_3(\vartheta, \varphi, \psi) &= V_1(\vartheta, \varphi, \psi) + V_2(\vartheta, \varphi, \psi) \ . \end{aligned}$$

We report here only a sample of numerical results; for a more complete collection, see [14].

Figure 35 refers to the potential V_1 and shows a few nonresonant motions, at

$\varepsilon = 10^{-2}$ (left) and 10^{-3} (right). The figure represents the motion of the tip of the unit vector $\mu = m/\|m\|$ on S^2; points are drawn at equal time intervals. The sphere is represented by means of the common polar stereographic projection ("north" emisphere up). Figure 36, very similar, refers to potential V_2. In both cases, by lowering ε, the region spanned by μ shrinks, as expected, to a line, and it would be easy to check that, in agreement with point (iii) of the above proposition 6.2, such a line is close to a line of constant averaged potential. The situation is instead substantially different, if we take initial data inside a low–order resonant region. For all of the above potentials, we found resonant chaotic motions in which μ fills wide regions on S^2, with no indication that such regions shrink to a line by lowering ε. Figure 37 shows the case of potential V_2, inside the resonance $\nu = (1, -1)$, for ε down to 10^{-6} (only the "north" emishpere, where the motion turns out to be confined, is here represented). The figures at different ε are not exactly similar to each other, but they do not need to be since, as stressed above, scaling breaks for degenerate systems. Another example of chaotic motion, relative to potential V_3, is reported in figure 38.

6.5 The perturbed Euler–Poinsot system: gyroscopic motions

A. Extending Proposition 6.2. As we already know (section 6.2), on proper rotations around e_3 the action–angle coordinates become singular. The singularity involves both the (G, g) and the (L, l) pair of conjugated variables and comes from the fact that, for m parallel to e_3, the nodal line n'' (figure 30) is no more defined. But as one easily guesses from the figure, the angle $\gamma = g + l$ remains well defined on the singularity, too, there coinciding with the Euler angle ψ (the nodal line $n = e_z \times e_3$ and $n' = e_z \times m$ do coincide on the singularity). It is then convenient to introduce in place of (G, L, g, l) the adapted action–angle coordinates $(\Gamma, \Lambda, \gamma, \lambda)$ defined, for each choice of the spatial frame, via the linear canonical transformation

$$\Gamma = G \ , \qquad \Lambda = G - L \ , \qquad \gamma = g + l \ , \qquad \lambda = -l \ ,$$

(J, j) remaining unchanged. As is illustrated by figure 39, the singularity is now confined to the (Λ, λ) pair, and is an ordinary polar–like singularity.

Remark. Quite clearly, Γ, Λ and λ are intrinsic, i.e. are independent of the spatial frame at hand, while γ and of course J, j do depend on it. Nevertheless, we shall not use the heavy notation $\gamma^{(i)}$ and so on: exactly as for non–gyroscopic motions, the perturbation development turns out to be transparent to the change of chart, and all of the functions we shall use, although written in coordinates, turn out to be representatives, in the chart at hand, of intrinsically defined ones.

The further change of coordinates to Cartesian ones:

$$p = \sqrt{2\Lambda} \cos \lambda \ , \qquad q = \sqrt{2\Lambda} \sin \lambda \ ,$$

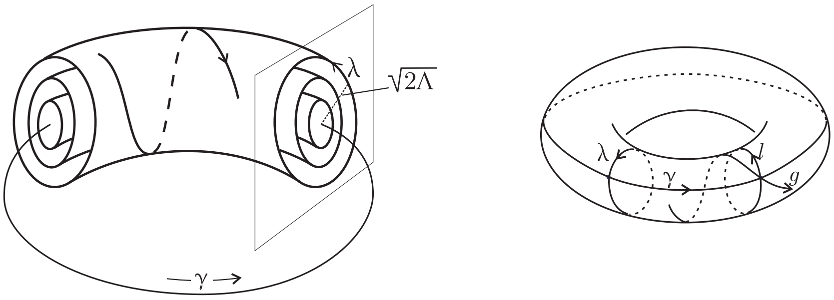

Figure 39: *Illustrating the change of variables* $(G,L,g,l) \mapsto (\Gamma,\Lambda,\gamma,\lambda)$.

removes the singularity, and it turns out that

$$p = \frac{m_2}{\sqrt{\Gamma - \frac{1}{2}\Lambda}} \ , \qquad q = -\frac{m_1}{\sqrt{\Gamma - \frac{1}{2}\Lambda}} \ . \tag{6.16}$$

More precisely,

PROPOSITION 6.3 *For any choice of the spatial frame* (e_x, e_y, e_z) *such that* $e_z \times m \neq 0$, *the coordinates* $(\Gamma, \gamma, p, q, J, j)$, *with* p, q *as in (6.16),* $\Gamma = G$ *and*

$$\gamma = g + l \quad \text{for} \quad L < G \ , \qquad \gamma = \psi \quad \text{for} \quad L = G \ ,$$

are analytic canonical coordinates in a neighborhood of the proper rotation around e_3. *Two charts of such coordinates, relative to two different spatial frames (with non parallel* z*–axes), provide an atlas of* $M^+ = \big\{(m,R) \in \mathbb{R}^3 \times SO(3) : m_3 > 0\big\}$.

In the following we shall still make use the symbol Λ, with however the agreement that it does not represent a coordinate, but a function of p and q, namely

$$\Lambda(p,q) = \frac{1}{2}(p^2 + q^2) \ .$$

As is immediately checked, the Hamiltonian has now the form

$$K(\Gamma,\gamma,p,q,J,j) = k(\Gamma,\Lambda(p,q)) + \varepsilon\Phi(\Gamma,\gamma,p,q,J,j) \ , \tag{6.17}$$

with

$$k(\Gamma,\Lambda) = \frac{1}{A_3}(\Gamma^2 + \beta\Lambda^2 - 2\beta\Gamma\Lambda) \ , \qquad \beta = \frac{\eta}{1+\eta} = 1 - \frac{A_3}{A_1} \ , \qquad -1 \le \beta < 1$$

and suitable Φ; the unperturbed frequency is then

$$\Omega(\Gamma,\Lambda) = \frac{\Gamma}{A_3}(1,-\beta) + \frac{\beta\Lambda}{A_3}(-1,1) \ .$$

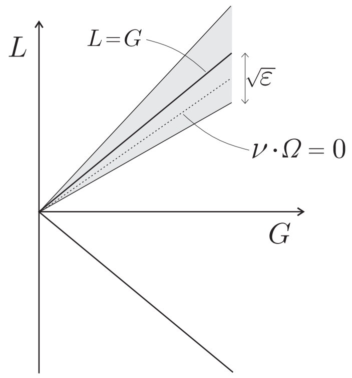

Figure 40: *At most one resonant line $\nu \cdot \Omega = 0$ can stay in a $\sqrt{\varepsilon}$–neighborhood of the proper rotation $\Lambda = 0$, either side of it.*

On the proper rotation, i.e. for $\Lambda = 0$, the resonance properties of the motion depend only on β, that is on the body. As for non gyroscopic motions, we shall use a cut–off $N \sim \varepsilon^{-1/4}$. Then in a neighborhood of order $\varepsilon^{1/2}$ of the gyroscopic rotation (the set left aside in proposition 6.2, where we are now obliged to work) there can be at most one resonant line, of equation

$$(\nu_1 - \beta\nu_2)\Gamma - \beta(\nu_1 - \nu_2)\Lambda = 0 \ ,$$

and this requires

$$\beta = \frac{\nu_1}{\nu_2} + \mathcal{O}(\varepsilon^{1/2}) \ ; \tag{6.18}$$

see figure 40 (warning: the resonant line can stay either side of the border $\Lambda = 0$ of the action space). So, in a neighborhood of size $\varepsilon^{1/2}$ of the proper rotation, it is not necessary to further distinguish between resonant and nonresonant motions, rather (for each choice of the cut–off and thus of ε) we should distinguish between resonant and non resonant bodies, and construct either the resonant or the nonresonant normal forms, depending only on ε and β.

