Uspekhi Mat. Nauk 47:6 (1992), 59–140 Russian Math. Surveys 47:6 (1992), 57–133

Canonical perturbation theory via simultaneous approximation

P. Lochak

"D'ailleurs, ce qui nous rend ces solutions périodiques si précieuses, c'est qu'elles sont, pour ainsi dire, la seule brèche par où nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable."

Henri Poincaré, Les méthodes nouvelles de la mécanique céleste (§36)

Contents

Chapter I

Introduction

In this article, we wish to present a new method to deal with problems related to canonical (that is, symplectic or Hamiltonian) perturbation theory. The familiar model situation consists in the perturbation of an integrable Hamiltonian system; that is, one considers the system in phase space governed by

$$(1) \qquad H(p,q) = h(p) + \varepsilon f(p,q), \quad (p,q) \in \mathbb{R}^n \times \mathbb{T}^n, \quad \mathbb{T} = \mathbb{R}/\mathbb{Z}.$$

As usual, (p, q) denote action-angle variables of the integrable Hamiltonian h, and ε is a small parameter.

Our main result here will be a substantial improvement, both quantitative and qualitative, of Nekhoroshev's results ([43], [44]) about the stability of the action variables over exponentially long times, when the unperturbed Hamiltonian h is *quasi-convex*, by which we mean (following Nekhoroshev) that the energy surface $h(p) = E$ is strictly convex (for a certain range of the energy E). Of course, any convex function is a fortiori quasi-convex.

Under such a hypothesis, the following general estimate holds:

$$(2) \qquad \|p(t) - p(0)\| \leq R(\varepsilon) \text{ when } |t| \leq \mathcal{T}(\varepsilon) \text{ and } |\varepsilon| \leq \varepsilon_0,$$

where $R(\varepsilon) = O(\varepsilon^b)$ and $\mathcal{T}(\varepsilon)$ has the orderof $\exp(c/\varepsilon^a)$. We call $\mathcal{T}(\varepsilon)$ the *stability time*, $R(\varepsilon)$ the *radius of confinement*, and $\varepsilon_0 > 0$ the *threshold of validity*. A gross but important evaluation of the size of $\mathcal{T}(\varepsilon)$ and $R(\varepsilon)$ is provided by the numbers (a, b), $0 < a, b \leqslant 1$, which we call the *stability exponents*.

(2) was proved by Nekhoroshev ([43], [44]) under the assumption that H is *analytic* (this cannot be dispensed with) and h is *steep*, which is a weak, C^∞ generic condition. Here we shall have to work under the more stringent hypothesis that h is quasi-convex and we have to insist that this does not merely allow for a technical simplification; the proof method used below simply does not carry over to the steep case. It is interesting to note in this respect that recent works also point to the specificity of quasi-convex systems (we return to this in Chapter IV, §4). We also mention that Nekhoroshev's original proof has been rewritten precisely in the convex case (see [6] and [7]).

Below (see Chapters II and III), we shall improve on the known values of the exponents in the quasi-convex case. In particular, we show that $a > 1/(2n+1) - \eta$ for any $\eta > 0$, and believe that in general $a \leqslant 1/(2n)$. The latter assertion essentially means that over a longer timescale Arnol'd's diffusion may, and in fact generically will, switch on, so that stability results are effectively broken. This of course seems quite hard to prove and we shall content ourselves with a heuristic argument which points in that direction (see Chapter V, §2). Another striking qualitative phenomenon which we explore is the fact that over finite but exponentially long times, resonance stabilizes the motion. In fact, we prove a precise local version of (2) which implies that the stability time is increased when the initial condition is resonant.

These results are easy consequences of the proof itself, which is substantially different from the usual one, and less cumbersome. To appreciate the difference, it suffices to say that the usual ingredients of canonical perturbation theory, such as small divisors, ultraviolet cut-off, resonance surfaces, or even

Fourier series in general, simply do not enter the proof at all. This stems from a change of viewpoint, and we devote the end of this introduction to a more general discussion because, as suggested in the title, we believe that this method may have a wide range of applications, some of which are explored or suggested in Chapters IV and V.

Loosely speaking, the usual point of departure of canonical perturbation theory consists in the observation that the perturbed system could be reduced to an integrable one, using the apparatus of normal form theory, in a region of phase space which is free from resonances. Since no such *open* domain of phase space exists in general (leaving aside the linear or isochronous case, that is, the perturbation of harmonic oscillators), non-integrability is the rule rather than the exception, as was brought to light essentially by Poincaré, and the aim is to understand what results can be obtained in spite of the unavoidable existence of resonances. Of such nature for instance was Kolmogorov's remarkable insight about the existence of invariant tori.

Here, in some sense, things are turned inside out, as one tries to view resonances not as a hindrance but as an opportunity. To this end, one focuses first on the fully resonant situation, which is embodied in closed orbits. Let us be more specific; at the level of "classical" perturbation theory, which culminates in results of Nekhoroshev type, one should consider the objects (resonant surfaces, closed orbits, and so on) related to the *unperturbed* system. Referring to (1) and denoting the frequency vector by $\omega(p) = \nabla h(p) \in \mathbb{R}^n$, the closed orbits of the unperturbed system simply correspond to *rational* vectors ω, that is, vectors which are multiples of integer ones: if $\omega_0 = \omega(p_0)$ is rational, the torus $p = p_0$ is filled with closed orbits of the unperturbed system, with common period T such that $T\omega_0 \in \mathbb{Z}^n$.

In order to prove estimates such as (2), one first studies stability in the neighbourhood of these periodic tori. This is the object of Chapter II, and all the analysis it requires is *one* phase or time averaging. Then, given an arbitrary point in phase space, or rather in action space, it may be approximated by points corresponding to periodic tori. The rate of approximation and the growth of the corresponding periods are related, for a generic point, by the simplest approximation result, namely Dirichlet's theorem. This procedure will enable us (in Chapter III) to prove (2) and its local version, which depends on the properties of the initial conditions.

Returning to more general considerations, one realizes that the approach relies on a kind of duality (using the word with a non-technical meaning), which may be expressed in several ways. First, at the dynamical level, there is the relationship between time and phase averaging, which lies at the root of ergodic theory. For linear flows on tori, it can of course be made much more explicit, using in particular the notion of approximate recurrence times. This corresponds, in the theory of Diophantine approximation, to the "duality" between *linear* and *simultaneous* approximations. Given $\omega \in \mathbb{R}^n$, the former deals with the size of the linear forms $\omega \cdot k$ ($k \in \mathbb{Z}^n$; ordinary dot product), the latter with the approximation of the straight line $T\omega$ ($T \in \mathbb{R}$) by the integer lattice $\mathbb{Z}^n$; that is, one is interested in the distance $\mathrm{dist}(T\omega, \mathbb{Z}^n)$ as T varies, say along the positive semiaxis. In dynamical terms, the first describes the distribution of the small divisors, the latter that of the closed orbits. In some sense, both contain the same arithmetical information about ω, as asserted by *transfer principles* which originated in the work of A. Khinchin (the simplest and most useful ones are recalled in Appendix 1). Notice however that the information is encoded in a more compact way using simultaneous approximation: it is always "one-dimensional", whatever the dimension of the ambient space. Finally, transfer principles are a—not straightforward—reflection of the projective duality between a linear subspace and its orthogonal complement, and in fact, as far as linear and simultaneous approximations are concerned, between lines and hyperplanes.

As a final word in this introduction, we mention that Appendix 2 has been inserted in order to clarify the discussion in Chapter IV, §2. Also we have tried to keep the reference list to a reasonable length and have accordingly refrained from quoting some classical—and less classical—works, the references to which can be found, for example, in the bibliographies of several of the articles we refer to.

It is a pleasure to acknowledge interesting conversations with N.N. Nekhoroshev and A.I. Neistadt in connection with this work. I wish to thank M.B. Sevryuk for making judicious remarks which helped me prepare the final draft of this paper and also for his contribution to the preparation of the Russian version of this text.

CHAPTER II

STABILITY IN THE NEIGHBOURHOOD OF A PERIODIC TORUS

We shall be interested in Hamiltonians of the type

$$(1) \qquad H(p,q) = h(p) + f(p,q). \quad (p,q) \in \mathbb{R}^n \times \mathbb{T}^n, \quad \mathbb{T} = \mathbb{R}/\mathbb{Z}.$$

$\omega(p) = \nabla h(p)$ denotes the frequency vector of the unperturbed system and we assume in this chapter that $\omega_0 = \omega(0)$ is *rational* of (minimal) period T, that is, $T\omega_0 \in \mathbb{Z}^n$. We use the notation $\Omega = \|\omega_0\|$ (Euclidean norm). h and f are assumed to be defined and *analytic* in a neighbourhood of the origin, more precisely on a complex domain $D = D(R, \rho, \sigma)$ ($\rho > 0$, $\sigma > 0$) defined as follows: let B_R be the real ball of radius R around the origin, then

$$(2) \qquad D(R,\rho,\sigma) = \{(p,q) \in \mathbb{C}^{2n},\ \mathrm{dist}(p, B_R) \le \rho,\ |\mathrm{Im}\, q| \le \sigma\},$$

where $|\mathrm{Im}\, q| = \sup_i |\mathrm{Im}\, q_i|$; h and f are supposed to be continuous on the boundary of D. The real part of D is of course nothing but $B_{R+\rho} \times \mathbb{T}^n$.

When $0 \leqslant \delta \leqslant \rho$, $0 \leqslant \xi \leqslant \sigma$, we denote by $D-(\delta, \xi)$ the domain $D(R, \rho-\delta, \sigma-\xi)$.

The norm $\|\cdot\|_D$ is simply the sup norm (L^∞) over D and we write $\|f\|_D = \varepsilon E$. Notice that we have not written the small parameter $\varepsilon \geqslant 0$ explicitly in (1) and we introduce the letter E essentially because in this way ε becomes a non-dimensional quantity and all the formulae we get will be dimensionally correct. The reader who does not find it useful may set $E = 1$ in the sequel; he may also set $\Omega = 1$, using a rescaling of the time variable, but again we find the formulae more suggestive this way. Finally, to define the size of the perturbation, one could also compare the norm of ∇f with Ω; we do not even assume, as one could, that f has zero mean with respect to q, for any p.

We consider the case when h is *convex*; the slight modifications needed when h is only assumed to be *quasi*-convex will be indicated at the end of this section. We denote the Hessian matrix by $A(p) = \nabla^2 h(p)$ and suppose that it is positive definite (if not, change t into $-t$ and H into $-H$), more precisely that m (respectively M) is a lower (respectively upper) bound of the spectrum of A over D. Explicitly:

$$\|A(p)v\| \leq M\|v\|, \quad (A(p)v, v) \geq m\|v\|^2, \quad \text{for any} \quad p \in D \cap \mathbb{R}^n \quad \text{and} \quad v \in \mathbb{R}^n,$$

where $0 < m \leqslant M$ and the dot denotes the usual scalar product.

We shall first prove an iterative lemma which consists in a simple one-phase averaging procedure and constitutes the only analytical result that will be needed in all this paper. From it there will easily follow three allied statements which describe the stability near periodic tori. To give a more precise idea of the type of results we have in mind, let us state Theorem 1B (see below) in a slightly informal way.

Let $H(p, q) = h(p)+f(p, q)$ be a perturbation of a convex integrable Hamiltonian such that $p = 0$ is a periodic torus of period T. Let $(p(t), q(t))$ denote the trajectory starting at $(p(0), q(0))$. Then, if $\|p(0)\| \leqslant r_0\varepsilon^{1/3}$, the estimate $\|p(t)\| \leqslant R_0\varepsilon^{1/3}$ holds when $\varepsilon \leqslant \varepsilon_0$ and $|t| \leqslant \mathcal{T}(\varepsilon) = \mathcal{T}_0 \exp((\tau/T)\varepsilon^{-1/3})$.

All the constants will be explicitly computed as simple functions of the parameters Ω, m, M, ρ, σ, E and T, the physical meaning of which is clear. The number n of degrees of freedom will not appear.

In order to state the iterative lemma, we need yet another simple notion. With a function $g(q)$ on the torus one associates its *time average* $\langle g\rangle$ along the orbits of the linear flow defined by ω_0:

$$\langle g\rangle(q) = \frac{1}{T}\int_0^T g(q+\omega_0 t)dt.$$

We shall say that g is *resonant* (with respect to ω_0) if $g = \langle g\rangle$, which simply means that g is constant along the orbits of the flow directed along ω_0.

Iterative lemma. *Let $H(p, q)$ be an analytic Hamiltonian on $D = D(R, \rho, \sigma)$ such that*

$$H(p,q) = h(p) + Z(p,q) + N(p,q), \tag{3}$$

where Z is resonant with respect to ω_0 (p comes in as a parameter) whereas $\langle N \rangle = 0$. Suppose that $\|Z+N\|_D \leqslant \varepsilon E$ and $\|N\|_D \leqslant \eta E$. Let ξ and δ satisfy

$$0 < \delta < \rho, \quad 0 < \xi < \sigma \quad \text{and} \quad \xi\delta \geq 2TE\eta; \tag{4}$$

then there exists a canonical transformation $C : D' \to D$ with $D' = D-(\delta, \xi)$ such that C is one-to-one and its image $C(D')$ satisfies

$$D - (\delta/2, \xi/2) \supset C(D') \supset D - (3\delta/2, 3\xi/2).$$

C is analytic and preserves reality, that is, $C(D' \cap \mathbb{R}^{2n}) \subset D \cap \mathbb{R}^{2n}$. Moreover if $C(p', q') = (p, q)$, one has the estimates $\|p'-p\| \leqslant \delta/2$, $\|q'-q\| \leqslant \xi/2$, and denoting $H' = H \circ C$, the function H' can be written in the form (3) *(using primed letters) with*

$$\varepsilon' \leq \varepsilon + \frac{1}{2}\eta', \tag{5$_1$}$$

$$\eta' \leq 5T\eta\left[M(R+\rho)\xi^{-1} + 4E\varepsilon(\xi\delta)^{-1}\right]. \tag{5$_2$}$$

To anticipate a bit, the idea is that by using the lemma one will progressively make the *non*-resonant part N as small as possible, keeping the size of the resonant term Z roughly constant. One then uses Hamilton's equation $\dot{p} = -\dfrac{\partial H}{\partial q}$ and the crucial fact that $\omega_0 \cdot \dfrac{\partial Z}{\partial q} = 0$ (here one may think of the "standard" case when $\omega_0 = (\nu, 0, ..., 0)$, with period $T = 1/\nu$). This implies that $\omega_0 \cdot \dot{p} = -\omega_0 \cdot \dfrac{\partial N}{\partial q}$, which will be very small. Thus, one has almost eliminated the possibility of a drift along the direction of ω_0. This will in turn guarantee stability, using a simple geometric argument (see (21) and the reasoning below).

Turning to the proof of the iterative lemma, C is built with the help of a Lie series, that is, as the time 1 map of an auxiliary Hamiltonian $\chi(p, q)$. We shall need a lemma in order to estimate gradients and Poisson brackets, which represents a slightly elaborate use of the Cauchy formula. Below, $\partial f/\partial p$ and $\partial f/\partial q$ (n-vectors) denote of course the gradients of f with respect to the variables p and q; $\{\cdot,\cdot\}$ is the Poisson bracket:

$$\{f,g\} = \frac{\partial f}{\partial p}\frac{\partial g}{\partial q} - \frac{\partial f}{\partial q}\frac{\partial g}{\partial p} \quad \text{(dot products).}$$

Finally the norm of a vector-valued function on D is defined as the supremum over D of the Euclidean norm of its value.

***Lemma* 1.** *Let f and g be analytic on D (continuous at the boundary); then*

$$\left\|\frac{\partial f}{\partial q}\right\|_{D-(0,\xi)} \leq \frac{1}{\xi}\|f\|_D, \quad \left\|\frac{\partial f}{\partial p}\right\|_{D-(\delta,0)} \leq \frac{1}{\delta}\|f\|_D. \tag{6}$$

Suppose that $0 \leqslant \xi' < \xi$ *and* $0 \leqslant \delta' < \delta$; *then*

$$\|\{f,g\}\|_{D-(\delta,\xi)} \leq \left(\inf[\xi(\delta-\delta'),\delta(\xi-\xi')]\right)^{-1}\|f\|_D\|g\|_{D-(\delta',\xi')}. \tag{7}$$

In particular,

$$\begin{aligned} &\|\{f,g\}\|_{D-(\delta,\xi)} \leq \frac{1}{\xi\delta}\|f\|_D\|g\|_D \quad \text{and} \\ &\|\{f,g\}\|_{D-(\delta,\xi)} \leq \frac{2}{\xi\delta}\|f\|_D\|g\|_{D-\left(\frac{\delta}{2},\frac{\xi}{2}\right)}. \end{aligned} \tag{8}$$

In order to prove the first inequality in (6), one observes that at a given point (p, q)

$$\left\|\frac{\partial f}{\partial q}(p,q)\right\| = \sup_{\|e\|=1}\left\|\frac{d}{dt}\Big|_{t=0} f(p,q+te)\right\|.$$

One then applies Cauchy's formula to the function $t \to f(p, q+te)$ of the complex variable t, defined for $|t| \leqslant \xi$ and continuous at the boundary, when $(p, q) \in D-(0, \xi)$; the second inequality in (6) is proved analogously.

To prove (7), one writes, in a similar way:

$$\{f,g\}(p,q) = \frac{d}{dt}\Big|_{t=0} g\left(p - t\frac{\partial f}{\partial q}, q + t\frac{\partial f}{\partial p}\right)$$

and again applies Cauchy's formula to the function of t appearing on the right-hand side. Here one uses the circle $|t| = \inf[\xi(\delta-\delta'), \delta(\xi-\xi')]\cdot(\|f\|_D)^{-1}$. To justify this, one first notices that $|g(p-t\partial f/\partial q, q+t\partial f/\partial p)| \leqslant \|g\|_{D-(\delta',\xi')}$ when $(p, q) \in D-(\delta, \xi)$ if t satisfies $|t\partial f/\partial q| \leqslant \delta-\delta'$ and $|t\partial f/\partial p| \leqslant \xi-\xi'$; one then applies (6) to show that the function at hand is indeed analytic inside the circle mentioned above (and continuous at the boundary). □

Let us now come back to the construction of the canonical transformation C, as the time 1 map of the Hamiltonian χ. One demands that it satisfy $\|\chi\|_D \leqslant \xi\delta/4$, so that (6) provides the following evaluations for the Hamiltonian vector field:

$$\left\|\frac{\partial \chi}{\partial q}\right\|_{D-\left(0,\frac{\xi}{2}\right)} \leq \frac{2}{\xi}\|\chi\|_D \leq \frac{\delta}{2}, \quad \left\|\frac{\partial \chi}{\partial p}\right\|_{D-\left(\frac{\delta}{2},0\right)} \leq \frac{2}{\delta}\|\chi\|_D \leq \frac{\xi}{2}. \tag{9}$$

This allows one to define C and ensures that the inclusions and inequalities following (4) are satisfied. Let L_χ denote Liouville's operator ($L_\chi(f) = \{\chi, f\}$). The transformation $C = \exp L_\chi$ acts on functions defined on D' and one can compute

$$H' = \exp(L_\chi)H = h + Z + N + \{\chi, h\} + \{\chi, Z+N\} + H' - H - \{\chi, H\}, \tag{10}$$

to be estimated on D' (below, however, we write (p, q) instead of (p', q') for simplicity). One has $\{\chi, h\} = -\omega\cdot\partial\chi/\partial q$ ($\omega = \omega(p)$) and we shall define χ such that it satisfies $\omega_0\cdot\partial\chi/\partial q = N$, which is made possible by the condition $\langle N\rangle = 0$ (see Lemma 2 below). One may then write $H' = h+Z+N''$ with

$$N'' = (\omega_0-\omega)\frac{\partial\chi}{\partial q} + \{\chi, Z+N\} + H' - H - \{\chi, H\}. \tag{11}$$

Then, one defines $Z' = Z+\langle N''\rangle, N = N''-\langle N''\rangle$ and finds that

$$\|Z'+N'\|_{D'} \le \|Z\|_{D'} + \|N\|_{D'} \le \varepsilon E + \|N''\|_{D'},$$
$$\|N'\|_{D'} = \left\|N''-\langle N''\rangle\right\|_{D'} \le 2\|N''\|_{D'},$$

so that $\varepsilon' \leqslant \varepsilon+(1/2)\eta'$, which is $(5)_1$, with $\eta' \leqslant 2\|N''\|_{D'}/E$. We have used the fact that the operation $\langle\cdot\rangle$ of averaging is a projection of unit norm, which means that for any function g, $\|\langle g\rangle\|_D \leqslant \|g\|_D$. This is obvious from the definition of $\langle\cdot\rangle$.

It remains to evaluate $\|N''\|_{D'}$. To this end one writes

$$\|N''\|_{D'} \le \|\omega_0-\omega\|_{D'}\left\|\frac{\partial\chi}{\partial q}\right\| + \left\|\{\chi, Z+N\}\right\|_{D'} + \frac{1}{2}\left\|\{\chi,\{\chi,H\}\}\right\|_{D-(\frac{\delta}{2},\frac{\xi}{2})}. \tag{12}$$

The last term comes from Taylor's formula and the fact that $C(D')\subset D-(\delta/2, \xi/2)$.

The first two terms on the right-hand side are easily estimated:

$$\|\omega_0-\omega\|_{D'} \le M(R+\rho) \quad\text{and}\quad \left\|\frac{\partial\chi}{\partial q}\right\|_{D'} \le \xi^{-1}\|\chi\|_D$$

(using (6));

$$\left\|\{\chi, Z+N\}\right\|_{D'} \le (\xi\delta)^{-1}\|\chi\|_D\|Z+N\|_D \le \varepsilon E(\xi\delta)^{-1}\|\chi\|_D$$

(using (8) and the definition of ε).

To estimate the third term, one takes advantage of the definition of χ and writes

$$\{\chi, H\} = -N + (\omega_0-\omega)\frac{\partial\chi}{\partial q} + \{\chi, Z+N\},$$

so that

$$\left\|\{\chi,\{\chi,H\}\}\right\| \le \|\{\chi,N\}\| + \left\|\left\{\chi,(\omega_0-\omega)\frac{\partial\chi}{\partial q}\right\}\right\| + \left\|\{\chi,\{\chi,Z+N\}\}\right\|,$$

and there are again three terms to be estimated over $D-(\delta/2, \xi/2)$. The first is dealt with as above:

$$\left\|\{\chi,N\}\right\|_{D-(\frac{\delta}{2},\frac{\xi}{2})} \le 4(\xi\delta)^{-1}\|\chi\|_D\|N\|_D \le 4\eta E(\xi\delta)^{-1}\|\chi\|_D.$$

The other two necessitate a repeated application of the inequalities (8):

$$\begin{aligned}\left\|\left\{\chi,(\omega_0-\omega)\frac{\partial\chi}{\partial q}\right\}\right\|_{D-(\frac{\delta}{2},\frac{\xi}{2})} &\le 8(\xi\delta)^{-1}\|\chi\|_D\left\|(\omega_0-\omega)\frac{\partial\chi}{\partial q}\right\|_{D-(\frac{\delta}{4},\frac{\xi}{4})}\\ &\le 8(\xi\delta)^{-1}\|\chi\|_D M(R+\rho)4\xi^{-1}\|\chi\|_D\\ &= 32M(R+\rho)\xi^{-2}\delta^{-1}\|\chi\|_D^2;\end{aligned}$$

and

$$\left\|\{\chi,\{\chi,Z+N\}\}\right\|_{D-\left(\frac{\delta}{2},\frac{\xi}{2}\right)} \leq 8(\xi\delta)^{-1}\|\chi\|_D\left\|\{\chi,Z+N\}\right\|_{D-\left(\frac{\delta}{4},\frac{\xi}{4}\right)}$$
$$\leq 8(\xi\delta)^{-1}\|\chi\|_D^2 16(\xi\delta)^{-1}\|Z+N\|_D$$
$$= 128\varepsilon E\xi^{-2}\delta^{-2}\|\chi\|_D^2.$$

Gathering terms together, this elementary (but admittedly forbidding) computation furnishes

$$\|N''\|_{D'} = \eta'\frac{E}{2} \leq M(R+\rho)\xi^{-1}\|\chi\|_D + \varepsilon E(\xi\delta)^{-1}\|\chi\|_D + 2\eta E(\xi\delta)^{-1}\|\chi\|_D + 16M(R+\rho)\xi^{-2}\delta^{-1}\|\chi\|_D^2 + 64\varepsilon E\xi^{-2}\delta^{-2}\|\chi\|_D^2. \tag{13}$$

To go further, we must compute χ, to which end we use the following lemma, where the variable p is omitted because it plays the role of a dummy parameter.

***Lemma* 2.** *Let $g(q)$ be a function with zero mean value (with respect to ω_0): $\langle g\rangle = 0$; then the equation*

$$\omega_0\frac{\partial\chi}{\partial q} = g \tag{14}$$

possesses the explicit solution

$$\chi(q) = \frac{1}{T}\int_0^T g(q+\omega_0 t)t\,dt. \tag{15}$$

In particular, it satisfies

$$\|\chi\| \leq \frac{T}{2}\|g\| \tag{16}$$

for any translation-invariant norm $\|\cdot\|$ defined on the space of measurable functions on the torus.

The two-line proof reduces to an integration by parts:

$$\omega_0\frac{\partial\chi}{\partial q} = \frac{1}{T}\int_0^T \omega_0\frac{\partial g}{\partial q}(q+\omega_0 t)t\,dt = \frac{1}{T}\int_0^T \frac{d}{dt}g(q+\omega_0 t)dt$$
$$= \frac{1}{T}g(q+\omega_0 t)t\Big|_0^T - \frac{1}{T}\int_0^T g(q+\omega_0 t)dt = g(q). \quad \square$$

Coming back to (13), χ satisfies (14) with $g = N$, and choosing the solution (15), we get $\|\chi\|_D \leqslant (T/2)\eta E$, so (13) becomes

$$\eta' \leq \eta T\left[M(R+\rho)\xi^{-1} + \varepsilon E(\xi\delta)^{-1} + 2\eta E(\xi\delta)^{-1} + 8MET(R+\rho)\xi^{-2}\delta^{-1}\eta + 32E^2T\xi^{-2}\delta^{-2}\varepsilon\eta\right]. \tag{17}$$

All these estimates are valid under the hypothesis that $\|\chi\|_D \leqslant \xi\delta/4$, that is, $2TE\eta \leqslant \xi\delta$. It allows for a simplification of (17), which also underlines its dimensional correctness. In fact, after inserting this inequality in (17), one gets the second inequality in (5):

$$(5)_2 \qquad \eta' \leq 5T\eta\left[M(R+\rho)\xi^{-1} + 4E\varepsilon(\xi\delta)^{-1}\right].$$

This finishes the proof of the iterative lemma. □

A few remarks may be in order.

1. The contribution of the second order term (that is, $\{\chi, \{\chi, H\}\}$) has been reduced to a numerical factor.

2. η'/η is bounded by a quantity proportional to T: the larger the frequency over which one averages, the better the resulting estimate.

3. The appearance of the terms on the right-hand side of $(5)_2$ is easy to understand. The second stems from the "quadratic error" and has the size of the Poisson bracket $\{\chi, f\}$ of the auxiliary Hamiltonian (or generating function) with the perturbation. The first represents a frequency shift and comes in because we always solve (14), instead of adapting the frequency, that is solve the same equation with ω_0 replaced by $\omega = \omega(p)$. The above algorithm is thus in principle worse than the usual Picard method, not to mention Newton's.

We shall now apply the iterative lemma a certain number of times, say $s = s(\varepsilon)$, so as to eliminate the resonant part of the Hamiltonian to a high order. We start from $H = H^{(0)} = h+f$, defined on $D^{(0)}$ (with $D^{(0)} \subset D(R, \rho, \sigma)$; see below) and in the decomposition (3) set

$$Z = \langle f \rangle,\ N = f - \langle f \rangle,\ \varepsilon = \varepsilon_0,\ \eta = \eta_0 = \left\| f - \langle f \rangle \right\|_D / E \leq 2\varepsilon.$$

In the end, we get a Hamiltonian $H' = H^{(s)}$, defined on $D' \subset D = D^{(0)}$ and characterized by $\varepsilon' = \varepsilon_s$ and $\eta' = \eta_s$, having gone through a sequence of intermediate quantities $H^{(j)}$, $D^{(j)}$, ε_j, η_j, $j = 0, \ldots, s$.

Because of the frequency shift, one cannot work on a domain of order 1, so we use the smaller domain

$$D = D^{(0)} = D\big(R(\varepsilon), \rho(\varepsilon), \sigma\big) \subset D(R, \rho, \sigma).$$

Any trajectory with initial condition (in action space) lying in the (real) ball of radius $r(\varepsilon)$ around the origin will stay, until time $T(\varepsilon)$, in the ball of radius $R(\varepsilon) \leqslant R$. One has $D^{(j)} = D^{(j-1)} - (\delta_j, \xi_j)$ and all the pairs (δ_j, ξ_j) are chosen to be equal for $j = 1, \ldots, s : \xi_j = \xi$, $\delta_j = \delta$; this choice implies that

$$D' = D^{(0)} - (s\delta, s\xi) = D\big(R(\varepsilon), \rho(\varepsilon) - s\delta, \sigma - s\xi\big).$$

We build a sequence of canonical transformations $C^{(j)} : D^{(j)} \to D^{(j-1)}$ using the iterative lemma and denote by C their composition, from D' into D.

The iterative lemma ensures that

$$D^{(0)} - \left(\frac{s\delta}{2}, \frac{s\xi}{2}\right) \supset C(D') \supset D^{(0)} - \left(\frac{3s\delta}{2}, \frac{3s\xi}{2}\right) = D\left(R(\varepsilon), \rho(\varepsilon) - \frac{3s\delta}{2}, \sigma - \frac{3s\xi}{2}\right).$$

There are two requirements on the domains:

1) The system is defined on $D = D^{(0)}$, that is, $D \subset D(R, \rho, \sigma)$, which implies that $R(\varepsilon) \leqslant R$ and $\rho(\varepsilon) \leqslant \rho$.

2) The image of C contains the real ball of radius $R(\varepsilon)$ centred at the origin in action space. For this to be true, it is enough that $3s\delta/2 \leqslant \rho(\varepsilon)$ and $3s\xi/2 \leqslant \sigma(\varepsilon)$, which leads to the choice

$$\delta_j = \delta = \frac{\rho(\varepsilon)}{2s}, \quad \xi_j = \xi = \frac{\sigma}{2s}.$$

Moreover, in the sequel, we choose $R(\varepsilon) = \rho(\varepsilon)$, which is no essential restriction; $r(\varepsilon) < R(\varepsilon)$ will be of the same order (with respect to ε). Finally, to simplify the notation, we shall often write r, R and ρ, without making the dependence on ε explicit. This slight ambiguity should cause no confusion as the original quantities R and ρ will not play any role and we shall rewrite the conditions $R(\varepsilon) \leqslant R$ and $\rho(\varepsilon) \leqslant \rho$ at the very end.

Rewriting formula $(5)_2$ with these values of the parameters, we get

$$\eta_j \leq \eta_{j-1} 20\sigma^{-1} T\left(M\rho s + \frac{4E\varepsilon_j s^2}{\rho}\right), \quad j = 1, \ldots, s. \tag{18}$$

This is in some sense the fundamental inequality and it reflects rather accurately the data of the problem in its dependence with respect to the various parameters. We shall exploit it in three ways, but first go on with what is common to the three variants.

Each time we perform s transformations and ask that $s \geqslant 2$ (in fact otherwise the construction is of no interest). We also require that the sequence η_j decrease at least geometrically with ratio $1/e$ ($e = 2.718\ldots$), which is of course somewhat arbitrary. From $\eta_j \leqslant \eta_0 e^{-j}$, $\eta_0 \leqslant 2\varepsilon$ and $\varepsilon_j \leqslant \varepsilon_{j-1} + (1\backslash 2)\eta_j$, one finds that $\varepsilon_j \leqslant 2\varepsilon_0 = 2\varepsilon$. Inserting this into (18) we find that $\eta_j \leqslant (1/e)\eta_{j-1}$ is guaranteed provided that

$$X \stackrel{\text{def}}{=} 20\sigma^{-1} T\left(M\rho s + \frac{8E\varepsilon s^2}{\rho}\right) \leq \frac{1}{e}. \tag{19}$$

The final remainder η' is thus estimated as

$$\eta' = \eta_s \leq \eta_0 e^{-s} \leq 2\varepsilon e^{-s}. \tag{20}$$

We now have at our disposal a "resonant normal form", in which the non-resonant harmonics of the original Hamiltonian have been eliminated to a high order s (still to be determined), with the help of a canonical transformation C such that $C(p', q') = (p, q)$.

To make use of it, we now add the ingredients of energy conservation and convexity of the unperturbed Hamiltonian. The functions $s = s(\varepsilon)$, $r(\varepsilon)$ and $R(\varepsilon)$ are still free parameters, and the reasoning below concerns the *real* parts of the various domains. We denote by $c(\varepsilon)$ the size of the canonical transformation in action space: $\|p'-p\| \leqslant c(\varepsilon)$; from the construction $c(\varepsilon) \leqslant (1/2)\rho(\varepsilon)$, but we shall need a somewhat more precise estimate.

We thus start from the initial condition $(p(0), q(0))$ with $\|p(0)\| \leqslant r(\varepsilon)$; one has

$$h(p'(t)) = h(p'(0)) + \omega(p'(0)) \cdot (p'(t) - p'(0)) + \frac{1}{2}\Big(A(p^*)(p'(t) - p'(0)) \cdot (p'(t) - p'(0))\Big),$$

where p^* is situated between $p'(t)$ and $p'(0)$ (the domain is convex). The convexity of h then implies that

$$(21) \quad \frac{1}{2}m\|p'(t) - p'(0)\| \leq \Big|h(p'(t)) - h(p'(0))\Big| + \Big|\,\omega(p'(0)) \cdot (p'(t) - p'(0))\,\Big|.$$

The first term on the right-hand side is estimated using conservation of energy for the full system: $H'(p'(t), q'(t)) = H'(p'(0), q'(0))$ implies that

$$(22) \quad \Big|h(p'(t)) - h(p'(0))\Big| \leq \Big|Z'(p'(t))\Big| + \Big|Z'(p'(0))\Big| + \Big|N'(p'(t))\Big| + \Big|N'(p'(0))\Big| \leq 2\varepsilon E + 2\varepsilon E + \eta' E + \eta' E \leq 5\varepsilon E.$$

We have used $\|Z'\| \leqslant 2\varepsilon E$ and $\eta' \leqslant (1/2)\varepsilon$ (which comes from $s \geqslant 2$).

To estimate the second term on the right-hand side of (21), one considers the projections of the vectors $\omega(p'(0))$ and $p'(t)-p'(0)$ on ω_0 and on the orthogonal complement. We denote the corresponding projection operators as Π and $\Pi^\perp$ respectively. First:

$$\Big\|\Pi^\perp \omega(p'(0))\Big\| = \Big\|\Pi^\perp(\omega(p'(0)) - \omega(0))\Big\| \leq \Big\|\omega(p'(0)) - \omega(0)\Big\| \leq M\|p'(0)\| \leq 2rM.$$

$c(\varepsilon) \leqslant r(\varepsilon)$ has been used, and that will be yet another requirement to keep in mind.