In fact, the whole perturbation procedure runs as smoothly as in the non gyroscopic case, up to technical questions related to the use of the Cartesian coordinates. The only point which deserves a remark are Fourier series: Fourier series are indeed a very basic instrument, so they need to be rewritten in Cartesian coordinates. Following the formal development of [90], we first replace the real p, q variables by the so–called Birkhoff complex variables:

$$w = \frac{p + iq}{\sqrt{2}} \ , \qquad z = \frac{p - iq}{i\sqrt{2}} \ ;$$

clearly $\Lambda = izw$, $z = \sqrt{\Lambda}\, e^{i\lambda}$ and $w = -i\sqrt{\Lambda}\, e^{-i\lambda}$. The Fourier series of any function F of $\Gamma, \gamma, w, z, J, j$ writes then

$$F(\Gamma, \gamma, w, z, J, j) = \sum_{\nu \in \mathbb{Z}^2} F_\nu(\Gamma, \Lambda, J, j)\, e^{i\nu_1\gamma} E_{\nu_2}(w, z) \ ,$$

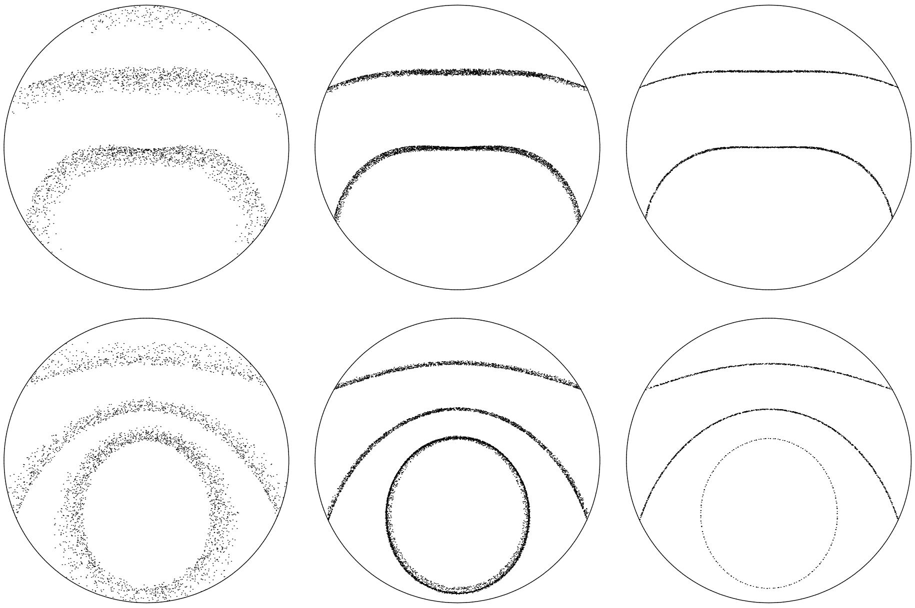

Figure 41: *Resonant gyroscopic motions for the potential* V_3*, in the resonance* $(-1,1)$*;* $\beta=-1$*,* $(\Gamma,\Lambda)=(1,0)$ *and (left to right)* $\varepsilon=10^{-2}$*,* 10^{-3}*,* 10^{-4}*.*

with

$$E_{\nu_2}(w,z)=\begin{cases} z^{\nu_2} & \text{for } \nu_2\geq 0 \\ w^{-\nu_2} & \text{for } \nu_2<0\ . \end{cases} \tag{6.19}$$

As for the case of non gyroscopic motions, such Fourier series turn out to be intrinsic, i.e. independent of the chart (of the spatial frame) at hand.

The result is the following proposition [17], which extends to gyroscopic motions the content of proposition 6.2:

PROPOSITION 6.4 *Let* K *be as in (6.17); assume the perturbation* Φ *is analytic, the initial datum* Γ^o *is bounded away from zero and* $\Lambda^o<\delta\varepsilon^{1/2}$*. If* ε *is sufficiently small, then for* t *up to*

$$|t|=\mathrm{const}\, e^{(\varepsilon^*/\varepsilon)^{1/4}} \tag{6.20}$$

the actions Γ *and* Λ *satisfy the estimate*

$$|\Gamma(t)-\Gamma^o|<\mathrm{const}(\varepsilon/\varepsilon^*)^{b_1}\ , \qquad |\Lambda(t)-\Lambda^o|<\mathrm{const}(\varepsilon/\varepsilon^*)^{b_2}\ ,$$

with $b_1=3/4$*,* $b_2=1$ *in the nonresonant case,* $b_1=b_2=1/2$ *in the resonant case. Moreover, in the nonresonant case the Hamiltonian (6.17) is canonically conjugated to the normal form Hamiltonian*

$$\begin{aligned} K'(\Gamma',\gamma',w',z',J',j') &= k(\Gamma',\Lambda(w',z')) \\ &+ \varepsilon u(\Gamma',\Lambda(w',z'),J',j')+\mathcal{O}(\exp-(\varepsilon^*/\varepsilon)^{1/4})\ , \end{aligned}$$

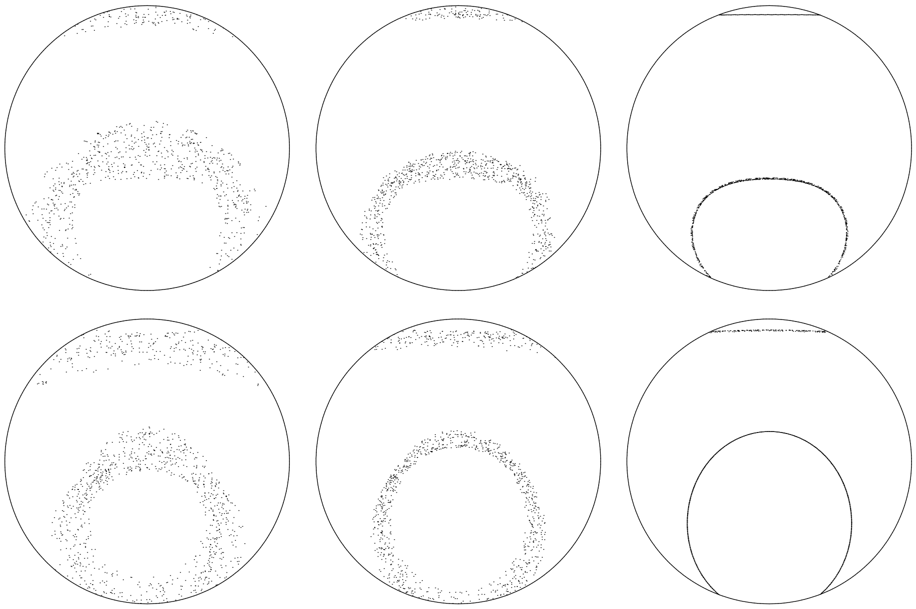

Figure 42: *Resonant motions for the potential V_3, in the resonance $\nu = (1, 2)$, at fixed $\varepsilon = 2 \times 10^{-3}$. Left to right: $\beta = 0.60,\ 0.55,\ 0.50$; initial data on the line $\nu \cdot \Omega = 0$, corresponding respectively to $\Lambda = 1/3,\ 1/6,\ 0$.*

with

$$u(\Gamma, \Lambda, J, j) = \Phi_{0,0}(\Gamma, \Lambda, J, j) + \mathcal{O}(\varepsilon^{1/4}) \ ,$$

and the canonical transformation is ε–close to the identity.