Projecting now on ω_0 we use the crucial fact that $\Pi\left(\dfrac{\partial}{\partial q}Z'\right) = 0$, because Z' is resonant. This is the only place where use is made of the normal form; to rephrase this, one can say that we have eliminated *one* degree of freedom by time averaging over the motion of period T. Because of the Hamiltonian character of the equation, this corresponds to motion orthogonal to the (unperturbed) energy surface. Convexity then provides quadratic potential wells, which prevent motion tangent to the surface. So from

$$\Pi\dot{p}' = -\Pi\Big(\frac{\partial Z'}{\partial q} + \frac{\partial N'}{\partial q}\Big) = -\Pi\frac{\partial N'}{\partial q}$$

we get

$$\Big\|\Pi(p'(t) - p'(0))\Big\| \leq |t|\Big\|\frac{\partial}{\partial q}N'\Big\| \leq \frac{2}{\sigma}|t|\|N'\| \leq \frac{2}{\sigma}T(\varepsilon)\eta' E.$$

We have applied the Cauchy inequality to *real* points of the domain D' and used the fact that the analyticity width σ' of H' is equal to $\sigma/2$ because of the choice of ξ. The time of validity $\mathcal{T}(\varepsilon)$ is still a free parameter at this point. We thus obtain

$$\left|\omega(p'(0))\cdot(p'(t)-p'(0))\right| \leq 2rM\|p'(t)-p'(0)\| + \frac{2}{\sigma}\mathcal{T}(\varepsilon)\eta' E\left\|\omega(p'(0))\right\|.$$

Since $p'(0) \in B_{2r}$, one has $\|\omega(p'(0))\| \leqslant \Omega+2rM \leqslant 2\Omega$ if $2rM \leqslant \Omega$, which is a condition on $r = r(\varepsilon)$. Writing $a = \|p'(t)-p'(0)\|$, we collect the above estimates as

$$\frac{1}{2}ma^2 \leq 5\varepsilon E + \frac{4}{\sigma}\mathcal{T}(\varepsilon)\Omega\eta' E + 2rMa. \tag{23}$$

We choose $\mathcal{T}(\varepsilon)$ such that $(4/\sigma)\mathcal{T}(\varepsilon)\Omega\eta' E \leqslant \varepsilon E$, that is, $\mathcal{T}(\varepsilon) \leqslant \varepsilon(\sigma/4\Omega)\eta'^{-1}$; since $\eta' \leqslant 2\varepsilon e^{-s}$, this is larger than

$$\mathcal{T}(\varepsilon) = \frac{\sigma}{8\Omega}e^{s(\varepsilon)}, \tag{24}$$

which is the value we finally adopt. Putting this into (23), we find that

$$a \leq 2r\frac{M}{m} + \frac{1}{m}(12m\varepsilon E + 4r^2M^2)^{\frac{1}{2}}. \tag{25}$$

One may notice that the quantity on the right-hand side is at least of the order of $\sqrt{\varepsilon}$, which could have been predicted. In fact, convexity provides *quadratic* potential wells, so that the energy increases as a^2 (the square of the distance to the bottom); adding a perturbation of order ε, both terms have to be at least of the same size to ensure confinement. This implies that the confinement radius is at least of the order of $\sqrt{\varepsilon}$, so that the second stability exponent satisfies $b \leqslant 1/2$.

We now require that the second term in the bracket be the larger: $12m\varepsilon E \leqslant 4r^2M^2$. Under this condition (25) implies that $a \leqslant 5rM/m$, so

$$\|p(t)-p(0)\| \leq 2c(\varepsilon) + \|p'(t)-p'(0)\| \leq 7r\frac{M}{m},$$

which entails $\|p(t)\| \leqslant 8rM/m$.

Before gathering everything together, we make the condition $r(\varepsilon) \geqslant c(\varepsilon)$ explicit, by estimating the latter quantity as follows:

$$c(\varepsilon) \leq \sum_{j=1}^{s}\frac{2}{\varepsilon_j}\|\chi_j\|_{D_j} \leq \sum_{j=0}^{s-1}\frac{4s}{\sigma}\frac{T}{2}\eta_j E = \frac{2ET}{\sigma}s\sum_{j=0}^{s-1}\eta_j.$$

From $\eta_j \leqslant 2\varepsilon e^{-j}$ we deduce the bound: $c(\varepsilon) \leqslant (8ET/\sigma)s\varepsilon$.

Everything we have done above is valid under the assumptions (4) of the iterative lemma. Now the first two are satisfied by construction and it is easy to see that the inequality $\xi\delta \geqslant 2TE\eta$ holds if $(5)_2$ is satisfied. This is shown by looking back at $(5)_2$ and estimating the right-hand side from below, keeping only the second term in the bracket. To compare the resulting inequalities, one uses the fact that $\eta \leqslant 2\varepsilon$ (we apply the iterative lemma with $\eta = \eta_j$, $\eta' = \eta_{j+1}$ and $\varepsilon_j \leqslant 2\varepsilon$).

We can now produce a statement from which Theorems 1A, 1B, 1C below will easily follow, by specifying the parameters in different ways.

Model statement. *In the situation above, suppose* (19) *is satisfied, so that one can perform* $s(\varepsilon)$ *steps of the algorithm. We choose* $R(\varepsilon) = \rho(\varepsilon)$ *and* $r(\varepsilon) = \frac{m}{8M}R(\varepsilon)$. *Then if the initial condition satisfies* $\|p(0)\| \leqslant r(\varepsilon)$, *one has* $\|p(t)\| \leqslant R(\varepsilon)$ *for* $|t| \leqslant T(\varepsilon) = \frac{\sigma}{8\Omega}e^{s(\varepsilon)}$, *provided that the following conditions obtain:*

i) $R(\varepsilon) = \rho(\varepsilon) \leqslant \inf(R, \rho)$; *R and* ρ *are the original quantities (see* (2)). *One simply requires that the Hamiltonian be defined over the domain one is working on.*

ii) $2Mr \leqslant \Omega$; *the frequency should not vary too much over the ball where the initial conditions are chosen.*

iii) $12m\varepsilon E \leqslant 4M^2r^2$, *that is,* $r^2 \geqslant 3\frac{m}{M^2}\varepsilon E$; *the energy of the perturbation is balanced by that arising from the quadratic wells for the "kinetic" part.*

iv) $r(\varepsilon) \geqslant c(\varepsilon)$, or $r \geqslant \frac{8ET}{\sigma}s\varepsilon$: *the size of the ball prescribed for the initial conditions is larger than that of the canonical transformation (in action space).*

v) $s \geqslant 2$: *one can perform at least two steps in the algorithm.*

Before stating Theorem 1A, we define three quantities which will be useful in the sequel; they are dimensionally correct and their occurrence, except for a numerical factor, is easy to understand; so we let

$$\lambda = 10^{-3}\frac{\sigma m}{M^2}, \quad \tau = 3\cdot 10^{-3}\frac{\sigma}{\sqrt{EM}}, \quad \mathcal{T}_0 = 4\cdot 10^{-2}\frac{\sigma}{\Omega}. \tag{26}$$

We retain, of course, the setting defined at the beginning of this section.

***Theorem* 1A.** *Let* α *be such that* $0 < \alpha \leqslant 1/3$; *assume that* $\|p(0)\| \leqslant r(\varepsilon) = \lambda\varepsilon^\alpha/T$ *with T satisfying* $1 \leqslant T \leqslant \tau\varepsilon^{-\frac{1}{2}(1-3\alpha)}$. *Then* $\|p(t)\| \leqslant R(\varepsilon) = 8M/m\cdot r(\varepsilon) \leqslant$ $\leqslant 10^{-2}\sigma/M\cdot\varepsilon^\alpha/T$ *when* $|t| \leqslant T(\varepsilon) = \mathcal{T}_0\exp(\varepsilon^{-\alpha})$, *provided that* ε *satisfies the following inequalities:*

$$\begin{aligned} &\varepsilon^\alpha \leq 100\frac{M}{\sigma}\inf(R,\rho), \quad &&\varepsilon^\alpha \leq 10^3\frac{\Omega M}{2\sigma m}, \\ &\varepsilon^\alpha \leq 4\cdot 10^{-2}\frac{m}{M}, &&\varepsilon^{\frac{1}{2}(1-3\alpha)} \leq \tau. \end{aligned} \tag{27}$$

This may not look like the most natural statement in the framework of this chapter, but its merit lies in that the time of validity is independent of the chapter of the orbit when this is short enough. This will be crucial in the next section. To prove the result, set

$$\rho = \rho_0\frac{T_0}{T}\varepsilon^\alpha, \quad s = [s_0\varepsilon^\alpha], \quad T \leq T_0\varepsilon^{-\beta}, \quad \beta = \frac{1}{2}(1-3\alpha). \tag{28}$$

ρ_0, T_0 and s_0 are to be determined; $[x]$ denotes the integer part of a real number x. The factor X in (19) becomes

$$X = \frac{20}{\sigma}\left[M\rho_0 s_0 T_0 + 8ET^2\frac{s_0^2}{\rho_0 T_0}\varepsilon^{1-3\alpha}\right].$$

The first term in the bracket is larger than the second provided that

$$M\rho_0 s_0 T_0 \geq 8E\frac{T_0 s_0^2}{\rho_0}.$$

Choose $s_0 = 1$ and $\rho_0 = 2(2E/M)^{1/2}$, which ensures equality. (19) is reduced to

$$\frac{40}{\sigma}M\rho_0 s_0 T_0 \leq \frac{1}{e}.$$

Assuming equality again, we find that

$$T_0 = \frac{1}{80e\sqrt{2}}\frac{\sigma}{\sqrt{EM}} > \tau,$$

so we may adopt the value $T_0 = \tau$ and compute $\rho(\varepsilon) = R(\varepsilon)$ and $r(\varepsilon)$. Notice that

$$R(\varepsilon) \leq \rho_0 T_0 \varepsilon^{\alpha} \text{ and } R(\varepsilon) \geq \rho_0 \varepsilon^{\alpha+\beta} = \rho_0 \varepsilon^{\frac{1}{2}(1-\alpha)} > \rho_0 \varepsilon^{\frac{1}{2}}.$$

Theorem 1A is then a consequence of the "model statement" above. Concerning the time of validity, one writes $s \geqslant s_0\varepsilon^{-\alpha}-1$, which leads to the value of T_0 defined in (26), by slightly reducing the quantity $\frac{\sigma}{8e\Omega}$, obtained from this inequality and the model statement.

The computation of the thresholds is straightforward, using the five conditions of the model statement. In the first two, one should use *upper* bounds for $r(\varepsilon)$ and $R(\varepsilon)$, that is, set $T = 1$, whereas in the third one has to use the *lower* bound for $r(\varepsilon)$. This leads to the first three inequalities in (27). Conditions iv) and v) are both weaker than iii). The last threshold ensures that the upper bound for T is indeed larger than 1. □

The next result will sound more natural in this setting; we write again

$$\rho_0 = R_0 = 2\sqrt{\frac{2E}{M}}, \quad r_0 = \frac{m}{8M}R_0 = \frac{m}{4M}\sqrt{\frac{2E}{M}}. \tag{29}$$

***Theorem* 1B.** *If* $\|p(0)\| \leqslant r_0\varepsilon^{1/3}$, *then* $\|p(t)\| \leqslant R_0\varepsilon^{1/3}$ *if* $|t| \leqslant T_0 \exp\left(\frac{\tau}{T\varepsilon^{1/3}}\right)$ *(however, if* $T \leqslant \tau$, *one should replace the factor* τ/T *by* 1) *provided that* ε *satisfies*

$$\begin{aligned} &\varepsilon^{\frac{1}{3}} \leq \frac{1}{2}\sqrt{\frac{M}{2E}}\inf(R,\rho), && \varepsilon^{\frac{1}{3}} \leq \frac{\Omega}{\sqrt{EM}},\\ &\varepsilon^{\frac{1}{3}} \leq 4\cdot 10^{-2}\frac{m}{M}, && \varepsilon^{\frac{1}{3}} \leq \frac{\tau}{2T} = 1.5\cdot 10^{-3}\frac{\sigma}{T\sqrt{EM}}. \end{aligned} \tag{30}$$

The proof is similar, and in fact simpler. One defines

$$\rho = \rho_0 \varepsilon^{\frac{1}{3}}, \quad s = [s_0 \varepsilon^{-\frac{1}{3}}];$$

inserting these values in (19) gives

$$X = \frac{20}{\sigma} T \left[M \rho_0 s_0 + \frac{8E}{\rho_0} s_0^2 \right] \leq \frac{1}{e}.$$

We assume that $s_0 \leqslant 1$ (hence the restriction in brackets in the statement), which allows one to replace s_0^2 by s_0, slightly strengthening the condition. Assuming equality, one finds that

$$s_0 = \frac{\sigma}{20eT} \left(M \rho_0 + \frac{8E}{\rho_0} \right)^{-1}.$$

Upon maximizing this expression with respect to ρ_0, one finds the value in (29) and

$$s_0 = \frac{1}{80e\sqrt{2}} \frac{\sigma}{T\sqrt{EM}} > \frac{\tau}{T}.$$

It remains to determine the thresholds of validity, which does not pose any particular problem. Again iv) is weaker than iii), and ii) has been slightly strengthened for aesthetic purposes. □

We are concerned with three parameters of physical interest, T, s and r, which are connected, respectively, with the *period* of the linear flow on the given torus, the *time of validity* of the stability estimate, and what we shall call the radius of the *influence zone* of the torus. In our last statement, we shall put the latter quantity in a privileged position and treat r as a free variable, trying to make it as large as possible. We may assume that $r \geqslant r_0 \varepsilon^{1/3}$, since otherwise Theorem 1B applies.

***Theorem* 1C.** *Let $(p(0), q(0))$ be an initial condition such that $\|p(0)\| \leqslant r$; then $\|p(t)\| \leqslant 8(M/m)r$ if $|t| \leqslant T_0 \exp[\lambda/(rT)]$ (in the case $\lambda/(rT) \geqslant \varepsilon^{-1/3}$ one should use the latter quantity), provided that the following conditions hold*:

a) $r \geqslant r_0 \varepsilon^{1/3}$, *where* $r_0 = \dfrac{m}{4M}(2E/M)^{1/2}$.

b) r *satisfies the following four inequalities*:

$$\text{(31)} \quad r \leq \frac{m}{8M} \inf(R, \rho), \; r \leq \frac{\Omega}{2M}, \; r^2 \geq \frac{3m}{M^2} \varepsilon E, \; r \leq \frac{\lambda}{2T} = 10^{-3} \frac{\sigma m}{2TM^2}.$$

Combining the last two inequalities provides a threshold for ε, and Theorems 1C and 1B connect nicely in the vicinity of $r = r_0 \varepsilon^{1/3}$. The proof is again straightforward, using the model statement. One has $\rho = \dfrac{8M}{m} r \geqslant \rho_0 \varepsilon^{1/3}$, so in (19) one finds that

$$X = \frac{20}{\sigma} T \left[M \rho s + 8E \frac{\varepsilon s^2}{\rho} \right] \leq \frac{20}{\sigma} T \left[M \rho s + 2\sqrt{2EM} (\varepsilon^{\frac{1}{3}} s)^2 \right].$$

One assumes that $s \leqslant \varepsilon^{-1/3}$ (hence the restriction in brackets in the statement) which is in fact true if for instance $\tau \leqslant T$, because $r \geqslant r_0\varepsilon^{1/3}$. Under this assumption, the first term in the bracket is the larger and one proceeds as in Theorems 1A and 1B. □

The three above results constitute the "building bricks" from which the general stability theorems over finite times will be obtained in the next section, using only simple approximation properties, without any additional work. Since these arithmetical considerations cannot be improved upon, all the non-optimal features of the results should be blamed on the above.

It remains in this chapter to indicate briefly the necessary modifications when h is only assumed to be *quasi*-convex. This is the natural geometrical assumption: the unperturbed energy surface is convex in angle-action variables (when considered in action space). It allows one in particular to include the case of periodic perturbations of convex Hamiltonians (see below) and the related situation of symplectic maps with sign-definite twist matrices (see Chapter IV, §2).

We still denote the Hessian matrix as $A(p) = \nabla^2 h(p)$ and M its largest eigenvalue on the given domain: $\|A(p)v\| \leqslant M\|v\|$ for any $v \in \mathbb{R}^n$, $p \in D$. Now $m > 0$ is defined by the inequality

$$A(p)v\cdot v \geq m\|v\|^2 \quad \text{if} \quad \omega(p)\cdot v = 0,\ v \in \mathbb{R}^n,\ p \in D\cap\mathbb{R}^n,$$

that is, v is tangent to the unperturbed energy surface ($\omega(p) = \nabla h(p)$).

For illustrative purposes, let us compute this quantity for a periodically perturbed convex Hamiltonian, that is, let $H(p, q, t) = h(p)+f(p, q, t)$, where h is convex and has associated quantities m and M when considered as a convex functional; f is assumed to be 1-periodic with respect to t. One may regard this as an autonomous problem in $n+1$ dimensions, with Hamiltonian

$$H_1(p_1, q_1) = h_1(p_1) + f_1(p_1, q_1),$$

where $p_1 = (p, e)$, $q_1 = (q, t)$, $h_1 = h(p)+e$ and $f_1(p_1, q_1) = f(p, q_1)$. The frequency is $\omega_1 = (\omega, 1)$. The Hessian matrix is singular, since we have added a constant frequency, but the system is isoenergetically non-degenerate (see the beginning of Chapter III). The function h_1 is quasi-convex, as we shortly see; let m_1 and M_1 denote the associated quantities. Obviously one has $M_1 = M$, and m_1 is computed as follows: if $v_1 = (v, w) \in \mathbb{R}^{n+1}$, $v \in \mathbb{R}^n$, $w \in \mathbb{R}$, the condition $v_1\cdot\omega_1 = 0$ reads $\omega\cdot v+w = 0$. By definition,

$$A_1 v_1\cdot v_1 = Av\cdot v \geq m\|v\|^2$$

and under the assumption $v_1\cdot\omega_1 = 0$

$$\|v_1\|^2 = \|v\|^2 + |w|^2 \leq (1+\|\omega\|^2)\|v\|^2,$$

hence

$$A_1 v_1\cdot v_1 \geq m(1+\|\omega\|^2)^{-1}\|v_1\|^2.$$

Working in a domain such that, say, $\|\omega(p)\| \leqslant 2\Omega = 2\|\omega(0)\|$, we may take

$$m_1 = m(1+4\Omega^2)^{-1}.$$

Returning to the proof of the results, the difference between convexity and quasiconvexity occurs in the geometrical reasoning only; the iterative lemma remains untouched (notice that m does not appear in it). (21) can still be written

$$\frac{1}{2}\Big(A(p^*)\big(p'(t)-p'(0)\big)\cdot\big(p'(t)-p'(0)\big)\Big) \le \Big|h\big(p'(t)\big)-h\big(p'(0)\big)\Big| + \Big|\,\omega\big(p'(0)\big)\cdot\big(p'(t)-p'(0)\big)\Big|,$$

where p^* lies somewhere between $p'(t)$ and $p'(0)$.