The immediate consequence is that, in the nonresonant case, J', j' follow, within accuracy ε, the level lines of a smooth function on the sphere $\|m\| = \Gamma'(0)$. The unprimed physical variables J, j do the same up to the deformation produced by the canonical transformation, that is still up to ε. In turn, this line is $\varepsilon^{1/4}$–close to a level line of the averaged potential energy $\Phi_{0,0}$. Such a mechanism of confinement does not work, instead, in the resonant case: precisely as for the non gyroscopic motions, the above proposition 6.4 allows for the presence of chaotic motions.

However, accurate numerical results [14] show that *proposition 6.4 does not provide an optimal description of the resonant case.* This is evident, for example, in figure 41, which refers to[18] $\beta = -1$, $\nu = (-1, 1)$, and ε ranging from 10^{-2} to 10^{-4}. The difference with respect to figures 37 and 38 is striking: wide chaotic motions here do not appear. Another evidence comes from figure 41:

[18] There are theoretical reasons, see [14], to believe that this is the most promising resonance for the search of chaotic motions.

here $\varepsilon = 2 \times 10^{-3}$ is fixed, and the initial data are taken inside the resonance $\nu = (1,2)$; the corresponding resonant line stays on the gyroscopic rotation for $\beta = \frac{1}{2}$, and gets far from it for significantly different ε. The figure shows clearly that, by approaching the value $\beta = \frac{1}{2}$, the width of the strip where J, j move shrinks to a line. So, *gyroscopic resonant motions are significantly more regular than non gyroscopic ones,* and proposition 6.4 needs to be improved.

B. Stability of all proper rotations. As remarked in [14], a theoretical element was not fully exploited in the proof [17] of proposition 6.4. Indeed, the resonant normal form associated to a resonance $\nu \cdot \Omega = 0$ is (quite obviously) of the form

$$\begin{aligned} K'(\Gamma', \gamma', w', z', J', j') &= k(\Gamma', \Lambda(w', z')) \\ &+ \varepsilon u(\Gamma', \gamma', w', z', J', j') + \mathcal{O}(e^{-(\varepsilon^*/\varepsilon)^{1/4}}) \end{aligned}$$

with

$$\begin{aligned} u(\Gamma', \gamma', w', z', J', j') &= u_0(\Gamma, \Lambda(w, z), J, j) \\ &+ \sum_{0 \neq n \in \mathbb{Z}} u_{n\nu}(\Gamma, \Lambda(w, z), J, j)\, e^{in\nu_1\gamma} E_{n\nu_2}(w, z) \ . \end{aligned}$$

But $E_{n\nu_2}(w, z)$, see (6.19), is proportional to either $w^{|n\nu_2|}$ or $z^{|n\nu_2|}$, $|n| \geq 1$. On the other hand, from (6.18) and $\|\beta\| \leq 1$ it follows $|\nu_2| \geq 1$: so, *all terms entering the normal form, but the average u_0, have in front a factor w or z, which is small in the vicinity of the proper rotation.*

How to use this small factor is not evident, since the Fourier term itself is small, but its derivatives with respect to either w or z, which enter the equations of motion, are not. The way out was found recently [19]. We provide here only a sketch of the result, focusing the attention to the case of the resonance $\nu = (-1, 1)$, $\beta = -1$ (actually the worst case). First of all, it is possible to simplify the resonant normal form by observing that, for $\nu = (-1, 1)$, only Γ, Λ, J, j and the combinations of variables

$$\tilde{z} = z e^{-i\gamma} \ , \qquad \tilde{w} = w e^{i\gamma} \tag{6.21}$$

enter the function u. The variables $\tilde{w}, \tilde{z}$ are then conveniently used in place of w, z; to have a canonical transformation, it is necessary to complete (6.21) by

$$\tilde{\Gamma} = \Gamma + \Lambda(w, z) \ , \qquad \tilde{\gamma} = \gamma$$

while the pair (J, j) is left unchanged.

Remark. A little reflection shows that $\tilde{\Gamma}, \tilde{\gamma}$ are the "fast" variables associated to the resonance $(-1, 1)$, while the "corotating" variables $\tilde{w}, \tilde{z}$ are the Cartesian equivalent of the "slow" variables (see section 4.3). Once more, the slow and fast variables appear to be the key to understand the resonant motions.

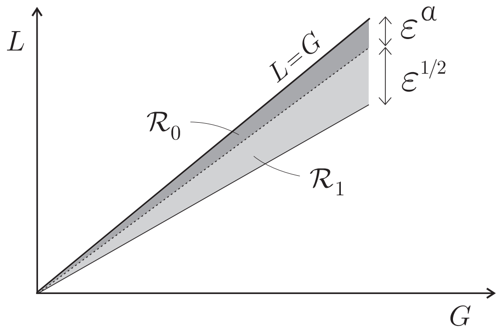

Figure 43: *Dark grey: the region $\mathcal{R}_0$ of size ε^{α}, $\frac{1}{2} < \alpha < \frac{2}{3}$, where chaotic motions are possible. Light grey: the region $\mathcal{R}_1$, of size $\varepsilon^{1/2}$, where it is possible to construct a nonresonant normal form.*

The Hamiltonian takes now the form

$$\begin{aligned}\tilde{K} \quad = \quad & \frac{1}{2A_3}(\tilde{\Gamma}^2 - \tilde{\Lambda}^2) + \varepsilon[\mathcal{U}(\tilde{\Gamma}, \tilde{\Lambda}, J, j) + \tilde{w}\mathcal{V}_w(\tilde{\Gamma}, \tilde{w}, \tilde{z}, J, j) \\ & + \tilde{z}\mathcal{V}_z(\tilde{\Gamma}, \tilde{w}, \tilde{z}, J, j)] + \mathcal{O}(e^{-(\varepsilon^*/\varepsilon)^{1/4}}) \ .\end{aligned}$$