The right-hand side is estimated as above by the right-hand side of (23). Let $\omega^* = \omega(p^*)$, $A^* = A(p^*)$, $u = p'(t)-p'(0)$, $\|u\| = a$. Choosing $\mathcal{T}(\varepsilon)$ again as in (24), one thus finds that

$$\frac{1}{2}(A^*u,u) \le 6\varepsilon E + 2rMa, \tag{32}$$

and it only remains to estimate the left-hand side from below. To this end, let Π^* be the orthogonal-projection operator on ω^*, and $\Pi^{*\perp}$ the projection operator on the orthogonal complement; we write the explicit decomposition

$$A^*u\cdot u = A^*\Pi^*u\cdot\Pi^*u + A^*\Pi^{*\perp}u\cdot\Pi^{*\perp}u + 2\,A^*\Pi^*u\cdot\Pi^{*\perp}u\,.$$

Then

$$\frac{1}{2}A^*\Pi^*u\cdot\Pi^{*\perp}u \ge \frac{1}{2}m\|\Pi^{*\perp}u\|^2 = \frac{1}{2}m\big(a^2-\|\Pi^*u\|^2\big).$$

Hence

$$\begin{aligned}\frac{1}{2}A^*u\cdot u &\ge \frac{1}{2}ma^2 - \frac{1}{2}m\|\Pi^*u\|^2 - \frac{1}{2}M\|\Pi^*u\|^2 - Ma\|\Pi^*u\| \\ &\ge \frac{1}{2}ma^2 - 2Ma\|\Pi^*u\|,\end{aligned}$$

which yields

$$\frac{1}{2}ma^2 \le 6\varepsilon E + 2rMa + 2Ma\|\Pi^*u\|. \tag{33}$$

With Π still denoting the projection on $\omega_0 = \omega(0)$ and $\Omega = \|\omega_0\|$, one has

$$\|\Pi^*u\| = \|\Pi u\| + \big\|(\Pi-\Pi^*)u\big\|.$$

We then use

$$\|\Pi u\| \le \frac{2}{\sigma}\mathcal{T}(\varepsilon)\eta' E \le \frac{\varepsilon E}{2\Omega},$$

which still holds. From $\|p^*\| \leqslant 2r+a$, one concludes that

$$\|(\Pi - \Pi^*)u\| \leq 2a(2r+a)\frac{M}{\Omega},$$

and in the end

$$\|\Pi^* u\| \leq \varepsilon\frac{E}{2\Omega} + 2a(2r+a)\frac{M}{\Omega}. \tag{34}$$

We now insist on obtaining the *same* radius of confinement $R(\varepsilon)$, in order not to have to alter the domains in the iterative lemma. A simple way to achieve this is to require that $\|\Pi^* u\| \leqslant r$. In view of (33), one may then leave R unchanged, replacing r by $r/2$. We therefore define

$$R(\varepsilon) = \rho(\varepsilon) \text{ and } r(\varepsilon) = \frac{m}{16M}R(\varepsilon) \quad \left(\text{instead of } \frac{m}{8M}R(\varepsilon)\right).$$

We need $\|\Pi^* u\| \leqslant r$, knowing that $a \leqslant R$. This amounts to a simple bootstrap argument: (34) yields a condition on R (or r), which is satisfied in particular if

$$r \geq \varepsilon\frac{E}{\Omega} \text{ and } r \leq 10^{-3}\frac{m^2\Omega}{M^3}.$$

Model statement (version for quasi-convex Hamiltonians). *It differs from the version for convex Hamiltonians only in the following points:*

1) *One still defines* $R(\varepsilon) = \rho(\varepsilon)$, *but now* $r(\varepsilon) = \dfrac{m}{16M}R(\varepsilon)$.

2) *Condition* ii) *is replaced by*

$$\text{ii bis)} \quad r \leq 10^{-3}\frac{m^2\Omega}{M^3};$$

which is stronger.

3) *One adds the condition* $r \geqslant \varepsilon\dfrac{E}{\Omega}$, *which for small* ε *is weaker than* iii).

We leave it to the interested reader to modify Theorems 1A, 1B, 1C accordingly. The modifications are of minor significance.

Chapter III

STABILITY FOR ARBITRARY INITIAL CONDITIONS

In this chapter, we use Theorems 1A, 1B, 1C, especially Theorem 1A, to obtain information about stability of points in phase space. The general idea is to apply one of these results whenever a given point lies in the influence zone of some periodic torus. This amounts to studying the distribution of rational points in frequency space, which corresponds to that of unperturbed periodic tori, provided that the frequency map $p \to \omega(p)$ enjoys some non-degeneracy condition. We mention that this is really all that is needed for the approximation process; in particular, analyticity and quasi-convexity are used only inasmuch as the theorems of Chapter II are applied.

To make all this precise, let us consider again the Hamiltonian (1) of Chapter II; the only difference here is that we look at the neighbourhood of some arbitrary point (or rather torus) $p = p^*$, so that in (2) (Chapter II) one should use a ball with centre at p^*.

If the unperturbed Hamiltonian h is convex, the Hessian matrix $A(p) = \nabla^2 h(p)$ is non-singular ($\det A(p) \neq 0$), and the frequency map is a local diffeomorphism. Simultaneous approximation will however deal rather with the ratios of the frequencies to one of them, which corresponds to isoenergetic non-degeneracy. For the sake of completeness, we briefly recall the definition and show that it is satisfied by quasi-convex (in particular convex) Hamiltonians.

Let Σ be the unperturbed energy surface $h(p) = h(p^*)$; one wants the $n-1$ ratios of the frequencies to a given one to yield a local chart of Σ. This is the same as requiring that the map

$$p \in \Sigma \to \omega \in \mathbb{P}\mathbb{R}^{n-1}$$

be a local diffeomorphism near p^*, where the frequency is considered in projective space. To check this condition, one should make sure that the Hessian matrix of the "homogeneous" map

$$(p, \lambda) \in \mathbb{R}^n \times \mathbb{R} \to \lambda h(p)$$

is non-singular at $(p^*, 1)$. The matrix reads

$$\mathcal{A}(p^*) = \begin{pmatrix} A & \omega \\ \omega & 0 \end{pmatrix},$$

where $\omega = \omega(p^*)$ is written as a column on the right and as a row at the bottom. The isoenergetic non-degeneracy condition thus reads $\det \mathcal{A} \neq 0$ (this is sometimes called "the Arnol'd determinant"). Suppose now that h is quasi-convex, and let $u \in \mathbb{R}^{n+1}$ with $\mathcal{A}u = 0$. We write $u = (v, w)$, $v \in \mathbb{R}^n$, $w \in \mathbb{R}$. The condition $\mathcal{A}u = 0$ splits into

$$Av + w\omega = 0 \quad \text{and} \quad (\omega, v) = 0.$$

Taking the dot product of the first equality with v, we find that

$$(Av, v) = 0 \quad \text{and} \quad (\omega, v) = 0,$$

which implies that $v = 0$ by the very definition of quasi-convexity, and then $w = 0$, so $u = 0$.

Quasi-convex Hamiltonians are thus isoenergetically non-degenerate. We recall that in the opposite direction, and for low dimensions, one has the following simple results: when $n = 2$, isoenergetic non-degeneracy is equivalent to quasi-convexity; when $n = 3$ quasi-convexity is equivalent to the condition $\det \mathcal{A} < 0$, so that, so to speak, "half" the non-degenerate Hamiltonians are quasi-convex. We also recall, as a word of caution, that outside the realm of convexity and quasi-convexity, non-degeneracy and isoenergetic non-degeneracy are independent conditions; neither of them follows from the other.

Below we shall again, for simplicity, treat the case of convex Hamiltonians; the modifications needed in the quasi-convex case are of minor interest. Geometrically speaking, in both cases the locus, in action space, where $\omega \in \mathbb{P}\mathbb{R}^{n-1}$ is constant, is a smooth curve which intersects the unperturbed energy surface transversally. So on this curve the n-vector ω varies along a straight line. Now in the convex case it does indeed vary and the frequency itself constitutes a local parameter; in the *quasi*-convex case, however, the frequency may well be constant along the curve (think of the periodic perturbation of a convex Hamiltonian; compare the end of the previous chapter).

So let h be convex, p^* a hitherto arbitrary point, and $\omega^* = \omega(p^*)$. Of course we use the notation of Chapter II. By the usual implicit function theorem, one has the following. Let $B(p^*)$ be a ball centred at p^* with radius S, such that for $p \in B(p^*)$

$$\|A(p) - A(p^*)\| \leq \frac{m}{2}.$$

Here $\|\cdot\|$ denotes the usual operator norm associated with the Euclidean norm. One can take $S = \dfrac{m}{2|h|_3}$, where $|h|_3$ is an upper bound of the third derivative of h. Then the frequency map is one-to-one on $B(p^*)$, and $\omega(B(p^*)) \supset B(\omega^*)$, a ball with centre at ω^* and radius $(m/2)S$.

The above determines in a quantitative way the local inversion properties of the frequency map. Because of this, the procedure to determine the domain is as follows. One starts from a fixed ball $B_0(p^*)$ over which $H = h + f$ is defined with analyticity widths ρ and σ. One then determines m, M and $|h|_3$ on $B_0(p^*)$, and then restricts oneself to a ball of radius S, which may be assumed to be included in $B_0(p^*)$, decreasing m if necessary.

Let us now get closer to the heart of the matter, which will necessitate one more piece of notation. For real x one has

$$x = [x] + \{x\},$$

with $[x] \in \mathbb{Z}$ the integer part, and $\{x\} \in (0, 1)$. We use the notation

$$\|x\|_{\mathbb{Z}} \stackrel{\text{def}}{=} \inf(\{x\}, 1 - \{x\}) = \operatorname{dist}(x, \mathbb{Z}).$$

Although $\|\cdot\|_{\mathbb{Z}}$ is not a norm, this is a commonly used notation, even without the index $\mathbb{Z}$, which we have added to avoid confusion. If now $x \in \mathbb{R}^n$ with components $x^{(j)}$, one sets

$$\|x\|_{\mathbb{Z}} \stackrel{\text{def}}{=} \sup_{j=1,\ldots,n} \|x^{(j)}\|_{\mathbb{Z}} = \inf_{\zeta \in \mathbb{Z}^n} \|x - \zeta\|_\infty,$$

where $\|\cdot\|_\infty$ is the norm of the largest component, which naturally occurs in approximation theory. In particular,

$$\|x\|_{\mathbb{Z}} \leq \operatorname{dist}(x, \mathbb{Z}^n) \leq \sqrt{n}\|x\|_{\mathbb{Z}}.$$

With this notation, one has the following result.

Theorem (Dirichlet, see, for example, [11] or [52]). *Let $\alpha \in \mathbb{R}^n$ and Q a real number, $Q > 1$. There exists an integer q, $1 \leqslant q < Q$, such that*

$$\|q\alpha\|_{\mathbb{Z}} \leq Q^{-\frac{1}{n}}.$$

We shall apply this basic result of approximation theory with $\alpha = \omega^* = \omega(p^*)$ and $q = T$ playing the role of a period. Before we do this, however, it is important to notice that one can gain one dimension, which bears directly on the value of the stability exponents. This reflects the fact that simultaneous approximation corresponds to *inhomogeneous* linear approximation (see Appendix 1) or, very concretely, that q in the above theorem is an integer, so that approximating α is equivalent to approximating $(1, \alpha) \in \mathbb{R}^{n+1}$. One way to put this to use is precisely to consider that $\alpha \in \mathbb{P}\mathbb{R}^{n-1}$, that is, one should in fact approximate the ratios of the components to a fixed one. In the framework we are interested in, we may in fact simply rescale one of the components of ω^* to unity. To this effect, let $w > 0$ denote the modulus of a non-zero component of ω^*, for example, but not necessarily the largest one: $w = \|\omega^*\|_\infty$. One introduces the scaling

$$(1)\quad t' = wt,\ H' = \frac{H}{w},\ \omega^{*\prime} = \frac{\omega^*}{w},\ m' = \frac{m}{w},\ M' = \frac{M}{w},\ E' = \frac{E}{w},\ \varepsilon' = \varepsilon.$$

Relabelling if necessary, we are reduced to the case when the first component of the frequency is equal to unity. Below, for the sake of clarity, we write everything using the original quantities and at the very end remember that one should *first* perform the transformations (1) and change the results accordingly.

Let us now apply Dirichlet's theorem with $\alpha = \omega^*$; for any $Q > 1$ there exists an integer T, $1 \leqslant T < Q$, and $\zeta \in \mathbb{Z}^n$, such that

$$\|T\omega^* - \zeta\| \leq \sqrt{n}Q^{-\frac{1}{n}} \text{ (Euclidean norm).}$$

Thus, $\omega = T^{-1}\zeta$ is a rational vector of period T, satisfying

$$(2)\qquad \|\omega - \omega^*\| \leq \frac{\sqrt{n}}{TQ^{1/n}}.$$

We assume that ω is close enough to ω^* so that the frequency map can be inverted. From what was recalled above, to secure this it is enough to require that

$$\frac{\sqrt{n}}{Q^{1/n}} \leq |h|_3^{-1}\frac{m^2}{4},$$

where T has been removed from the left-hand side because it is $\geqslant 1$. Under this condition, there exists a point p such that $\omega = \omega(p)$ and

$$\|p - p^*\| \leq \frac{\sqrt{n}}{m} \frac{1}{TQ^{1/n}}, \tag{3}$$

where the factor $1/m$ estimates from above the norm of the inverse of the frequency map.

We wish to apply Theorem 1A *around the point p*, which is rational with period T, and we want p^* to lie in the influence zone of p, which will be the case if

$$r(\varepsilon) = \frac{\lambda \varepsilon^{\alpha}}{T} \geq \|p - p^*\|.$$

It follows from (3) that this is in turn guaranteed if we choose

$$Q^{\frac{1}{n}} = \frac{\sqrt{n}}{\lambda m} \varepsilon^{-\alpha}, \tag{4}$$

which *defines* the value of Q. This is where it is crucial to apply Theorem 1A, with its time of stability independent of the period and a radius of the influence zone inversely proportional to it. The latter feature yields (4), in which T does not appear. We shall point out below what can be inferred if one tries to use Theorems 1B and 1C.

To apply Theorem 1A, there remains yet another important condition to be satisfied: the period should not be too long. More precisely, since $T < Q$, it is enough to require that

$$Q \leq \tau \varepsilon^{-\frac{1}{2}(1-3\alpha)},$$

that is, referring to (4),

$$\left(\frac{\sqrt{n}}{\lambda m}\right)^n \varepsilon^{-n\alpha} \leq \tau \varepsilon^{-\frac{1}{2}(1-3\alpha)},$$

or

$$\varepsilon^{\frac{1-3\alpha}{2} - n\alpha} \leq \tau \left(\frac{\lambda m}{\sqrt{n}}\right)^n. \tag{5}$$

This simple reasoning is very important, because in fact it unveils the meaning of the first stability exponent, the more important one. The inequality (5) defines a threshold for ε provided that $\frac{1}{2}(1-3\alpha) - n\alpha > 0$, or

$$\alpha < \frac{1}{2n+3}.$$

The value on the right-hand side is thus not accessible, but any smaller value is; at this point one should remember that we shall then substitute $n-1$ for n, and the stability exponent $a(n)$ will be given by $a(n) = \alpha(n-1)$. For the time being, we go on with the initial quantities and write a statement with

$\alpha = \dfrac{1}{2n+4}$, which satisfies $\frac{1}{2}(1-3\alpha)-n\alpha = \frac{1}{2}\alpha$. This allows us to rewrite (5) as

(6) $$\varepsilon^{\alpha} \leq r^{2}\left(\frac{\lambda m}{\sqrt{n}}\right)^{2n}.$$

The radius of stability can be easily computed since Theorem 1A specifies the distance from p and $\|p-p^*\| \leqslant r(\varepsilon)$. In fact, returning to the text immediately above the "model statement" of Chapter II, one notices that we have proved slightly more than what was actually included in the statement itself; in fact, it was shown that

$$\|p(t) - p(0)\| \leq 7r\frac{M}{m} < R(\varepsilon).$$

We use this estimate to get

$$\|p(t) - p(0)\| < R(\varepsilon) < 10^{-2}\frac{\sigma}{M}\frac{\varepsilon^{\alpha}}{T} \leq 10^{-2}\frac{\sigma}{M}\varepsilon^{\alpha}.$$

To compute the threshold of validity, one should essentially copy inequalities (27) of Chapter II, add the invertibility condition and, most important, inequality (5), in the form (6) because of our choice of α. Let us quickly go into some details.

We leave the first of (27) unchanged: H is defined and analytic over the domain $D = D(R, \rho, \sigma)$ around p^*, and then one should possibly restrict the domain because of the invertibility condition, as explained above. Applying Theorem 1A around p (and not p^*) does not change anything.

In the second of (27), however, one should beware of the fact that Ω refers to p ($\Omega = \|\omega(p)\|$), and not p^* ($\Omega^* = \|\omega^*\|$). In order to express everything with parameters centred at p^*, we may add for example the following condition:

$$|\Omega - \Omega^*| \leq \frac{1}{2}\Omega^*,$$

which holds in particular if

$$M\|p - p^*\| \leq Mr \leq \frac{1}{2}\Omega^*.$$

But this is precisely equivalent to the second inequality in (27) (see the "model statement", condition ii) in Chapter II), with ω replaced by Ω^*. In short, to take care of these details, it is enough to substitute Ω^* for Ω in the second of (27). Then in the definition of T_0 (compare (26) in Chapter II) one replaces Ω by $\frac{3}{2}\Omega^*$.

The third inequality in (27) remains unaltered. As regards the fourth, one notices that the exponent $\frac{1}{2}(1-3\alpha)$ is larger than 1/5 (or even 1/4 if $n \geqslant 3$); here we have already taken into account the substitution $n \to n-1$, to be effected after the transformation (1). One then adds the invertibility condition together with (6), with $n-1$ instead of n. Summarising, we have proved the following statement.

Theorem 2. *For any initial point* $(p(0), q(0))$ $(p(0) = p^*)$, *the trajectory* $(p(t), q(t))$ *starting at* $(p(0), q(0))$ *satisfies*

$$\|p(t) - p(0)\| \leq 10^{-2}\frac{\sigma}{M}\varepsilon^{a} \quad \text{if} \quad |t| \leq T(\varepsilon) = T_0^* \exp(\varepsilon^{-a}),$$

where $a = a(n) = \dfrac{1}{2(n+1)}$, $\Omega^* = \|\omega(p(0))\|$, $T_0^* = 3 \cdot 10^{-2}\dfrac{\sigma}{\Omega^*}$.