The work then proceeds as follows. The neighborhood of the proper rotation we are working in, is decomposed into two parts (see figure 43): a thin inner sector $\mathcal{R}_0$ of size ε^{α}, with

$$\frac{1}{2} < \alpha < \frac{2}{3} ,$$

and a larger outer sector $\mathcal{R}_1$ of size $\varepsilon^{1/2}$. Inside $\mathcal{R}_1$ it turns out to be possible to use the smallness of $\tilde{w}, \tilde{z}$ to start a further perturbation procedure, which eliminates $\mathcal{V}_w$ and $\mathcal{V}_z$ up to a remainder of order

$$e^{-(\varepsilon^*/\varepsilon)^b} , \qquad b = 1 - \frac{3}{2}\alpha , \qquad 0 < b < \frac{1}{4} ,$$

and leads to a nicer nonresonant normal form:

$$K' = \frac{1}{2A_3}(\Gamma'^2 - \Lambda'^2) + \varepsilon u'(\Gamma', \Lambda', J', j') + \mathcal{O}(e^{-(\varepsilon^*/\varepsilon)^b}) \ .$$

In this way, inside $\mathcal{R}_1$ the regularity of motions is recovered, though for reduced times. Inside $\mathcal{R}_0$ instead the procedure does not apply and chaotic motions cannot be excluded. But *thanks to the smallness of $\mathcal{R}_0$, the coordinates J, j are no longer allowed to wander in a wide portion of the sphere.* This is evident from the form of the above Hamiltonian $\tilde{K}$, by simply using the conservation of energy: indeed Γ' stays practically constant, the remainder is irrelevant, while Λ and w, z are bounded by assumption by, respectively, ε^{α} and $\varepsilon^{\alpha/2}$. As a result, $u(\tilde{\Gamma}, 0, J, j)$ might oscillate, at most, by ε^{β}, with

$$\beta = \min(2\alpha - 1, \alpha/2) = 2\alpha - 1 , \qquad 0 < \beta < 1/3 \ .$$

That is: (J, j) *are confined to a level line of a regular function of the sphere, up to a positive power of* ε*; wide chaotic motions in gyroscopic conditions are excluded.*

In conclusion, the situation can be summarized as follows:

- For ordinary (i.e. non gyroscopic) motions, degeneracy is relevant and leads, in correspondence with resonant initial data, to wide chaotic motions on a time scale of order $1/\varepsilon$. Indeed inside resonances the actions G, L move at most by quantities of order $\sqrt{\varepsilon}$, as in nondegenerate systems, but thanks to degeneracy, J, j are allowed to move in a larger region, of size $\mathcal{O}(1)$.
- For gyroscopic motions, chaos is also possible in connection with resonances. But in spite of degeneracy, thanks to the extra smallness of the perturbation due to the small distance to the proper rotation, such motions have smaller size: in the worst case, the action Λ wanders irregularly on a region of size ε^{α}, with α as close as one wishes to $\frac{2}{3}$, while on the same time J, j follow a level line of a regular function on the sphere, up to oscillations of size $\varepsilon^{2\alpha-1}$. This is, qualitatively, the behavior of nondegenerate systems: chaotic motions do exist, but are spatially localized. In conclusion: *by approaching the proper rotations the special behavior due to degeneracy does disappear.*

This is the first proof of long time stability for all gyroscopic motions of the Euler perturbed rigid body. For details, including numerical tests confirming the theory, see [19].

7 Conclusions

We conclude these lectures by spending a few words on some selected applications of Hamiltonian perturbation theory, which we could not discuss in the previous sections.

A. Equilibrium times in gases. Consider a gas of interacting polyatomic molecules, for simplicity diatomic ones. The "Principle of equipartition of energy" — actually the cornerstone of (equilibrium) Statistical Mechanics — states that whenever energy is, for example, added to the system, eventually it goes equally distributed among all molecules and, for each molecule, among its degrees of freedom, proportionally to the number of (quadratic) terms entering the Hamiltonian of the molecule. The immediate consequence is that, if f denotes the number of terms in the Hamiltonian of a molecule, the constant volume specific heat of a mole of gas is

$$C_V = \frac{f}{2} R \ ,$$

R being the usual constant of gases. For diatomic molecules there are three translational and two rotational degrees of freedom, each contributing to the

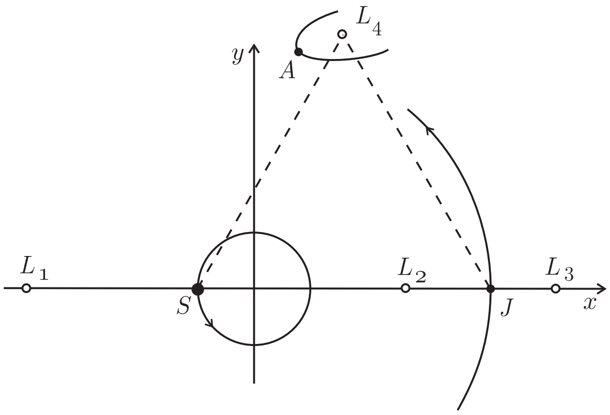

Figure 44: *The location of the Lagrangian equilibria $L_1, \ldots, L_4$, for the circular restricted three body problem, in the corotating frame.*

Hamiltonian with a kinetic term, and a vibrational degree of freedom, which instead contributes by both a kinetic and a potential term; so $f = 7$. Experience shows instead that, in ordinary conditions, $C_V = \frac{5}{2}R$, as for rigid nonvibrating molecules, while only at much larger temperatures the value $\frac{7}{2}R$ is recovered. The commonly accepted interpretation is based on quantum mechanics: if the frequency ω of the oscillator is sufficiently large ($\hbar\omega > k_B T$, where $\hbar$ and k_B are respectively Plank's and Boltzmann's constants and T is the absolute temperature), then there is not enough energy for a "quantum jump"; correspondingly the vibrational energy is, with great probability, constant or "frozen". Quantum freezing of large frequency oscillators is usually said to have no analog in classical physics.

As is not much known, a purely classical mechanism for the freezing of high frequency oscillators was proposed, before quantum mechanics, by Boltzmann and Jeans. The idea is that, for large ω, the exchange of energy between the vibrational degree of freedom and the remaining ones is extremely inefficient and slow, so the oscillators, in any experiment performed in a reasonable time, *appear* as if they where frozen. Such a possibility has been recently analyzed in a sequence of papers, both analytic and numeric, in the framework of Hamiltonian perturbation theory. The overall result, in the very essence, is that classical freezing does exist, more precisely the times necessary to practically share energy with the high frequency oscillators grow exponentially with the frequency (this is indeed the classical analog of the exponentially small probability of excitation of quantum mechanics). Among the papers on the subject, we suggest [10][28][12][32] and the review paper [11].

B. The stability of the Lagrangian equilibria L_4, L_5. Consider the circular restricted three body problem; to fix the ideas, let us call Sun and Jupiter, with masses respectively m_S and $m_J < m_S$, the two primaries, and Asteroid the zero mass body. It is well known that in the corotating frame in which the Sun and Jupiter are at rest, see figure 44, there are five equilibrium points for the Asteroid, called Lagrangian equilibria. Three of them, usually denoted L_1, L_2, L_3, are collinear with the primaries and are unstable, more precisely are hyperbolic. The remaining two, denoted L_4, L_5, lie symmetrically on opposite sides of the line Sun–Jupiter and form with the two primaries equilateral triangles; they are easily seen to be hyperbolic if the mass ratio $\mu = m_J/(m_S + m_J)$ exceeds the so–called Routh limit[19] $\mu_R = 0.038\ldots$, and are instead elliptic for $\mu < \mu_R$. The fact that L_4, L_5 are elliptic, however, does not imply their stability: indeed one of the three frequencies of the linearized problem is negative, so that the energy of the Asteroid is not positive definite in any neighborhood of the equilibria.