This holds provided that ε *satisfies the following inequalities*:

$$\begin{aligned} &\varepsilon^a \leq 100\frac{M}{\sigma}\inf(R, \rho), \quad && \varepsilon^a \leq 200\frac{M\Omega^*}{\sigma m}, \quad \varepsilon^a \leq 4 \cdot 10^{-2}\frac{m}{M}, \\ &\varepsilon \leq \tau^5, && \varepsilon^a \leq 200\frac{M^2}{\sigma|h|_3}, \quad \varepsilon^a \leq \tau^2\left(\frac{\lambda m}{\sqrt{n-1}}\right)^{2(n-1)}, \end{aligned} \tag{7}$$

where λ *and* τ *are defined in formula* (26) *of Chapter* II, *and* $|h|_3$ *is the maximum of the third derivative of* h *over the domain* D.

In this statement, all the parameters connected with the Hamiltonian, along with time t, *are those which are obtained after the rescalings* (1) *have been performed, that is, one should use the primed quantities in* (1).

Although the threshold conditions seem to proliferate somewhat dangerously, only the last one is really significant. In particular, all but this last one read exactly the same whatever the value of a, inside the interval $\left(0, \dfrac{1}{2n+1}\right)$. On the other hand, the last condition is essentially a rewriting of (5), which is a direct consequence of the Dirichlet estimate. In short, the value of the first exponent, which governs the time of stability, is a very direct descendant of the exponent $1/n$ which appears in Dirichlet's theorem. In particular, when the number of degrees of freedom increases, results deteriorate, not because of the invasion of the phase space by the resonance surfaces, but because of the relative scarcity of rational points, that is, periodic tori. We shall see below how this new point of view may be exploited further.

Returning to the last condition, we notice that this is also the only place in which n appears explicitly. The factor $\sqrt{n}$ (or $\sqrt{n-1}$) is simply the length of the diagonal of the unit cube, and occurs because one uses Euclidean norms, whereas the sup norm is more natural when dealing with approximation theory; this is of little importance. Apart from this, the last threshold is very sensitive to the value of α (or a), and it vanishes when $a = 1/(2n+1)$. If one chooses for instance $\alpha = 1/(2n+5)$, which satisfies $\frac{1}{2}(1-3\alpha) - n\alpha = \alpha$, one comes up with the condition

$$\varepsilon^a \leq \tau\left(\frac{\lambda m}{\sqrt{n-1}}\right)^{n-1}, \quad a = \frac{1}{2n+3},$$

which is much weaker than the last inequality in (7).

It is important to notice that the lack of optimality of Theorem 2 is exactly the same as that of Theorem 1A. Indeed, the only new ingredient we have added is Dirichlet's theorem, which is optimal (at least as far as the quantities we are interested in are concerned).

It is in our opinion quite remarkable that essentially the best possible perturbation result over finite times may be obtained using the most basic approximation result, but various important refinements and improvements of Theorem 2 are easily derived for certain classes of initial conditions. We devote the end of this chapter to some of them, using freely the notions and notations of Appendix 1.

Placing arithmetical conditions on the frequency is essentially equivalent to ensuring some comparison between q (or T) and Q in Dirichlet's theorem. In particular, the following statement is true.

***Corollary* 1.** *Assume that after the rescalings* (1) *one has* $\omega^* = \omega(p^*) = (1, \omega')$, *with* $\omega' \in \Omega_{n-1}(\delta, \gamma)$, ($\gamma$, $\delta > 0$; *we avoid the letter* τ, *which has been used already in this context*). *Then in Theorem* 2 *one may replace the radius of confinement by*

$$\|p(t) - p^*\| \leq 10^{-2}\frac{\sigma}{M}\frac{\varepsilon^a}{T} \leq 10^{-2}\frac{\sigma}{M}\left(\frac{\lambda m}{\sqrt{n-1}}\right)^{\frac{n-1}{1+\delta}}\frac{\varepsilon^b}{\gamma},$$

where $b = a\dfrac{n+\delta}{1+\delta}, a = \dfrac{1}{2(n+1)}$.

The time of stability and the threshold conditions remain the same.

In particular, almost all points in phase space admit, for any $\eta > 0$, *the stability exponents*

$$(a, b) = \left(\frac{1}{2n+1} - \eta, \frac{n}{2n+1} - \eta\right).$$

The proof is straightforward; we have in fact already written the first inequality on the norm $\|p(t)-p^*\|$. Now, when $\omega^* \in \Omega_n(\delta, \gamma)$, T may be estimated from below, since

$$Q^{-\frac{1}{n}} \geq \|T\omega^*\|_{\mathbb{Z}} \geq \left(\frac{\gamma}{T}\right)^{\frac{1}{n}(1+\delta)}.$$

Q is given by (4) and one only needs to substitute it; of course, (1) is used first in order to rescale one of the components. The last assertion comes from the proposition in Appendix 1, together with the fact that in Theorem 2 one may replace $1/(2n+2)$ by $\dfrac{1}{2n+1}-\eta$ for any $\eta > 0$ (and vanishing threshold when η goes to zero). $\square$

Note that the pair of exponents comes very close to the would-be optimal pair $\left(\dfrac{1}{2n}, \dfrac{1}{2}\right)$. The assertion may be slightly misleading, because although almost every point belongs to $\Omega_n(\delta)$ for any $\delta > 0$, it is also the case that for almost every point the corresponding constant γ approaches 0 together with δ.

One may devise a result of a slightly different nature; fix $\delta > 0$ and consider $\gamma = \gamma(\varepsilon)$, going to infinity as ε goes to 0, for example, $\gamma = \gamma_0\varepsilon^{-\xi}$, $\gamma_0 > 0$, $0 < \xi < b = b(n, \delta)$. One then gets the same result, over the set (after rescaling) $\Omega_{n-1}(\delta, \gamma(\varepsilon))$, whose relative measure goes to 1 as ε goes to 0, with a second exponent $b(n, \delta)-\xi$, and a *fixed* value γ_0.

Suppose now that we wanted to apply Theorem 1B or 1C in order to find results for general initial conditions. We would come across a kind of intermittency phenomenon, which is perhaps worth noticing. Let $(T_i)_{i\geqslant 0}$ be the sequence of the periods of ω^*, and $(\omega_i)_{i\geqslant 0}$ the related best approximations. The rational vectors ω_i converge to ω^*, so one has $T_i\omega_i \in \mathbb{Z}^n$, and the estimate

$$\|\omega_i - \omega^*\| \leq \frac{\sqrt{n}}{T_i T_{i+1}^{1/n}}.$$

For i large enough, let us define the corresponding points p_i, converging to p^* $(\omega(p_i) = \omega_i)$, and let

$$r_i = \|p_i - p^*\| \leq \frac{\sqrt{n}}{m}\frac{1}{T_i T_{i+1}^{1/n}}. \tag{8}$$

Finally, we define the sequence of values ε_i satisfying $r_i = r_0\varepsilon_i^{1/3}$, where r_0 still denotes the constant in formula (29) of Chapter II. With these definitions and the same setting as above, the following statement holds.

Let $\varepsilon > 0$, $\varepsilon_{i-1} \geqslant \varepsilon > \varepsilon_i$, i large enough; then

$$\|p(t) - p^*\| \leq \frac{8M}{m} r_{i-1}$$

if $|t| \leqslant \mathcal{T}_i^* = \mathcal{T}_0^* \exp(\mu T_i^{1/(n-1)})$, with $\mathcal{T}_0^* = 3\cdot 10^{-2}\dfrac{\sigma}{\Omega^*}, \mu = 10^{-2}\dfrac{\sigma}{\sqrt{n-1}}\left(\dfrac{m}{M}\right)^2.$

As usual, the parameters relate to the situation after transformation (1).

This holds for any initial condition $(p^* = p(0), q(0))$, but we do not state this rather unnatural assertion as a "theorem", nor bother to mention the thresholds, which could be easily computed. The proof is again very short; just apply Theorem 1C around the point p_{i-1}, which is valid because $r_{i-1} \geqslant r_0\varepsilon^{1/3}$, by the definition of ε and of the sequence (ε_i). Next, to estimate the exponent $\lambda/(r_{i-1}T_{i-1})$ from below, use (8):

$$\frac{\lambda}{r_{i-1}T_{i-1}} \geq \frac{\lambda m}{\sqrt{n}} T_i^{\frac{1}{n}}.$$

Finally, notice that the factor $\lambda m/\sqrt{n}$ is slightly larger than μ, after the substitution of $n-1$ for n. □

Both the radius of confinement and the time of stability remain constant when ε belongs to an interval $(\varepsilon_i, \varepsilon_{i-1})$. In fact, the most favourable situation occurs when ε is equal to one of the ε_i's. This is also apparent if one applies Theorem 1B: if $\varepsilon = \varepsilon_{i-1}$, it may be applied around the point p_{i-1}; but as soon as ε crosses this value, p^* leaves the influence zone of p_{i-1} and one must use the point p_i instead. This causes the time of validity to drop discontinuously from a quantity proportional to $\exp(\tau/(T_{i-1}\varepsilon^{1/3}))$ to one proportional to $\exp(\tau/(T_i\varepsilon^{1/3}))$.

We notice that the quantities T_i, ω_i, p_i, r_i have an obvious intrinsic meaning, but the sequence (ε_i) is rather artificial, again partially because of the exponent 1/3 instead of 1/2 in Theorem 1B. What is really important is that the distribution of the rational vectors around a given one may be quite erratic (for example, nothing can be said in general about the sequences T_i/T_{i+1} or r_i/r_{i+1}) and there arises a sequence of values for which the closed orbit approximation is relatively best possible.

With this in mind, it is not surprising that one can prove statements of the same type as Corollary 1, using Theorem 1B and the Diophantine sets $\Omega(\tau, \gamma)$; in fact, by the very definition of these sets (see Appendix 1), it is then possible to estimate, for example, the ratios T_i/T_{i+1} from below, which enables one to derive results that are valid for all sufficiently small perturbations over a set of large measure (or even almost everywhere). Recalling from Appendix 1 the inclusions

$$\Omega_n(\tau,\gamma) \subset \Omega(\tau,\gamma^{-(1+\tau)}),$$

one may then compare the statement obtained from Theorem 1B with Corollary 1, over sets of type $\Omega_n(\tau, \gamma)$, and make sure that they are essentially equivalent. We shall not go into the easy details.

As a final remark on this topic, we note that it may sound somewhat paradoxical to introduce Diophantine conditions while studying the behaviour of a system over *finite* times, because these arithmetical conditions are essentially of asymptotic nature. In fact, there are two additional flexibilities which we have not used.

1. We are interested in phenomena which occur over exponentially long times; on the other hand, the sequence of the periods (T_i) of *any* vector increases at least geometrically (see Appendix 1). Therefore we could restrict attention to indices i such that $i = O(\varepsilon^{-c})$ for some $c > 0$; this is the "simultaneous" analogue of the ultraviolet cut-off, and it is naturally interpreted in terms of approximate recurrence times (again see Appendix 1).

2. There is an additional freedom related to the initial condition. Suppose that we divide the radius $r(\varepsilon)$ of the influence zone into—say—two equal parts; it is then sufficient to find a point whose frequency has nice arithmetical properties and which lies within $r(\varepsilon)/2$ of the given initial condition p^*.

It may be that these two remarks can be combined to show that the estimate of the radius of confinement which appears in Corollary 1 holds in fact for *any* point in phase space, so the second stability exponent would indeed always be close to 1/2. In any case, the second remark will be put to use below, to derive Corollary 3.

Let us now pass to the interpretation of the resonances, and to what happens when the initial condition is resonant or even only *nearly* resonant. Loosely speaking, the important property that emerges is that the more resonant the initial condition, the more stable the corresponding trajectory will be. In fact it should be remembered that the more resonant points correspond precisely to rational vectors, that is, unperturbed closed orbits, and that we have already noticed that Theorems 1A, 1B, 1C are essentially independent of the number of dimensions. We strongly emphasize that this stabilization *through* resonance is very specific of quasi-convex systems and cannot possibly hold for generic steep Hamiltonians.

Let us first recall some usual notions. Let $\mathcal{M}$ be a submodule (or sublattice of $\mathbb{Z}^n$ of rank (dimension) r, generated over $\mathbb{Z}$ by the linearly independent vectors $k_1, \ldots, k_r$ of $\mathbb{Z}^n$. A vector $\omega \in \mathbb{R}^n$ is said to be resonant with *multiplicity* r and *associated module* $\mathcal{M}$ (we write $\mathcal{M}$-resonant) if $\omega \cdot k = 0$ for any $k \in \mathcal{M}$, which is of course equivalent to $\omega \cdot k = 0$, $i = 1, \ldots, r$. With $\mathcal{M}$ we also associate the corresponding *resonant surface* $\Sigma_{\mathcal{M}}$, consisting of the points p in action space whose corresponding frequency $\omega(p)$ is $\mathcal{M}$-resonant:

$$\Sigma_{\mathcal{M}} = \left\{p \in \mathbb{R}^n,\ \left(\omega(p), k_i\right) = 0,\ i = 1, \ldots, r\right\}.$$

Since the frequency map is a local diffeomorphism, $\Sigma_{\mathcal{M}}$ is a smooth manifold of dimension $d = n-r$. In this classical framework, we prove the following result.

***Corollary* 2.** *Let $\mathcal{M}$ a submodule of $\mathbb{Z}^n$ of rank r, $0 \leqslant r \leqslant n-1$, $p^* \in \mathbb{R}^n$, $\omega(p^*) = \omega^*$. Assume that ω^* is $\mathcal{M}$-resonant, that is, $p^* \in \Sigma_{\mathcal{M}}$. Then there exist positive constants $c(\mathcal{M})$ and $c'(\mathcal{M})$ (to be constructively defined in the proof) such that Theorem* 2 *and Corollary* 1 *remain valid at the points $(p(0) = p^*, q(0))$ ($q(0)$ arbitrary) with the following changes:*

a) *one replaces everywhere n by $d = n-r$ and $\sqrt{n-1}$ by $c(\mathcal{M})\sqrt{d-1}$.*

b) *in the transformation* (1) *one replaces $w = \|\omega^*\|_\infty$ by $c'(\mathcal{M})w$.*

c) *finally, in Corollary* 1, *the expression "almost all points" should be interpreted as "almost everywhere on the resonant surface $\Sigma_{\mathcal{M}}$", equipped with the natural superficial measure.*

The basic idea is that resonant surfaces should be viewed as loci which contain abnormally many rational vectors. Thus, Corollary 2 will follow almost immediately from the next lemma.

***Lemma* 3.** *Let $\mathcal{M}$ a submodule of $\mathbb{Z}^n$ of rank r, $0 \leqslant r \leqslant n$, and let $\alpha \in \mathbb{R}^n$, be an $\mathcal{M}$-resonant vector. Writing $d = n-r$, there exists $c(\mathcal{M})$ such that if $Q > 1$ is real one can find an integer q satisfying $1 \leqslant q < Q$ and such that*

$$\|q\alpha\|_{\mathbb{Z}} \leq c(\mathcal{M})Q^{-\frac{1}{d}}.$$

$c(\mathcal{M})$ is defined as the smallest constant with the above property and the integer $c(\mathcal{M})/c(\{0\})$ will be called the order of the resonance (or submodule).

The assertion is obvious when $\mathcal{M}$ defines the "standard" resonance, that is, when α has the form $\alpha = (0, \alpha')$, where $0 \in \mathbb{R}^r$ is the zero vector and $\alpha' \in \mathbb{R}^d$. The proof of the lemma is then nothing but an exercise in linear algebra, by means of which we can reduce everything to this case, but we shall go into some details, for the sake of completeness. Before this, we note that Dirichlet's theorem asserts that $c(\{0\}) \leqslant 1$, but that equality does *not* hold (see Appendix 1), which is the reason why we defined the order as above; since however $c(\{0\})$ is close to 1 and the difference is completely irrelevant for our purpose, we shall occasionally indulge in calling $c(\mathcal{M})$ itself the order of the resonance.

Let $K = (k_i^{(j)})$ be an $r \times n$ matrix whose rows are vectors k_i which generate $\mathcal{M}$ over $\mathbb{Z}$: $k_i = k_i^{(j)}$, $j = 1, \dots, n$. Since K has integer entries, a classical result from linear algebra asserts that it may be written $K = B\Delta A$, with $B \in Gl_r(\mathbb{Z})$ and $A \in Gl_n(\mathbb{Z})$ *invertible* square matrices; Δ has the form $\Delta = [D \mid 0_d]$. Here 0_d is the zero matrix of order $d = n-r$ and D is diagonal: $D = \mathrm{diag}(d_1, \dots, d_r)$; moreover, d_j is a multiple of d_i for $i \leqslant j$. The positive integers d_i are often called the invariants of $\mathcal{M}$. We say that the module is *primitive* when they are all equal to unity, which is the same as requiring that $d_r = 1$ or else that the determinants of all the $r \times r$ submatrices of K be mutually prime. One has then $D = \mathbb{1}_r$, and we write $\Delta = \Pi$, because this is a projection operator. Any module is contained in a unique primitive one (obtained by replacing the original Δ by Π) which defines the same resonance, so that one may restrict attention to primitive modules. This stems from the obvious equivalences

$$\alpha \ \mathcal{M}\text{-resonant} \iff K\alpha = 0 \iff \Delta A\alpha = 0 \iff \Pi A\alpha = 0.$$

So let $\mathcal{M}$ be primitive and denote by (e_i), $i = 1, \dots, n$, the standard basis of $\mathbb{Z}^n$. The $\mathcal{M}$-resonant vectors are generated over $\mathbb{R}$ by the vectors $A^{-1}e_i$, $i = r+1, \dots, n$. Let $\|u\|_\infty$ denote the sup norm as usual and if $M = (m_{ij})$ is a matrix, let $\|M\|_\infty$ denote the corresponding operator norm, that is:

$$\|M\|_\infty = \sup_{u, \|u\|=1} \|Mu\|_\infty = \sup_i \sum_j |m_{ij}|.$$

Now suppose one wants to approximate an $\mathcal{M}$-resonant vector α; $A\alpha$ lies in the subspace $\mathbb{R}^d$ spanned by (e_i), $i = r+1, \dots, n$. Apply Dirichlet's theorem in this space to find $q \in \mathbb{N}$ such that $\|qA\alpha\|_{\mathbb{Z}} \leqslant Q^{-1/d}$; then

$$\|q\alpha\|_{\mathbb{Z}} \leq \|A^{-1}\|_\infty Q^{-\frac{1}{d}},$$

because A has *integer* entries. This proves Lemma 3 and the estimate $c(\mathcal{M}) \leqslant \|A^{-1}\|_\infty$. □

Corollary 2 is a direct consequence of this lemma, because if ω^* is $\mathcal{M}$-resonant, then using Lemma 3 the estimate (2) may be changed to

$$\|\omega - \omega^*\| \leq \frac{\sqrt{d}c(\mathcal{M})}{TQ^{1/d}},$$

and the reader will check that everything follows from this, except for the preliminary rescaling. To be more specific, one uses the linear symplectic transformation $(p, q) \to (p', q') = ({}^tA^{-1}p, Aq)$ to reduce the situation to the standard resonance case. One then uses transformation (1) to gain one more dimension, and obtain a frequency vector of type $(0, \ldots, 0, 1, \omega')$ with $\omega' \in \mathbb{R}^{d-1}$. In this way, one has rescaled a component of $A\omega^*$, and this is how the factor $c'(\mathcal{M})$ arises. In fact, this shows that $c'(\mathcal{M}) \leqslant \|A\|_\infty$.