The question of the stability of L_4, L_5 is nontrivial.[20] For the planar case (initial position and velocity of the Asteroid in the corotating plane) the problem was positively solved during the sixties and the early seventies of last century, within KAM theory: for any $\mu < \mu_R$, in any arbitrarily small neighborhood of the equilibrium point and on each surface of constant energy, there are KAM tori which bound the motion and thus provide perpetual stability. The result however does not hold in the spatial case: KAM tori do exist, but due to the higher dimensionality of the energy surface, they do not provide topological obstruction and diffusion is possible.

Several works have been devoted to the spatial case, to prove that the equilibrium, though possibly not perpetually stable, is nevertheless stable for long times, namely for times growing exponentially with the distance ε of the initial datum from the equilibrium point. The idea of long time stability of L_4, L_5, in fact, is rather old [69], but only in the last fifteen years, after the Nekhoroshev theorem, it was the object of a systematic study. Some papers, like [53] and [52], concern general elliptic equilibria, and prove exponential stability under the strong[21] assumption that the frequencies of the linearized problem satisfy the Diophantine condition (3.8). A related point of view, in a sense more pragmatic, is considering, for example, the realistic Sun–Jupiter problem, with realistic masses and thus frequencies, and determining a neighbourhood of L_4, L_5 such that the escape time from it, computed by constructing a normal form of sufficiently high order and estimating the remainder, exceeds the Universe lifetime.[22] Such a program, started in [53], led to significant results: in particular, the region of "practical stability" around L_4, L_5, as produced in [58], is sufficiently large to contain some of the asteroids which are effectively known to

[19] Let us recall that for the realistic Sun–Jupiter system $\mu \simeq 0.00095$, while for the Earth-Moon system $\mu \simeq 0.011$.

[20] This is perhaps the oldest "elementary" unsolved problem of Celestial Mechanics, likely of Mechanics.

[21] In the framework of the three body problem such an assumption should be considered strong, because the three frequencies in the game are not independent parameters, but functions of the mass ratio μ: the point is that by varying μ, the frequency vector describes a particular curve in $\mathbb{R}^3$, and while it is well known that Diophantine frequencies are abundant in measure in any open subset of $\mathbb{R}^3$, essentially nothing can be said for that curve.

[22] The minimal incertitude in the frequency (which could be even rationally dependent) is irrelevant in this approach, as it does not affect the constructed finite order normal form.

oscillate around such points.

A further possibility is extending Nekhoroshev theory, so as to possibly prove exponential stability of the triangular Lagrangian points for all values of μ up to μ_R, including rationally dependent values of the frequencies. Such a program was carried on in [18]. The conclusion, in the essence, is as follows: L_4, L_5 are Nekhoroshev stable for $0 < \mu < \mu_R$, with the possible exclusion of just five isolated values of μ, for which further analysis is necessary. It is worthwhile to remark that in a small subinterval (containing the Earth-Moon mass ratio) the stability is expected to be weaker — the "Nekhoroshev exponents" are smaller — since quasi–convexity is there absent and a weaker steepness property must be invoked. A few more words on the subject can be found in [11].

C. Other applications to Celestial Mechanics. Besides the above quoted studies on the stability of Lagrangian equilibria, there are a few other papers where (mathematically rigorous) Hamiltonian perturbation theory is applied to concrete problems of Celestial Mechanics. Among them, [74] and [63] concern the stability of the asteroids in the main belt, in the light of Nekhoroshev theorem. The model considered is a massless asteroid perturbed by several planets. The former paper is devoted to a general setting of the problem, and to a review of the Nekhoroshev theorem from the point of view of astronomers; the latter explores the phase space of the asteroids, looking for the applicability of the theorem. A key point is the fact that (as we remarked in section 6.1) the unperturbed Kepler system is degenerate, and nevertheless the aim is to confine the degenerate variables (eccentricity and inclination of the asteroid), too. In order for this to be possible, it is necessary that the first terms of the normal form remove the degeneracy and provide quasi–convexity or steepness. The procedure does work, and so the orbital elements of the asteroid stay constant for long time, with the exclusion of a few regions of the phase space dominated by low order resonances. As is remarkable, in some cases the assumption of quasi–convexity fails, and to apply the theorem it is necessary to resort to a weaker steepness property. (As far as I know, this is the first example where steepness is shown to be relevant in connection with a physical system. The same steepness assumption was later used in [18], to prove the stability of L_4, L_5.)

Another paper devoted to the stability of the asteroids is [30]. The model here considered is the planar circular restricted three body problem — the Sun, Jupiter and a massless asteroid — to which KAM theorem is known, in principle, to apply, if the mass of Jupiter is sufficiently small. By means of a computer assisted proof, that is it a proof containing rigorously controlled numerical estimates,[23] the authors are able to show that some KAM tori do

[23] Computer assisted proofs were introduced, in the world of dynamical systems, in the eigthees; in the very essence, the computer is assumed to manipulate not real numbers, but (small) real intervals: any function f of, for example, two real variables x and y (like sum, multiplication...) is replaced by a corresponding function F which operates on intervals X and Y, and gives as result an interval Z such that $f(x, y) \in Z$ for any choice of $x \in X$ and $y \in Y$. In this way the error step by step is kept controlled (the intervals of course grow), and the overall work is rigorous in the mathematical sense.

survive for realistic values of Jupiter's mass and for orbital parameters of the asteroid close to realistic ones. The result was significantly improved in [31], where it was shown that nearby KAM tori do exist in the same constant energy surface; this is relevant because, thanks to the low dimension of the model, such invariant surfaces provide a topological obstruction for the motion of the asteroid, which turns out to be stable for any choice of the orbital paremeters in between.

Finally we mention [57], where the stability of three body planetary systems (a star and two planets) is studied in the light of KAM theorem. The system is treated in the so–called secular approximation, in which two fast angles, actually the mean anomalies of the planets, are averaged out; the resulting model, taking into account the overall conservation of momentum and angular momentum, has two degrees of freedom. By means of a computed assisted proof, the authors are able to prove that KAM theorem applies to it, and invariant tori do exist, for a range of values of the masses including the actual masses of the Sun–Jupiter–Saturn system and for orbital elements of the planets close to the realistic ones.

D. Symplectic integration schemes. Numerical integration schemes for ordinary differential equations

$$\dot{x} = F(x) \ , \qquad x \in \mathbb{R}^n \ ,$$

are based on replacing the continuous flow Φ_F^t by the iteration of a map $\Psi_\tau = \mathbb{I} + \mathcal{O}(\tau)$, $\tau \ll 1$, such that

$$\|\Psi_\tau - \Phi_F^\tau\| = \mathcal{O}(\tau^{k+1}) \ ;$$

k is said to be the order of the algorithm. It is a common experience that for Hamiltonian systems the best results are obtained by the so–called symplectic methods, that is when the map Ψ_τ shares with the exact Hamiltonian flow Φ_H^τ the property of being a canonical transformation. In some cases (see, for a review, [16]) the use of symplectic algorithms allowed to reach incredibly accurate results.[24]

A good understanding of symplectic algorithms can be acheived by means of perturbation theory. In essence, it can be proven [27] [64] that for each symplectic algorithm Ψ_τ there exists an "interpolating Hamiltonian" K_τ (also called the "modified Hamiltonian"), such that

$$\|\Psi_\tau - \Phi_{K_\tau}^\tau\| = \mathcal{O}(e^{-1/\tau}) \ ; \tag{7.1}$$

moreover K_τ is close to H, namely if Ψ_τ is of order k, then

$$\|H - K_\tau\| = \mathcal{O}(\tau^k) \ . \tag{7.2}$$

For reasonable values of τ, the $\mathcal{O}(e^{-1/\tau})$ term is (much) smaller than the unavoidable round–off errors, so if these, as is typical, are neglected, the exponentially small term should also be. Equations (7.1) and (7.2) are interpreted as

[24]For example, for a problem of molecular collisions, it was possible to measure energy exchanges between different degrees of freedom as small as one part over 10^{100}.

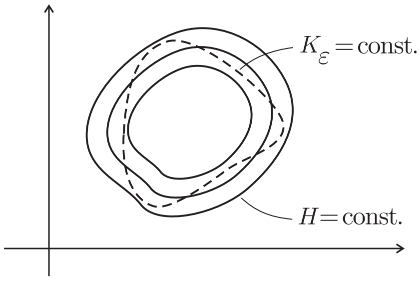

Figure 45: *The surfaces of constant energy of H and of K_τ.*

follows: whenever a Hamiltonian system is numerically integrated by means of a symplectic algorithm, the Hamiltonian H is replaced by a nearby one K_τ, and then the integration is essentially exact. This fact has several nice consequences, and explains in particular the accuracy in energy conservation, with no drift in time, that symplectic integrators are known to exhibit. Indeed the algorithm conserves K_τ, so the error in energy conservation after s iterations of the map,

$$\Delta E(s) = H(\Psi_\tau^s(x)) - H(x) \ ,$$

is not really an error but the difference between H and K_τ, and is bounded (uniformly in s) by

$$|\Delta E(s)| \le |H(\Psi_\tau^s(x)) - K_\tau(\Psi_\tau^s(x))| + |H(x) - K_\tau(x)| = \mathcal{O}(\tau^k) \ ;$$

this is in fact evident, if one looks at figure 45. For special problems like molecular collisions, or adiabatic invariants, results greatly improve, see [27] [16]. Of course, it is not obvious that replacing H by K_τ is an innocent operation: for example, if H is integrable, then K_τ generically is not. More generally, whenever H has some peculiar property, which is considered to be relevant and is not shared by nearby Hamiltonians, then numerical results risk to be unreliable (but non symplectic algorithms would hardly do better). Whenever instead the property one is looking at is structurally stable, i.e. small changes in the Hamiltonian do not alter it in an essential way, then symplectic algorithms should be considered as reliable. The latter is the typical situation of perturbation theory: if $H = h + \varepsilon f$ is the Hamiltonian one aims to study, then the effectively studied Hamiltonian has the form

$$K_\tau = h + \varepsilon f + \tau^k g \ , \tag{7.3}$$

and unless f has some special requirements, which might not be satisfied by g, the only attention to be paid[25] is that $\tau^k \ll \varepsilon$.

[25] As is remarkable, splitting algorithms like (6.15) have the nice additional property (see for example [13]) that the correction g in (7.3) is already proportional to ε, so in this case it is not even necessary to reduce the time step when ε is reduced.

In recent years the study of symplectic algorithms, more generally of the "structure preserving" algorithms, deserved great attention in the literature. Unfortunately we cannot further discuss the problem, and limit ourselves to suggest, beside the already quoted papers, the recent book [65], which is rather exhaustive and moreover contains all the important references.

E. Miscellanea. A very interesting subject, that we did not even touch in these lectures, is the possible extension of perturbation methods to systems with infinitely many degrees of freedom, either discrete (for example FPU–like systems in the thermodynamic limit) or continuos (vibrating strings, the electromagnetic field...). General theorems like KAM and Nekhoroshev do not extend to systems with infinitely many degrees of freedom, since many constants entering the statements have a bad dependence on the number n of degrees of freedom. The problem is rather difficult, and papers on the subject are by necessity somehow technical. Most results concern systems with infinitely many degrees of freedom, which however have only a finite number of independent frequencies, or in principle have infinitely many frequencies, but only a finite number of them is effectively excited. A pionering paper is [48], where the authors study a discrete system of infinitely many oscillators with random frequencies, weakly coupled by short range nonlinear terms. The linear system is known to have, with great probability, localized excitations, which can be regarded as finite–dimensional tori; by an extension of KAM theory, it is proven that some of them do persist in the nonlinear system, too. (See also, for long time stability of finite energy excitations in the same model, [21].) A more recent paper on a discrete model of infinitely many rotators is [35].

Concerning nonlinear wave equations, they have been studied within Hamiltonian perturbation theory by several authors. Recent papers on the subject include, for example, [36] [67] [85] [29] [8] [7]; other papers are there referenced. A suggested starting point is the review paper [8].

We finally mention a curious problem, known as the problem of the stability of the Riemann ellipsoids, which is interesting also for historical reasons, since authors like Laplace, Jacobi, Riemann, Dirichlet, Dedekind and Chandrashechar made some work on it. The question concerns the stability of the ellipsoidal configurations of a self–gravitating incompressible fluid. Modern perturbation theory, in the Nekhoroshev version, was successfully applied to this problem, too, leading to new results [46].

Bibliography

[1] R. Abrham and J.E. Marsden, *Foundations of Mechanics.* Benjamin Cummings, Reading 1981.

[2] H. Andoyer, *Cours de Méchanique Celeste.* Gautier-Villars, Paris 1923.

[3] V.I. Arnol'd, *Méthodes Mathématiques de la Méchanique Classique.* MIR, Moscow 1976.

[4] V.I. Arnol'd, *Instability of dynamical systems with several degrees of freedom.* Sov. Math. Doklady **5**, 581–585 (1964).

[5] V.I. Arnold, *Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian*, Uspeki Mat. Nauk. **18,** 13-35 (1963) [Russ. Math. Surv. **18**, 9 (1963)]; *Small denominators and problems of stability of motion in Classical and Celestial Mechanics*, Uspeki Mat. Nauk. **18**, 91 (1963) [Russ. Math. Surv. **18,** 85-191 (1963)].

[6] V.I. Arnol'd, V.V. Kozlov and A.I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics.* In: *Dynamical Systems III,* V.I. Arnol'd Ed. Encyclopedia of Mathematical Sciences, Springer, Berlin 1988.

[7] D. Bambusi, *Birkhoff normal form for some nonlinear PDEs.* Preprint (http://users.mat.unimi.it/ bambusi/birkhof.ps.gz).

[8] D. Bambusi and N.N. Nekhoroshev, *Long time stability in perturbations of completely resonant PDE's*, Acta Applicandae Matematicae **70**, 1–22 (2002).

[9] R. Barrar, *Convergence of the von Zeipel procedure.* Cel. Mech. **2**, 494–504 (1970).

[10] G. Benettin, *On the Boltzmann conjecture for a classical "freezing" of fast degrees of freedom.* Proceedings of the conference: Boltzmann's legacy 150 years after his birth, Atti Accad. Nazionale dei Lincei **131** 89–105 (1997).