Finally, the last assertion of the corollary about the interpretation of the expression "almost all points" which arises in Corollary 1 should be clear from the above. □

We shall add some simple remarks about the geometric meaning of the constants $c(\mathcal{M})$ and $c'(\mathcal{M})$ and obtain slightly better estimates for them.

By construction, the last d columns of A^{-1} provide integer vectors which are orthogonal to $\mathcal{M}$, and in fact they generate over $\mathbb{Z}$ the primitive module $\mathcal{M}^\perp$ orthogonal to $\mathcal{M}$. In other words, the $n \times d$ matrix E composed of the last d columns of A^{-1} defines a linear embedding of $\mathbb{Z}^d$ into $\mathbb{Z}^n$ whose image coincides with $\mathcal{M}^\perp$. One may obviously refine the estimate of $c(\mathcal{M})$ to

$$c(\mathcal{M}) \leq \|E\|_\infty = \sup_{i=1,\ldots,n} \sum_{j=r+1}^{n} |(A^{-1})_{ij}|.$$

Since A has determinant ± 1, its inverse is simply the cofactor matrix. One can still minimize this with respect to the possible matrices E, that is, with respect to the possible embeddings of $\mathbb{Z}^d$ into $\mathbb{Z}^n$ with image $\mathcal{M}^\perp$. In other words, E can be replaced by ET, where $T \in Gl_d(\mathbb{Z})$, which corresponds to changing A into

$$\begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & T^{-1} \end{pmatrix} A,$$

using block notation.

In a parallel way, the d last *rows* of A provide vectors which generate a module, or lattice, $\mathcal{M}'$ such that $\mathcal{M} \oplus \mathcal{M}' = \mathbb{Z}^n$ and one has the estimate

$$c'(\mathcal{M}) \leq \sup_{i=r+1,\ldots,n} \sum_{j=1}^{n} |A_{ij}|,$$

with a further minimization over the possible choices.

Of course the matrix B, whose value does not enter, simply corresponds to a possible change of basis of $\mathcal{M}$ itself: changing the basis changes K into PK with $P \in Gl_r(\mathbb{Z})$, and choosing $P = B^{-1}$ reduces the general situation to the case $B = \mathbb{1}_r$.

We also note, because it may sometimes be useful, that it is easy to write down explicitly a basis of *integer vectors* for the real subspace orthogonal to $\mathcal{M}$ ("resonant plane"). Assume in fact that $k_1, ..., k_r, e_{r+1}, ..., e_n$ span the whole space $\mathbb{R}^n$, which is always the case, up to a possible relabelling; then set

$$l_i = k_1 \wedge \cdots \wedge k_r \wedge e_{r+1} \wedge \cdots \wedge \widehat{e_{r+i}} \wedge \cdots \wedge e_n, \quad i = 1, \ldots, d,$$

where $\wedge$ denotes the ordinary exterior product and the vector under the "hat" is omitted. The l_i's generate over $\mathbb{Z}$ a module $\mathcal{L}, \mathcal{L} \subset \mathcal{M}^{\perp}$, and the span over $\mathbb{R}$ $(\mathcal{L} \otimes \mathbb{R})$ is the plane orthogonal to $\mathcal{M}$; in general $\mathcal{L}$ is not primitive, so $\mathcal{L} \neq \mathcal{M}^{\perp}$.

As a final remark, let us consider the case when $r = n-1$, that is, the "maximally resonant" case, or that of rational vectors. Then Corollary 2 should and indeed does reduce to Theorem 1A, except for a few minor losses which occur while going all the way round. When $r = n-1$, and writing ω instead of α, one has $A\omega = (0, ..., 0, \nu)$ with $\nu > 0$ (up to a possible change of sign in A), so

$$\nu = \sum_j A_{nj} \omega_j,$$

where A_{nj} is the last row of A. The period $T = 1/\nu$ and $T\omega \in \mathbb{Z}^n$ is the last column of A^{-1}. In other words, suppose that $\omega \cdot k_i = 0$, $i = 1, ..., n-1$, and the $n-1$ square matrices of size $n-1$ obtained by deleting a column from the matrix K have mutually prime determinants. If $k_n \in \mathbb{Z}^n$ satisfies $\det(k_1, ..., k_n) = \pm 1$, the period is given by $T = |\omega \cdot k_n|^{-1}$.

We shall now refine Corollary 2, showing that the initial point p^* need not be situated exactly on the resonant surface. This uses the remark we made above, that instead of approximating the initial point itself, one may use another, sufficiently close point. So, consider again the module $\mathcal{M}$, the associated resonant surface $\Sigma_{\mathcal{M}}$, and a point p^* lying at a distance from $\Sigma_{\mathcal{M}}$ less than $r(\varepsilon)/2$. Here (see Theorem 1A), one has $r(\varepsilon) = \lambda \frac{\varepsilon^{\alpha}}{T}$, and we want to estimate this from *below* as T runs through the values prescribed in Theorem 1A. This was done already in Chapter II, towards the end of the proof of Theorem 1A, to the effect that

$$r(\varepsilon) \geq r_0 \varepsilon^{\frac{1}{2}(1-\alpha)} > r_0 \varepsilon^{\frac{1}{2}},$$

with r_0 defined in (29) of Chapter II. So let p^* satisfy

$$\mathrm{dist}(p^*, \Sigma_{\mathcal{M}}) \leq \frac{r_0}{2} \varepsilon^{\frac{1}{2}};$$

we apply Theorem 2, in the version of Corollary 2, to a point of $\Sigma_{\mathcal{M}}$ as close to p^* as possible (it is not necessarily unique but it does not matter). The only difference is that the influence zones should be shrunk, so that the result effectively applies to p^*. So we also approximate the point on the surface by means of rational points within $r(\varepsilon)/2$, instead of $r(\varepsilon)$. Looking back to equation (3), one sees that formally this is equivalent to changing $\sqrt{n}$ into $2\sqrt{n}$, or rather $c(\mathcal{M})\sqrt{d-1}$ into $2c(\mathcal{M})\sqrt{d-1}$. This proves the following result.

***Corollary* 3.** *Let $\mathcal{M}$ be a submodule of $\mathbb{Z}^n$ of rank r, and $\Sigma_{\mathcal{M}}$ the associated resonant surface of dimension $d = n-r$; let p^* be a point in action space satisfying*

$$\operatorname{dist}(p^*, \Sigma_{\mathcal{M}}) \leq \frac{r_0}{2}\varepsilon^{\frac{1}{2}} = \frac{m}{8M}\sqrt{\frac{2E}{M}}\varepsilon^{\frac{1}{2}}.$$

Then for any point $(p(0) = p^*, q(0))$, Theorem 2 applies with the replacement of n by d and $\sqrt{n-1}$ by $2c(\mathcal{M})\sqrt{d-1}$; in the preliminary transformation (1), *$w = \|\omega^*\|_\infty$ should be changed to $c'(\mathcal{M})w$.*

Of course, one could also devise a — somewhat far-fetched — statement in the spirit of Corollary 1. We believe that Corollaries 2 and 3 should have important and far-reaching consequences. Roughly speaking, one may remember that initial conditions which belong to a tubular neighbourhood of thickness $O(\sqrt{\varepsilon})$ of a resonant surface of dimension d will be stable (in action space) for a time of the order of $\exp(c\varepsilon^{-1(2d)})$; but, of course, the order of the resonance comes into play and, given ε, this will break if this order is too high. We slightly elaborate on this heuristic picture in Chapter V, §2.

To put it differently, define subsets of phase space by

$$\mathcal{F}(d_0, c_0, \varepsilon) = \{(p^*, q^*) \in \mathbb{R}^n \times \mathbb{T}^n, \text{ such that there exists } \mathcal{M}, \text{ a submodule of } \mathbb{Z}^n, \operatorname{corank}\mathcal{M} \leqslant d_0, c(\mathcal{M}) \leqslant c_0, \text{ and } \operatorname{dist}(p^*, \Sigma_{\mathcal{M}}) \leqslant \frac{r_0}{2}\cdot\varepsilon^{1/2}\},$$

with $c_0 > 0$ and $d_0 \in \mathbb{N}$ $(1 \leqslant d_0 \leqslant n)$. Then on such a subset the stability of the action variables is essentially that of a system with d_0 degrees of freedom. It is of course tempting to let n tend to infinity (thermodynamical limit) or simply be infinite from the start (see Chapter IV, §3).

Corollaries 2 and 3 also demonstrate that there should be, *for quasi-convex systems*, a competition between stability over finite times and perpetual KAM stability, which applies, roughly speaking, to "very non-resonant" frequencies. This may be relevant in particular in celestial mechanics, as detailed below (Chapter IV, §1).

CHAPTER IV

TRANSPOSITIONS, APPLICATIONS, PROSPECTS

§1. Additional variables and an application to celestial mechanics

There is one important extension of the above results which does not require any extra work, namely one may add canonical variables in the perturbation. That is, let

$$H(p,q,I,\phi) = h(p) + \varepsilon f(p,q,I,\phi), \quad (p,q) \in \mathbb{R}^n \times \mathbb{T}^n, \quad (I,\phi) \in \mathbb{R}^m \times \mathbb{T}^m;$$

if h is quasi-convex, all the results above carry over, obtaining of course stability of the p variables only. To check this, just go through the proofs again and make sure that nothing is altered by the addition of "dummy" variables (this remark also applies in the general steep case; see [43], §1.5). It should be emphasized that such systems are degenerate from the standpoint of KAM theory, which extends to them only under some rather restrictive additional assumptions. Essentially, one should have

$$f(p,q,I,\phi) = f_1(p,I) + \varepsilon f_2(p,q,I,\phi),$$

along with the corresponding non-degeneracy condition.

As a first class of applications, one may treat in this way the "adiabatic-integrable" situations, that is, Hamiltonians of the form

$$H(p,q,\varepsilon t) = h(p) + \varepsilon f(p,q,\varepsilon t),$$

where f is periodic in $\tau = \varepsilon t$. Introducing the variable e, canonically conjugate to τ, brings this to the form

$$H(p,q,e,\tau) = h(p) + \varepsilon\left[e + f(p,q,\tau)\right],$$

which is of the type considered above.

Here we want to mention an important application, which may have far-reaching consequences in celestial mechanics: the problem of planetary systems. Since it is discussed at length by Arnol'd ([2]) in connection with the conservation of tori and by Nekhorochev ([43], §1.18 and §12) from the same viewpoint as ours, namely stability over exponential times, we shall be quite sketchy about the setting of the problem. Our results will however be not only quantitatively better than those of [43], but also qualitatively different, because Corollaries 2 and 3 seem indeed to open new perspectives, when applied in this context.

So, one wants to examine the particular case of the many-body problem in which one of them (the sun) is much heavier than the others (the planets). If the interactions among the planets is neglected, these travel along mutually independent Keplerian orbits, which are determined by their elliptic elements: the major semi-axis, the eccentricity, and the inclination, along with the corresponding angles. For reasons to be sketched below, one has to restrict attention to the case of small eccentricities and small mutual inclinations, that is, to a neighbour of the plane circular problem. The best suited variables are then the so-called Poincaré heliocentric variables. We refer to [45] (§§8–12) or to [2] (Chapter III, §2) for their definition. They read $(\Lambda, H, Z, \lambda, h, \zeta) \in (\mathbb{R}^n)^3 \times (\mathbb{T}^n)^3$; the action variables (Λ, H, Z) are simple functions of the semi-axes, eccentricities, and mutual inclinations; when eccentricities and inclinations are small, it is best to pass to symplectic polar coordinates in the pairs (H, h) and (Z, ζ), obtaining the variables $(\Lambda, \lambda, \xi, \eta, p, q)$ where, componentwise,

$$H = \frac{1}{2}(\xi^2 + \eta^2), \quad Z = \frac{1}{2}(p^2 + q^2).$$

The mass of the sun may be normalized to unity and those of the planets written as $m_i = \varepsilon\mu_i$, where ε is the ratio of the mass of the heaviest planet to that of the sun (for the solar system, $\varepsilon \approx 10^{-3}$). With these notations, the Hamiltonian reads

$$H = h(\Lambda) + \varepsilon f(\Lambda, \lambda, \xi, \eta, p, q, \varepsilon), \quad h(\Lambda) = -\frac{1}{2}\sum_{i=1}^{n} \frac{\mu_i^3}{\Lambda_i^2}.$$

Moreover, $\Lambda_i = \mu_i\sqrt{a_i}$, a_i being the major semi-axis of the ellipse osculating to the trajectory of the i-th planet at a given time; hence, controlling Λ is equivalent to controlling the semi-axes.

Before applying any theorem, one should make sure that the perturbation is indeed small, and this is true only so long as the $n+1$ bodies do not come too close to each other. Since we shall have control on the a_i's only, the only region in phase space when this implies estimates on the eccentricities and inclinations is near the plane circular problem. This is because plane circular motion with the planets travelling in the same direction achieves a maximum of the angular momentum of the system, and the latter is a conserved quantity. More precisely, let G_i be the angular momentum vector of the i-th planet ($\|G_i\| = m_i(a_i(1-e_i^2))^{1/2}$, e_i the eccentricity), $G = \Sigma_i G_i$, the total angular momentum, and $N = \varepsilon^{-1}G$, independent of ε. Pick $2n$ positive numbers α_i, β_i satisfying

$$0 \le \alpha_1 \le \beta_1 \le \alpha_2 \le \cdots \le \alpha_n \le \beta_n.$$

A *domain of planetary motion* $B(\alpha, \beta, \gamma)$ is a region of phase space such that

$$\alpha_i \le a_i \le \beta_i, \quad i = 1, \ldots, n \text{ and } \|N\| \ge \gamma;$$

γ is a number such that $0 < \gamma_0(\alpha, \beta) < \gamma < \gamma_m(\alpha, \beta)$. Here $\gamma_m(\alpha, \beta)$ is the maximal possible value of $\|N\|$ under the conditions imposed on the a_i's; it corresponds to plane circular motions with radii β_i, and $\gamma_0(\alpha, \beta)$ is the largest value of $\|N\|$ corresponding to possible collisions among the planets and/or with the sun. We refer to [43] (§12) for a detailed discussion of the fact that on a domain $B(\alpha, \beta, \gamma)$ one may indeed apply the results about stability over exponentially long times; this discussion carries over without any change.

Since $h(\Lambda)$ is a convex function, Theorem 2 applies, yielding stability of the major semi-axes over exponentially long times. The implications of Corollaries 2 and 3, however, are much more intriguing. Indeed, these assert that resonant, or even nearly resonant, trajectories are *privileged*, from the point of view of finite time stability. Resonance here simply means resonance between the inverses of the periods of the motions along the instantaneous ellipses, that is, the inverses of the "years"; the relation between these periods of revolution and the values of the semi-axes, that is, Kepler's third law, follows from the expression for $h(\Lambda)$.

But there has been a long-standing discussion about the fact that celestial bodies seem to pick resonant trajectories more often than could be expected from a mere statistical effect. These speculations about "harmonic motions" could be traced back to Pythagoras, Plato or Kepler, but in modern terms this was forcefully advocated by Molchanov (see [40], [41], [5], [27]) who noticed the existence of many "simple" resonance relations between the planets of the solar system and inside the satellite subsystems around Jupiter, Saturn and Uranus. He was immediately strongly criticized on the ground that these relations were not really "astonishing" and would often occur among numbers or vectors picked "at random"; he then replied to these criticisms, trying in particular to give a precise definition of the adjective "simple" used above. Since no repeatable experiment can be performed in this case, the evidence is bound to remain fragile. In any case, since then a lot of work has been devoted to the subject, including resonances which involve artificial satellites. Many of these resonances are ascribed to non-Hamiltonian causes, for example tidal effects, but there seems to remain some "mystery" buried in a mass of controversial observations. In Molchanov's terms: "Why are planets and satellites locked into simple resonances, whereas the rings of Saturn or the asteroid belt have gaps in these places?" Even if particular assertions may be challenged, this seems to ask us an authentic riddle.

The results above offer the first purely Hamiltonian partial explanation for this; if the bodies must linger much longer about resonant trajectories than elsewhere, after some time these will become indeed the most populated places. This is not so simple, however, and in accordance with the spirit of the above quotation, we have really set up a "competition" between finite time stability and perpetual stability of the KAM type, since the latter favours very *non*-resonant trajectories. According to the concrete situation at hand, it is quite possible that one or the other kind of stability actually prevails. In this context, we insist that the stability estimates imply that the bodies remain locked in resonance zones for exponentially long times, but of course they do not preclude small amplitude ($O(\sqrt{\varepsilon})$) "chaotic" motions inside such a zone, on much shorter timescales.

One should also note that the present results have a wider range of validity than KAM results (specifically the theorem proved in [2]). First, from a practical point of view, although it is not realistic, the threshold of validity which we obtain is not nearly as small as the one of KAM theory; it could even perhaps be pushed to some realistic value, using computer assisted estimates. Second, finding invariant tori of *maximal* dimension (half the dimension of the phase space) requires that the unperturbed system be integrable with respect to *all* the variables. Here, this translates into the fact that one must perturb from the *exact* plane circular problem (as in [2]) and so, in the perturbed problem, the eccentricities and the inclinations should be of the order of (a power of) the perturbation parameter, that is, extraordinarily small. It should be noted however that a version of KAM theory has been developed in which one looks for *low-dimensional* tori, that is, tori which are *not* of maximal dimension (see in particular [10], [47], [53], [57] and references therein); this in turn requires only *partial* integrability of the system, as is the case here. To our knowledge, this theory has never been applied specifically to the planetary problem, although the difficulties are probably of a technical nature only (see however [51]). More significant may be the fact that the set of tori one thus finds is of zero Lebesgue measure. We shall briefly comment on this when discussing Arnol'd's diffusion in Chapter V, §2.

Returning to the results on stability over finite times, these only require that the system be close enough to the plane circular problem so as to avoid collisions. This is the condition $\|N\| > \gamma_0$ in the definition of a domain of planetary motions, where γ_0 is independent of ε. It defines an "order 1" neighbourhood of the plane circular problem, the most favourable case arising when all the planets have the same mass; indeed no condition on the momentum can possibly prevent collision as the mass of at least one planet vanishes, as in the restricted three-body problem.

§2. Transposition to other contexts and degenerate cases

The results of Chapters II and III can be transposed, at least to some extent, to the other circumstances under which classical perturbation theory applies. We mention:

i) perturbation of an integrable Hamiltonian vector field;

ii) neighbourhood of an elliptic fixed point of a Hamiltonian vector field;

iii) neighbourhood of a Lagrangian torus over which a Hamiltonian vector field induces a flow conjugate to a linear one.

Each situation has its discrete analogue where Hamiltonian vector fields are replaced by symplectic maps. Of course, i) is the problem we have been dealing with, but we listed it for the sake of completeness. We refer to [4] (and [21] as far as iii) is concerned) for the elementary details. Continuous and discrete problems essentially correspond under the two inverse operations of *section* and *suspension*. Let us briefly illustrate this on i). It is well known how to construct a local Poincaré section for an autonomous Hamiltonian vector field. On the other hand, start from the discrete problem, which is described as follows: let B_δ be the open ball of radius $\delta > 0$ centred at the origin in $\mathbb{R}^n$, and $\mathbb{A}_\delta = \mathbb{T}^n \times B_\delta$ an annulus. Let f_0 be defined as

$$(1) \qquad (\theta, r) \to f_0(\theta, r) = \big(\theta + \omega(r) \bmod \mathbb{Z}^n, r\big), \quad (\theta, r) \in \mathbb{A}_\delta.$$

We assume that $\omega = \nabla h$ is the gradient of a function h; then f_0 is an integrable globally canonical map with generating function h. We consider the map f generated by a perturbation of h, $\Sigma(\theta, r') = h(r') + \sigma(\theta, r')$, where σ is small (of order ε), and f is implicitly defined by

$$(\theta', r') = f(\theta, r) = \left(\theta + \omega(r') + \frac{\partial \sigma}{\partial r'} \bmod \mathbb{Z}^n,\ r - \frac{\partial \sigma}{\partial \theta}\right).$$

Assume that h and σ are analytic and that h is a *convex* function; then one has the following stability result for the variable $r \in B_\delta$:

If $\varepsilon = \|\sigma\| \leqslant \varepsilon_0$, *then* $\|r_s - r\| \leqslant c\,\varepsilon^b$ *when* $|s| \leqslant c \exp(\varepsilon^{-a})$, $s \in \mathbb{Z}$; we use the notation $(\theta_s, r_s) = f^s(\theta, r)$, $\theta_0 = \theta$, $r_0 = r$.

The definition of the norms is as in Chapters II and III, and the exponents (a, b) are as in Theorem 2, *but* in dimension $n+1$. All the later refinements could be added.

To view such a result as a corollary of those of Chapter III, one must build a suspension of the map f, that is, realize it as the time 1 map of a flow associated with a Hamiltonian $H(\theta, r, t)$ which is periodic of period 1 in the time variable t. Here the real difficulty lies in the regularity assumption; in fact, the construction is quite easy in a C^∞ setting, much less so if one requires analyticity, as is necessary here. There is no obstruction however, and Kuksin proves (in [34]) the existence of H, which is an $O(\varepsilon)$-perturbation of h, of which f_0 is the time 1 map. So one is led to the quasi-convex case, having to deal with a periodic perturbation of a convex Hamiltonian (this is why h must be convex, not *quasi*-convex). Of course, it would still be useful to write a direct proof of the result above. Note that one has to cope with the fact that energy conservation is not available any more.

We shall now dwell a bit more on situation ii), which has been the subject of much study, for the past century at least. We shall not mention any more the discrete cases corresponding to ii) and iii). Situation ii) is degenerate from the point of view of perturbation theory, and before we turn to it, it is useful to look at another, slightly simpler but quite similar problem: the perturbation of harmonic oscillators; of course, this is also interesting for its own sake. So let

$$H(p, q) = \omega_0 \cdot p + \varepsilon h_1(p) + \varepsilon^2 f(p, q), \tag{3}$$

where $\omega_0 \in \mathbb{R}^n$ is a non-zero vector, h_1 and f are analytic functions, and h_1 is quasi-convex. We perform the scalings $t \to \varepsilon t$, $H \to \varepsilon^{-1} H$, and obtain, keeping the same notations for simplicity,

$$H(p, q) = \frac{\omega_0}{\varepsilon} \cdot p + h_1(p) + \varepsilon f(p, q). \tag{4}$$

We write $\omega_1 = \nabla h_1$; degeneracy manifests itself through the fact that the frequency $\omega = \varepsilon^{-1}\omega_0 + \omega_1$ is of the order of ε^{-1}. Fix $\varepsilon > 0$ small enough, suppose that $\omega(0)$ is rational of period T, and go through Chapter II again.

The quantities m and M which measure the non-linearity and convexity now refer to h_1 and are independent of ω_0; this implies that the iterative lemma carries over without any change. However, in the geometric reasoning leading to equation (23) of Chapter II, one should take into account the fact that $\|\omega(0)\|$ is of order ε^{-1}; so just replace Ω by $2\varepsilon^{-1}\Omega$, where Ω here stands for $\|\omega_0\|$ (indeed $\|\omega(0)\| \leqslant 2\varepsilon^{-1}\|\omega_0\|$ for ε small enough). Then (24) still defines $\mathcal{T}(\varepsilon)$, with the replacement $\Omega \to 2\varepsilon^{-1}\Omega$, and the rest is unaltered. So the "model statement" also carries over with only this modification.

Finally, Theorem 1A is valid for the Hamiltonian (4), except for the substitution $\Omega \to 2\varepsilon^{-1}\|\omega_0\|$. We said we have *fixed* $\varepsilon > 0$ so that $\omega(0)$ is rational of period T. Now the only requirement is that ε satisfy inequalities (27) of Chapter II. The second of these inequalities is very much weakened by the substitution on Ω, but one has to add the requirement $\|\omega(0)\| \leqslant 2\varepsilon^{-1}\|\omega_0\|$, that is, $\|\omega_1\| \leqslant \varepsilon^{-1}\|\omega_0\|$, which is a weak bound on ε. Note that Ω does not enter in the definition (26) of the quantities λ and τ.

Now let $p^* \in \mathbb{R}^n$ be a point in action space, $\omega_1^* = \omega_1(p^*)$, $\omega^* = \varepsilon^{-1}\omega_0 + \omega_1^*$, and we wish to approximate ω^*. Here comes the key observation: although we are working at *high* frequencies (of order ε^{-1}), there are always *low* frequency (of order 1) orbits close to a given one, and this phenomenon is uniform in ε as this quantity goes to zero. Indeed it only expresses the fact that for any value of $\varepsilon > 0$, $\varepsilon^{-1}\omega_0$ can be shifted back into the unit cube, using an integer vector. This simple but physically significant property will allow one to cope with the degeneracy. Let us now implement the above: with our notations, formula (2) of Chapter III is unchanged; *define* ω_1 by the equality $\omega = \varepsilon^{-1}\omega_0 + \omega_1$, so

$$\|\omega_1 - \omega_1^*\| \leq \frac{\sqrt{n}}{TQ^{1/n}}.$$

Since the map $p \to \omega_1(p)$ is locally invertible, one finds p close to p^* such that $\omega_1 = \omega_1(p)$, and the rest of the reasoning needs no modification at all. Of course the matrix A now denotes the Hessian matrix of h_1, and analogously for the other quantities. Let us state the result.

***Theorem* 3.** *Consider the Hamiltonian* (3) *above. Then the result stated as Theorem* 2 *of Chapter* III *holds, with the following qualifications:*

i) Ω^* *is replaced by* $2\varepsilon^{-1}\|\omega_0\|$, *with the additional threshold condition on* ε:

$$\|\nabla h_1\| \leq 2\varepsilon^{-1}\|\omega_0\|;$$

ii) *the constant* $\mathcal{T}_0^*$ *now has the value*

$$\mathcal{T}_0^* = 1.5 \cdot 10^{-2} \frac{\sigma}{\|\omega_0\|};$$

iii) *scalings* (1) *of Chapter* III *are not performed, so n should be changed to $n+1$ in the statement, and quantities m, M, and so on, refer to the original Hamiltonian h_1.*

Although this result has been obtained in an almost effortless way, we have stated it as a "theorem", because we believe it is quite significant: indeed, this is the first non-linear stability result over exponential times to be obtained in a degenerate case. Before we comment on this, let us briefly return to the statement above: ii) comes from the fact that we have been working with the Hamiltonian (4); returning to (3) involves a scaling of the time variable which gains back the factor ε that had been lost before. Scalings (1) of Chapter III cannot be performed because they involve the frequency, which is here of order ε^{-1}; hence iii). In particular, we obtain for the time of stability an exponent $a = \dfrac{1}{2n+3} - \eta$ for any $\eta > 0$.

The Hamiltonian (3) may arise naturally, for example in the following context: consider again a perturbation of a system of harmonic oscillators:

$$H(p,q) = \omega_0 \cdot p + \varepsilon g(p,q). \tag{5}$$

Assume that g contains only a finite number of harmonics, that is, it is a trigonometric polynomial in q. Then, away from a *finite* number of resonance surfaces, one can perform one step of the reduction to normal form, which leads (after a change of variables) to

$$H(p,q) = \omega_0 \cdot p + \varepsilon \langle g \rangle (p) + \varepsilon^2 f(p,q).$$

So, if the space average $\langle g \rangle$ is quasi-convex, we are reduced to the Hamiltonian (3).

To appreciate the significance of Theorem 3, one should beware of an important possible misunderstanding. We have proved a result which is completely independent of ω_0, in particular its arithmetical properties. In fact, if one sets $\omega_0 = 0$ in (3), it reduces to the non-degenerate case, and apart from some details which we leave to the reader to settle, we do recover the corresponding result. Now, if ω_0 is strongly non-resonant, say satisfies the usual Diophantine condition

$$\exists \gamma > 0,\ \tau > n-1, \text{ such that } \left| \omega_0 \cdot k \right| \geq \gamma |k|^{-\tau},\ \forall k \in \mathbb{Z}^n \setminus \{0\}, \tag{6}$$

it is easy to derive a stability result over exponential times. Indeed, starting from the less explicit form (5), one simply builds up the Birkhoff series, and makes use of (6) to control the process, using either an iterative method or a majorant series. This completely algebraic construction allows us to prove a stability result over exponential times, but one which is very sensitive to the arithmetics of ω_0.

Such elementary estimates are derived for example in [7], and we propose to call them *Gevrey type estimates*, the reason for this terminology being clarified in Appendix 2. In results of this kind, one considers the *linear* Hamiltonian $h(p) = \omega_0 \cdot p$ as the unperturbed system. Theorem 3 lies definitely deeper: one considers the *non-linear* Hamiltonian $h(p) = \omega_0 \cdot p + \varepsilon h_1(p)$ as the unperturbed part and takes advantage of the non-linearity (anharmonicity) and convexity to derive an estimate which is independent of the unperturbed frequency ω_0; such estimates we propose to call *Nekhoroshev type estimates*. A similar result should be valid (with other exponents) if h_1 is only assumed to be *steep*, but this seems very cumbersome to obtain if one applies Nekhoroshev's original method.

We finally note that Zaslavskii and coworkers (see [58]) have recently studied, mostly from a physical and numerical standpoint, systems which are perturbations of Hamiltonians of type

$$h(p) = \omega_0 \cdot p_0 + h_1(p_1), \quad p = (p_0, p_1) \in \mathbb{R}^{l+m} = \mathbb{R}^n,$$

where h_1 is non-degenerate (say convex); in such a situation, instabilities usually occur on much shorter time-scales, and Nekhoroshev type results are excluded in general.

We now return to the problem of studying a Hamiltonian vector field in the neighbourhood of an elliptic fixed point, and we shall use Theorem 3 in order to derive a result for this situation. We denote by $\pm i\alpha_j (i = \sqrt{-1}), j = 1, \ldots, n$, the eigenvalues of the linearized system at the fixed point, which we take as the origin of the coordinate system (x_j, y_j), $j = 1, \ldots, n$, of $\mathbb{R}^{2n}$; we write $z = (x, y) \in \mathbb{R}^{2n}$. We assume that the linear part can be diagonalized and that there is no resonance of order $\leqslant s$ (a positive integer), which means, writing $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$, that

$$\forall k \in \mathbb{Z}^n \backslash \{0\} \quad \alpha \cdot k \neq 0 \quad \text{if} \quad |k| = |k_1| + \cdots + |k_n| \leq s.$$

Let $r_j = (1/2)(x_j^2 + y_j^2)$, $r = (r_1, \ldots, r_n) \in \mathbb{R}^n_+$. Following Birkhoff, one can perform a canonical transformation so as to put the Hamiltonian in the form

$$H(z) = H(x, y) = H^{(s)}(r) + O(\|x, y\|^{s+1}). \tag{7}$$

$H^{(s)}(r)$ is a polynomial of degree at most $[s/2]$ in the r_j's; we assume that H is analytic, so that the rest is a convergent power series whose terms are of degree at least $s+1$ in x_j, y_j. We suppose that $s > 4$, so $H^{(s)}$ has the form

$$\begin{aligned} H^{(s)}(r) &= \sum_j \alpha_j r_j + \frac{1}{2} \sum_{i,j} \alpha_{ij} r_i r_j + O(\|r\|^3) \\ &= \alpha \cdot r + \frac{1}{2}(Ar \cdot r) + O(\|r\|^3), \end{aligned} \tag{8}$$

where $A = (\alpha_{ij})$ is a symmetric matrix. This way of writing determines the sign of the α_j's; note that if they are all of the same sign, the stability problem is immediately settled (positively), because the origin is a local maximum (or minimum) of the Hamiltonian. Here we shall work again under the convexity assumption that A is a, say, positive matrix and we let $m > 0$

(respectively, $M \geqslant m$) be its smallest (respectively, largest) eigenvalue. Under these assumptions we prove the following theorem.

Theorem 4. *Consider a trajectory $z(t)$ under the evolution governed by $H(z)$ (see (7) and (8)). There is a constant $\nu > 0$ such that if $z = z(0)$ is small enough and satisfies*

$$r_j = \frac{1}{2}(x_j^2 + y_j^2) \geq \nu \|z\|^{2+\frac{1}{n+2}}, \quad j = 1, \ldots, n, \tag{9}$$

($r_j = r_j(0)$, and so on), then one has

$$\left\| r_j(t) - r_j \right\| \leq \frac{\nu}{2} \|z\|^{2+\frac{1}{n+2}}, \quad j = 1, \ldots, n,$$

provided that t satisfies

$$|t| \leq \mathcal{T} \exp\left(\|z\|^{-\frac{1}{n+2}}\right),$$

where $\mathcal{T}$ is some strictly positive constant.

Before making some comments, we show how this is an easy consequence of Theorem 3; as the reader will see, we prove in fact a more precise and slightly stronger statement. First, introduce the usual symplectic polar coordinates (r, θ) defined as

$$x_j = \sqrt{2r_j} \cos \theta_j, \quad y_j = \sqrt{2r_j} \sin \theta_j.$$

We fix $z = z(0)$ and set $\varepsilon = \Sigma_j r_j = (1/2) \| z \|^2$. Then perform the scaling $r = \varepsilon\rho$, $H = \varepsilon K$, which multiplies the symplectic form by the factor ε and leaves the equations invariant. Then

$$K(\rho, \theta) = \alpha \cdot \rho + \frac{1}{2}\varepsilon(A\rho \cdot \rho) + \varepsilon^2 f(\sqrt{\varepsilon\rho}, \theta),$$

where we write "componentwise" $\sqrt{r} = (\sqrt{r_1}, \ldots, \sqrt{r_n}) \in \mathbb{R}^n_+$. The function f is analytic, periodic in θ, and we let σ be its analyticity width in θ.

We may now apply Theorem 3, provided that we keep away from the singularities at $\rho_j = r_j = 0$, $j = 1, \ldots, n$. Now recall that in Theorem 2 (or even Theorem 1), the analyticity width in the action variables need not be of order 1, but only at least equal to the confinement radius. This is quite an important feature in the present context, because it says how close we may approach the singularities. From Theorem 3 (or rather Theorem 2), we compute the confinement radius:

$$\left\| \rho_j(t) - \rho_j \right\| \leq 10^{-2} \frac{\sigma}{M} \varepsilon^a = 10^{-2} \frac{\sigma}{M} \left(\frac{1}{2}\|z\|^2\right)^a < 10^{-2} \frac{\sigma}{M} \|z\|^{2a}.$$

Here $a = \dfrac{1}{2n+4}$ and we set $\nu = 10^{-2} \dfrac{\sigma}{M}$. We thus get the inequality on the drift of the action variables, with a time of validity

$$\mathcal{T}(z) = \mathcal{T} \exp(\varepsilon^{-a}) > \mathcal{T} \exp(\|z\|^{-2a}), \quad \mathcal{T} = 1.5 \cdot 10^{-2} \frac{\sigma}{\|\alpha\|}.$$

All this is valid provided that the inequalities $\rho_j(t) > \nu\varepsilon^a$ keep holding during the time $\mathcal{T}(z)$, which is guaranteed by inequality (9). We have thus proved the theorem and computed the quantities ν and $\mathcal{T}$. The threshold of validity, that is, the maximal possible value of $\|z\|$, could of course also be computed explicitly, using Theorems 2 and 3. □

We add some short comments about this result. First, a similar, slightly weaker estimate holds with $s \geqslant 4$ (the minimal order of a possible resonance), which is also the condition under which KAM theory applies. We took $s > 4$ for convenience only.

Second, one may improve on this result if s is really larger, by performing some steps of the Birkhoff normal form algorithm and applying this type of reasoning afterwards. In fact, this strategy may also be used in the contexts of Theorems 2 and 3, at least under certain circumstances. This is a combination of the usual method and the closed orbit method we put forward in this paper, and this may be useful in trying to improve the estimates, possibly in a computer assisted way.