[11] G. Benettin, *Physical applications of Nekhoroshev theory and exponential estimates.* Lectures given at the CIME School: "Theory and applications of Hamiltonian dynamics", Cetraro (Italy), 1998.

[12] G. Benettin, A. Carati and G. Gallavotti, *A rigorous implementation of the Landau–Teller approximation for adiabatic invariants.* Nonlinearity **10**, 479-505 (1997).

[13] G. Benettin, A.M. Cherubini and F. Fassò, *A "changing chart" symplectic algorithm for rigid bodies and other Hamiltonian systems on manifolds.* SIAM Journ. on Sc. Computing **23**, 1189–1203 (2001).

[14] G. Benettin, A.M. Cherubini and F. Fassò, *Regular and chaotic motions of the fast rotating rigid body: a numerical study.* In print in: Discete and Continuous Dynamical Systems B.

[15] G. Benettin and F. Fassò, *Fast rotations of the symmetric rigid body: a general study by Hamiltonian perturbation theory. Part I.* Nonlinearity **9**, 137–186 (1996).

[16] G. Benettin and F. Fassò, *From Hamiltonian perturbation theory to symplectic integrators, an back.* Appl. Num. Math. **29**, 73-87 (1999).

[17] G. Benettin, F. Fassò and M. Guzzo, *Fast rotations of the symmetric rigid body: a study by Hamiltonian perturbation theory. Part II, Gyroscopic rotations.* Nonlinearity **10**, 1695–1717 (1997).

[18] G. Benettin, F. Fassò and M. Guzzo, *Nekhoroshev–stability of L4 and L5 in the spatial restricted three-body problem.* Regular and Chaotic Dynamics **3**, 56-72 (1998).

[19] G. Benettin, F. Fassò and M. Guzzo, *Nekhoroshev stability of all gyroscopic motions of the perturbed symmetric Euler–Poinsot rigid body.* Preprint 2002.

[20] G. Benettin, G. Ferrari, L. Galgani and A. Giorgilli: *An Extension of the Poincaré–Fermi Theorem on the Non-Existence of Invariant Manifolds in Nearly-Integrable Hamiltonian Systems.* Nuovo Cimento **72 B**, 137-148 (1982).

[21] G. Benettin, J. Froehlich and A. Giorgilli, *A Nekhoroshev-type theorem for Hamiltonian systems with infinitely many degrees of freedom.* Comm. Math. Phys. **119**, 95–108 (1988).

[22] G. Benettin, L. Galgani and A. Giorgilli, *Classical Perturbation Theory for a System of Weakly Coupled Rotators.* Nuovo Cimento **89 B**, 89–102 (1985).

[23] G. Benettin, L. Galgani and A. Giorgilli, *Numerical Investigation on a Chain of Weakly Coupled Rotators, in the Light of Classical Perturbation Theory.* Nuovo Cimento **89 B**, 103–119 (1985).

[24] G. Benettin, L. Galgani and A. Giorgilli, *A Proof of Nekhoroshev Theorem for Nearly-Integrable Hamiltonian Systems.* Celestial Mechanics **37**, 1-25 (1985).

[25] G. Benettin, L .Galgani, A. Giorgilli and J.–M. Strelcyn, *A proof of Kolmogorov theorem using canonical transformations defined by the Lie method.* Nuovo Cimento B **79**, 201–223 (1984).

[26] G. Benettin and G. Gallavotti, *Stability of Motions near Resonances in Quasi Integrable Hamiltonian Systems.* Journ. Stat. Phys. **44**, 293 (1986).

[27] G. Benettin and A. Giorgilli, *On the Hamiltonian interpolation of near to the identity symplectic mappings, with application to symplectic integration algorithms.* Journ. Stat. Phys. **73**, 1117–1144 (1994).

[28] G. Benettin, P. Jort and P. Sempio, *Exponentially long equilibrium times in a one dimensional collisional model of a classical gas.* Journ. Stat. Phys. **94**, 871-892 (1999).

[29] J. Bourgain, *On diffusion in high dimensional Hamiltonian systems and PDE's*, J. Anal. Math. **80**, 1–35 (2000).

[30] A. Celletti and L. Chierchia, *On the stability of realistic three body problems,* Comm. Math. Phys. **186**, 413–449 (1997).

[31] A. Celletti and L. Chierchia, paper in preparation.

[32] A. Carati and L. Galgani, *Analog of Plank's formula and effective temperature in classical statistical mechanics far from equilibrium*, Phys. Rev. **E 61**, 4791-4794 (2000).

[33] L. Chierchia and C. Falcolini, *Compensations in small divisor problems.* Comm. Math. Phys. 175, 135–160 (1996).

[34] L. Chierchia and G. Gallavotti, *Smooth prime integrals for quasi–integrable Hamiltonian systems.* Nuovo Cimento B **67**, 277–295 (1982).

[35] L. Chierchia and P. Perfetti, *Second order Hamiltonian equations on T^∞ and almost-periodic solutions.* J. Differential Equations 116, 172–201 (1995).

[36] W. Craig and C.E Wayne, *Newton's method and periodic solutions of nonlinear wave equations.* Comm. Pure Appl. Math. **46**, 1409–1498 (1993).

[37] A. Deprit, *Free rotation of a rigid body studied in phase plane.* Am. J. Phys. **55**, 424 (1967).

[38] L.H. Eliasson *Absolutely convergent series expansions for quasi–periodic motions.* Math. Phys. Electronic J. **2**, paper 4, 33 pp. (1996).

[39] Y. Elskens and D.F. Escande, *Slowly pulsating separatrices sweep homoclinic tangles where islands must be small: an extension of classical adiabatic theory.* Nonlinearity **4**, 615–667 (1991).

[40] Y. Elskens and D.F. Escande, *Infinite resonance overlap: a natural limit for Hamiltonian chaos.* Physica **62** D, 66–74 (1993).

[41] F. Fassò, *Lie series method for vector fields and Hamiltonian perturbation theory*, J. Appl. Math. Phys. (ZAMP) **41**, 843–864 (1990).

[42] F. Fassò, *Hamiltonian perturbation theory on a manifold.* Cel. Mech. & Dyn. Astr. **62,** 43–69 (1995).

[43] F. Fassò, *The Euler–Poinsot top: a non-commutatively integrable system without global action–angle coordinates.* J. Appl. Math. Phys. **47**, 953-976 (1996).

[44] F. Fassò, *Notes on Finite Dimensional Integrable Dynamical Systems.* Unpublished lectures, 1999.

[45] F. Fassò, M. Guzzo and G. Benettin, *Nekhoroshev-stability of elliptic equilibria of Hamiltonian systems.* Comm. Math. Phys. **197**, 347–360 (1998).

[46] F. Fassò and D. Lewis: *Stability properties of the Riemann ellipsoids*, Archive for Rational Mechanics and Analysis 158, 259-292 (2001).

[47] E. Fermi, *Beweis, dass ein mechanisches Normalsystem im allgemainen quasi–ergodisch ist.* Physik. Zeitschr. **24**, 261–265 (1923). Italian version in: Nuovo Cimento **26**, 105–113 (1923).

[48] J. Fröhlich, T. Spencer and C.E. Wayne, *Localization in disordered, nonlinear dynamical systems.* J. Statist. Phys. **42**, 247–274 (1986).

[49] G. Gallavotti, *Quasi–Integrable Mechanical Systems,* in *Critical phenomena, Random Systems, Gauge Theories,* K. Osterwalder and R. Stora editors, *Les Houches,* Session XLIII, 1984 (North–Holland, Amsterdam 1986).

[50] G. Gallavotti, *Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable hamiltonian systems. A review.* Reviews Math. Phys. **6**, 343–411 (1994).

[51] G. Gentile, *A proof of existence of whiskered tori with quasi flat homoclinic intersections in a class of almost integrable systems.* Forum Math. 7, 709–753 (1995).

[52] A. Giorgilli, *Rigorous results on the power expansions for the integrals of a Hamiltonian system near an elliptic equilibrium point.* Ann. Inst. Henri Poincaré - Physique Thèorique **48**, 423–439 (1988).

[53] A. Giorgilli, A. Delshams, E. Fontich, L. Galgani and C. Simó, *Effective Stability for a Hamiltonian System near an Elliptic Equilibrium Point, with an Application to the Restricted three Body Problem.* J. Diff. Eq. **77**, 167-198 (1989).

[54] A. Giorgilli and G. Galgani, *Rigorous estimates for the series expansions of Hamiltonian perturbation theory.* Celestial Mech. **37** 95–112 (1985).

[55] A. Giorgilli and U. Locatelli, *Kolmogorov theorem and classical perturbation theory.* ZAMP **48**, 220-261 (1997).

[56] A. Giorgilli and U. Locatelli, *On classical series expansions for quasi-periodic motions.* Math. Phys. Electronic J. **3**, paper 5 (1997).

[57] A. Giorgilli and U. Locatelli, *Invariant tori in the secular motions of the three-body planetary systems.* Cel. Mech., to appear.

[58] A. Giorgilli and Ch. Skokos, *On the stability of the Trojan asteroids*, Astron. Astroph. **317**, 254-261 (1997)

[59] W. Gröbner, *Serie di Lie e loro applicazioni.* Cremonese, Roma 1973.

[60] M. Guzzo: *Nekhoroshev stability of quasi–integrable degenerate Hamiltonian Systems.* Regular and Chaotic Dynamics 4, 78-102 (1999).

[61] M. Guzzo and G. Benettin, *A spectral formulation of the Nekhoroshev theorem and its relevance for numerical and experimental data analysis.* Discr. Cont. Dyn. Sys. B **1**, 1–28 (2001).

[62] M. Guzzo, F. Fassò and G. Benettin, *On the stability of elliptic equilibria.* Math. Phys. Electronic J. **4**, paper 1, 16 pp. (1998).

[63] M. Guzzo and A. Morbidelli, *Construction of a Nekhoroshev like result for the asteroid belt dynamical system.* Cel. Mech. & Dyn. Astr. **66,** 255-292 (1997).

[64] E. Hairer, *Backward analysis of numerical integrators and symplectic methods.* Ann. Num. Math. **1**, 107–132 (1994).

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

[66] A.N. Kolmogorov, *On the conservation of quasi-periodic motions for a small change in the Hamiltonian function*, Dokl. Akad. Nauk. SSSR **98,** 527-530 (1954); English translation in G. Casati and G. Ford (Editors), Lect. Notes Phys. No. 93 (Springer, Berlin, 1979).

[67] S. Kuksin, *A KAM-theorem for equations of the Korteweg-de Vries type.* Rev. Math. Math. Phys. **10**, (1998).

[68] V.F. Lazutkin, *The existence of caustics for a billiard problem in a convex domain.* Math. USSR Izvestija **7**, 185–214 (1973).

[69] J.E. Littlewood, Proc. London Math. Soc. (3)**9**, 343 (1959); **9**, 525 (1959).

[70] P. Lochak, *Canonical perturbation theory via simultaneous approximation.* Russ. Math. Surv. **47**, 57-133 (1992).

[71] P. Lochak, *Simultaneous Diophantine approximation in classical perturbation theory: Why and what for?*. Preprint.

[72] P. Lochak and A.I. Neishtadt: *Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian.* Chaos **2**, 495–499 (1992).

[73] Mishenko and Fomenko, *Generalized Liouville method of integration of Hamiltonian systems.* Funct. Anal. Appl. **12**, 113–121 (1978).

[74] A. Morbidelli and M. Guzzo, *The Nekhoroshev theorem and the asteroid belt dynamical system,* Cel. Mech. & Dyn. Astr. **65,** 107–136 (1997).

[75] J. Moser, *On invariant curves of area–preserving maps of an anulus,* Nach. Akad. Wiss. Göttingen, Math. Phys. Kl., **1**, 1-20 (1962).

[76] A.I. Neishtadt, *Change of the adiabatic invariant at a separatrix,* Sov. J. Plasma Physics **12**, 558–573 ('986).

[77] A.I. Neishtadt, *On the change in the adiabatic invariant on crossing of a separatrix in systems with two degrees of freedom,* PMM USSR **51**, 586–592 (1987).

[78] A.I. Neishtadt, V.V. Sidorenko and D.V. Treshev, *Stable periodic motions in the problem of passage through a separatrix,* Chaos **7**, 2–11 (1997).

[79] N.N. Nekhoroshev, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems.* Usp. Mat. Nauk **32**:6, 5-66 (1977) [Russ. Math. Surv. **32**:6, 1-65 (1977)].

[80] N.N. Nekhoroshev, *Action–angle variables and their generalizations.* Trudy Moskov Mat. Obsc. **26**, 181–198 (1972) [Trans. Moskow Math. Soc. **26**, 180-198 (1972)].

[81] L. Niederman, *Nonlinear stability around an elliptic equilibrium point in an Hamiltonian system.* Nonlinearity **11** 1465–1479 (1998).

[82] L. Niederman, *Exponential stability for small perturbations of steep integrable dynamical systems.* Prépublication d'Orsay numéro 2000-73.

[83] J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets.* Comm. Pure Appl. Math. **35**, 653–696 (1982).

[84] J. Pöschel, *Nekhoroshev estimates for quasi–convex Hamiltonian systems.* Math. Z. **213**, 187–216 (1993).

[85] J. Pöschel, *A KAM-theorem for some nonlinear partial differential equations.* Ann. Scuola Norm. Sup. Pisa Cl. Sci. **23**, 119–148 (1996).

[86] J. Pöschel, *On Nekhoroshev's estimate at an elliptic equilibrium.* Internat. Math. Res. Notices 1999, no. 4, 203–215.

[87] H. Poincaré, *Les Méthodes Nouvelles de la Méchanique Céleste,* Vol. 1 (Gautier–Villars, Paris, 1892).

[88] F. Scheck, *From Newton Laws to Deterministic Chaos.* Springer, Berlin 1990.

[89] C.L. Siegel, *Iterations of analytical function.* Ann. Math. **43,** 607–612 (1942).

[90] C. Siegel and J. Moser, *Lectures on Celestial Mechanics.* Springer, Berlin 1971.

[91] J. Touma and J. Wisdom, *Lie–Poisson integrators for rigid body dynamics in the Solar System.* Astr. J. **107**, 1189–1202 (1994).