Third, if the matrix A has no definite sign, steepness cannot be decided from the knowledge of α and A alone and one must compute more Birkhoff invariants (hence s must be larger, at least $\geqslant 6$); then, in the steep case, it would in principle be possible to apply a variant of the strategy of [43] to prove a result of the type of Theorem 4, but again this looks very cumbersome indeed.

Fourth, we have not proved an exponential "exit time" estimate because of the seemingly artificial and spurious requirement (9) on the initial conditions. This stems from the fact that we had to use the action-angle variables (r, θ), which present singularities on the coordinate planes $r_j = 0$. Exactly the same difficulty is encountered (and left unsolved) in KAM theory (see, for example, [46], last paragraph). We do not know if and how it may be overcome and accordingly we have had to leave out small cusp-shaped regions with vertices at the fixed point.

Fifth, there are some obvious generalizations which may be useful. For example, one may require *quasi*-convexity only: in this context, it means that the quadratic form $Ar \cdot r$ has to be of definite sign, but only when restricted to the plane $\alpha \cdot r = 0$. Alternatively, one may consider *periodic* perturbations: A must have definite sign but the Hamiltonian may depend periodically on time.

Lastly, the same comment is in order concerning Gevrey type estimates, as was discussed in connection with Theorem 3 (we again refer the reader to Appendix 2). If α is a Diophantine vector, that is, if it satisfies inequalities (6) (with α in place of ω_0), one obtains exponential stability estimates in an elementary, purely algebraic way, by controlling the growth of the Birkhoff series (see, for example [24] and [25]). Again, these estimates depend strongly on the arithmetics of α. We note that if one wants to derive Nekhoroshev type estimates, as we did, one cannot use the usual complex coordinates ($w = x+iy$ and the complex conjugate vector) because these are not directly related to the action-angle variables of the unperturbed part, unless the latter is taken to be *linear*, as is the case in Gevrey type estimates.

In some sense, case iii) mentioned at the beginning of this section, namely the neighbourhood of an invariant Lagrangian torus, is easier to disentangle. To start with, by the symplectic tubular neighbourhood theorem, one may symplectically describe the neighbourhood of the torus as $\mathbb{T}^n \times B_\delta$, where B_δ is again the open ball of radius δ centred at the origin in $\mathbb{R}^n$. We still denote the corresponding coordinates as $(\theta, r) \in \mathbb{T}^n \times B_\delta$; the invariant torus has the equation $r = 0$, and after conjugation the flow on it is linear with vector $\alpha \in \mathbb{R}^n$. The crux of the matter is that if α is *not* Diophantine, the situation is structurally unstable and it seems quite hard to say anything at all. Indeed normal theory *at first order* already requires that α be strongly irrational.

Suppose now that α is indeed Diophantine; as a side remark we note that this implies, under weak regularity assumptions, that the torus is Lagrangian, so this need not be part of the hypothesis any more. Then one is reduced to a situation very similar to that of the elliptic point, namely, after a change of coordinates, to the Hamiltonian

$$
\begin{aligned}
&H(\theta,r) = H^{(s)}(r) + O(\|r\|^{s+1}),\\
(10)\qquad &H^{(s)}(r) = \sum_j \alpha_j r_j + \frac{1}{2}\sum_{i,j} \alpha_{ij} r_i r_j + O(\|r\|^3)\\
&\qquad\quad = \ \alpha\cdot r \ + \frac{1}{2}(Ar\ r) + O(\|r\|^3).
\end{aligned}
$$

Here s is arbitrary and $r \in B_\delta$ runs through a neighbourhood of the origin. No singularities occur and one may derive, without any convexity or steepness assumptions, Gevrey type estimates, because α is highly non-resonant. This is done as in the case of the elliptic fixed point, except that here complex coordinates cannot be used ((θ, r) do not arise as polar coordinates); this situation is similar to the case of the elliptic fixed point, with the latter "blown-up". These estimates seem not to have yet been written out in detail, although they describe in particular the time needed to move away from a Kolmogorov invariant torus. Note that the latter is a problem with *two* small parameters: ε describing the perturbation from integrability and $\|r\|$ measuring the distance from the torus.

§3. Systems with (infinitely) many degrees of freedom

Corollaries 2 and 3 of Chapter III are perhaps of great relevance to a class of problems with a large—possibly infinite—number of degrees of freedom. Here we are thinking of simple statistical models, such as spin lattices, chains, crystals, and so on, as well as some particular PDE's, mainly in one space dimension. These problems have been studied during the past few years with varied successes and a special emphasis on KAM theorem; the bibliography of [48] contains some of the important references on the subject.

Rather than being too vague or abstract, it is perhaps best to consider a simple example which displays the main features and difficulties: a one-dimensional chain of rotators with nearest neighbours interactions. We thus look at the Hamiltonian

$$H(p,q) = \sum_{i=1}^{N} \Big(\frac{1}{2}p_i^2 + \varepsilon V(q_{i+1} - q_i)\Big).$$

Here V is a potential with a critical point at some value $a > 0$ ($V'(a) = 0$) representing the average distance between two free rotators. If N is finite, one should add boundary conditions (for example, periodicity, say $q_{N+1} = q_1$) and then look for results which do not depend on N, at least asymptotically when this tends to infinity (thermodynamical limit). Alternatively, one may set $N = \infty$ from the start, with a suitable mathematical setting.

Now, the point we want to make in this short section is that *localization is resonance* and that, by the results of Chapter III, convexity and resonance together imply stability, because of a local abundance of periodic orbits. From this, it should be possible to derive strong "non-linear localization results". Indeed, suppose that at time $t = 0$ one jiggles d of the N rotators (assume that N is finite and impose periodic boundary conditions for simplicity), that is, we have the following initial conditions:

$$p_i(0) \text{ arbitrary}, \quad i = 1,\ldots,d; \quad p_i(0) = 0, \; i = d+1,\ldots,N;$$
$$q_i(0) \text{ arbitrary}, \quad i = 1,\ldots,N.$$

This is a resonant situation, since the frequency vector is none other than

$$\omega(0) = \big(p_1(0),\ldots,p_d(0),0,\ldots,0\big),$$

so we start on a d-dimensional resonant surface. Now apply Corollary 2 of Chapter III and conclude that the action variables are stable over an interval of time essentially of the order of $\exp(\varepsilon^{-1/(2d)})$ for ε small enough, independently of the number N of degrees of freedom. Corollary 3 adds the important flexibility that one may even allow for some energy to be fed into the remaining $N-d$ rotators, still getting essentially the same result.

But all this is cheating, of course! What is it that is lacking? Not much really; only the fact that the number of degrees of freedom is buried in the definition of the norms we use, for example, to measure the strength of the perturbation. These do not take advantage of the fact that the problem displays some locality property in *real* space, to wit that the interaction involves nearest neighbours only. Pöschel, starting from the work of various authors (including himself) has abstracted a general scheme to deal with these *local structures* (see [48]). We hope that combining this with the principles of the present paper could lead to interesting results, of the kind that were carelessly stated above.

Again convexity is here an essential ingredient. In particular, chains of perturbed harmonic oscillators could be treated, using Theorem 3 in §2, only insofar as the non-linear perturbation presents some convexity property *in action-angle variables* ("normal modes" coordinates). Unfortunately, this is a rather unnatural requirement in this context (compare the Fermi–Pasta–Ulam model).

We recall also that KAM theory has been recently extended to some classes of infinite-dimensional systems; under some technical assumptions, one proves the existence of either *finite*-dimensional invariant tori (Kuksin, Pöschel, Wayne, et al.) that is, quasi-periodic motions with finitely many frequencies and/or *infinite*-dimensional invariant tori (Vittot, Pöschel, et al.). We refer to [48] and [35] for a bibliography. Our last remark is that, among other conditions, KAM theory requires a priori some form of non-degeneracy condition, as is usual, but that in the context of statistical mechanics, this essentially *implies* convexity. This stems from the fact that the unperturbed integrable system is usually assumed to be an ensemble of non-interacting *identical* objects. Non-degeneracy means that each of the microscopic entities is "truly non-linear" (for example, a rotator, rather than a harmonic oscillator). But then the Hessian matrix of the unperturbed Hamiltonian will obviously be diagonal with identical non-zero entries, which implies convexity.

§4. Steepness, quasi-convexity, and closed orbits

We have repeatedly emphasized that the closed orbit method we use in the present paper is restricted to the quasi-convex case and that stability in the general steep situation is just not amenable to it. Maybe this could provoke a renewal of interest for the latter case, which has been very little investigated? All the more since in the analytic framework, taking advantage of the rigidity of analytic objects, Il'yashenko has given a completely algebraic characterization of steepness (in [30]), which was originally introduced as a C^∞-notion. It would thus be quite interesting to rewrite Nekhoroshev's proof ([43], [44]), trying to clarify the relationship with geometry and singularity theory, from which, incidentally, steepness originally emerged. One could also try to isolate interesting subclasses of steep functions, beyond the quasi-convex one, which, we recall, is the only one where steepness can be read off the 2-jet of the function.

On the other hand, quasi-convexity has been recognized, in the past few years, to imply very specific properties, and from this standpoint the stability properties explored in the present paper fit well into the picture. It may thus be useful to mention some of these features. (Quasi-)convexity is naturally appealing first of all, because the kinetic energies which one comes across in physics usually enjoy this property. This is also linked to the fact that even in a non-perturbative framework, Hamiltonians are usually derived from Lagrangians, and that convexity with respect to the action variables goes along with the existence and nice properties of the Legendre transform.

Then, convexity is also the natural and simplest framework of variational methods, for example if one tries to prove the existence of closed orbits "in the large", that is, for arbitrary Hamiltonians with compact energy surfaces. The assumption that the energy surface is convex enormously simplifies the problem and much more precise results are known than in the general case.

Some rather subtle specificities of convexity have been revealed recently. Let us consider, as in §2 (formula (1)), a globally canonical integrable map of the annulus:

$$(\theta, r) \to (\theta + \omega(r) \bmod \mathbb{Z}^n, r), \quad (\theta, r) \in \mathbb{A}^n = \mathbb{T}^n \times \mathbb{R}^n, \quad \omega = \nabla h.$$

Here we shall need only a finite order of differentiability, so that everything is really local in the r variables. One considers a globally canonical perturbation of the above (see (2) in §2). If $n = 1$, under the twist condition $\omega'(r_0) \neq 0$, Birkhoff showed that any invariant curve Γ close to the circle $r = r_0$ is the graph of a continuous function, that is, there exists $\psi \in C^0(\mathbb{T}^1, \mathbb{R})$ such that $\Gamma = \Gamma_\psi = \{(\theta, \psi(\theta)), \theta \in \mathbb{T}^1\}$. Moreover, ψ is in fact Lipschitz, and its derivative (which exists almost everywhere) satisfies an a priori estimate; Birkhoff's theory is in fact *not* of perturbative nature, but we restrict ourselves to this case for simplicity.

Now, if $n > 1$, Herman ([28]) shows that various pathologies may arise, unless one restricts consideration to *Lagrangian* tori homotopic to $r = 0$ *and* if one assumes *monotone twisting* (convexity), that is, that the matrix $A(r) = \partial\omega/\partial r = \nabla^2 h$ has a definite sign. Only in that case can Birkhoff's regularity theory be generalized to more than one dimension, at least in a perturbative way.

We shall devote the end of this section to a brief discussion of the existence of (exact) closed orbits for near integrable Hamiltonians; we first recall the old perturbative result, essentially due to Poincaré ([45], Chapters III and IV), and emphasize how the specificity of quasi-convexity is already quite visible at this level, something which never seems to be pointed out in the literature.

Let us go back to the setting of Chapter II. Let $H = h(p) + \varepsilon f(p, q)$ be a perturbed Hamiltonian; ε is written explicitly and there is no loss of generality in assuming that it is $\geqslant 0$. When $\varepsilon = 0$, $p = 0$ is an invariant periodic torus of period T and rational frequency $\omega(0) = \omega_0$. We do *not* assume quasi-convexity for the moment, only non-degeneracy: $\det A_0 \neq 0$ ($A(p) = \nabla^2 h$, $A_0 = A(0)$). As in Chapter II, if $g(q)$ is a function on $\mathbb{T}^n$, $\langle g \rangle$ denotes its average along ω_0. Let $\langle f \rangle(q) = \langle f \rangle(0, q)$ be the average of the perturbation of the torus $p = 0$. It is constant on the orbits of the linear flow along ω_0 and can be thought of as a function on the space of orbits $\mathcal{O} = \mathbb{T}^{n-1}$. We suppose that it is a Morse function on this space. Viewed on the torus $\mathbb{T}^n$, it means that the Hessian matrix

$F_0 = \dfrac{\partial^2}{\partial q^2}\langle f\rangle(0, q^{(0)})$ at a critical point $q^{(0)}$ has a *one-dimensional* kernel spanned by ω_0 (critical points are in fact critical orbits). The following assertion holds.

Theorem. *Let $H(p, q) = h(p)+\varepsilon f(p, q)$ ($\varepsilon \geqslant 0$) be such that $p = 0$ is, for $\varepsilon = 0$, a periodic torus of frequency ω_0 and period T. Assume that h is non-degenerate at $p = 0$ ($\det \nabla^2 h(0) \neq 0$) and that the average $\langle f\rangle(0,q)$ has a one-dimensional null space (spanned by ω_0) at its critical points.*

Then for $\varepsilon > 0$ small enough there exist, in an $O(\varepsilon)$ neighbourhood of $p = 0$, at least 2^{n-1} orbits of period T, including multiplicity, of which at least n are geometrically distinct.

Moreover, if h is quasi-convex, one may specify the spectral type of these orbits and assert that there are at least $\binom{n-1}{k}$ k-hyperbolic orbits, $k = 0, 1, \ldots, n-1$.

Here we call an orbit k-hyperbolic if it has k pairs of Floquet exponents which are not purely imaginary; recall that μ is a Floquet exponent of some orbit of period T if $\lambda = e^{\mu T}$ is an eigenvalue of the linearized return map. The last item thus says that *in the quasi-convex case* one may predict the linear stability of the orbits which are born from a periodic torus. For instance, there will arise at least one linearly stable (that is 0-hyperbolic or elliptic) orbit. This comes from the fact that the Floquet exponents, which are paired in pairs of opposite signs, may be expanded in powers of $\sqrt{\varepsilon}$ (two of them vanish); at first order, they are of the form $\pm(\varepsilon\Omega_j)^{1/2}$, $j = 1, \ldots, n$, where the Ω_j's are the eigenvalues of the matrix $-A_0F_0\left(A_0 = \nabla^2 h(0), F_0 = \dfrac{\partial^2}{\partial q^2}\langle f\rangle(0, q^{(0)})\right)$. In the non-degenerate case one uses Morse inequalities to specify the number and spectral type of the critical points (or rather orbits of $\langle f\rangle$, that is, the spectral type of F_0. Then, adding the assumption of quasi-convexity, one may use the following elementary assertion.

Proposition. *Let A and B be two symmetric matrices, $A > 0$. Then, the spectrum of the product AB is real and of the same type as that of B, that is, it contains the same number of positive, zero and negative eigenvalues (including multiplicity).*

Indeed, if $P^2 = A$, $P > 0$, the spectrum of AB coincides with that of PBP, which is symmetric, and subspaces over which B is > 0 (respectively, $= 0$, < 0) are carried over by P into corresponding subspaces of PBP. □

To apply this proposition, one considers the orthogonal complement of ω_0, so only *quasi*-convexity is required. The upshot is that even at the perturbative level, only in the quasi-convex situation can one predict the stability of (at least some of) the periodic orbits which are born from a periodic torus.

The theorem above applies when $0 < \varepsilon \leqslant \varepsilon_0 = \varepsilon_0(h, T)$. It took a century to prove that one may at least partially remove the dependence of ε_0 on T, and this again in the quasi-convex situation only. Loosely speaking, Bernstein and Katok proved (in [8]) that *if h is quasi-convex*, for ε small enough, *independently of T*, there survive at least n closed orbits in a $O(\varepsilon^{1/3})$ neighbourhood of a torus of period T. This is a deeper result, strongly connected with the multidimensional version of "Poincaré's last geometric theorem", as proved by Conley and Zehnder (in [14]). This is also the first step in trying to understand what happens under perturbation, when a sequence of rational tori accumulates to a given limiting torus (in the unperturbed situation), that is, in trying to generalize the Aubry–Mather theory of "cantori" to higher dimensions. Once more, all this requires quasi-convexity, not only because the methods are often variational, but also because many "wild" phenomena seem to be liable to occur otherwise (see again [28]). Of course, in order to prove stability results over exponentially long times, we only had to use some simple arithmetics related to these results; we shall go deeper into the arithmetics in the next section (Chapter V, §1).

CHAPTER V

ROBUST TORI; ARNOL'D DIFFUSION

§1. Robust tori and "renormalization"

We do not know precisely how simultaneous approximation can be used to prove KAM type results, although this is certainly possible. Note that such a method would be closer to the ideas of "renormalization" and especially the original intuition of Greene in [26]. This could be useful in several respects, for example, for the study of lower-dimensional invariant tori. The only thing we wish to mention in this direction is a simple proposition which makes more precise the convergence of the time averages to the space average for linear flows on the torus. It should probably be used, in some form at least, on the way towards KAM type results via simultaneous approximation.

Let $\mathcal{A}_\rho$ denote the space of functions on $\mathbb{T}^n$ which extend analytically to the strip $|\operatorname{Im} q| < \rho$ and are continuous at the boundary; $\mathcal{A}_\rho$ is provided with the norm $\|\cdot\|_\rho$ of the maximum over the closed strip. On the other hand, let $\omega \in \mathbb{R}^n$, let $(T_j)_{j\geqslant 0}$ be the sequence of its period, and $(\omega_j)_{j\geqslant 0}$ the corresponding sequence of best approximations (ω_j has period T_j; see Appendix 1). We denote $\eta_j = \|\omega_j - \omega\|$ and one has the estimate

$$\eta_j \leq \frac{\sqrt{n}}{T_j T_{j+1}^{\frac{1}{n}}} < \frac{\sqrt{n}}{T_j^{1+\frac{1}{n}}}. \tag{1}$$

Lastly, we introduce the operators M_j of time average along ω_j and M_∞ the space average; for a function $g(q)$ on the torus

$$M_j(g) = \frac{1}{T_j}\int_0^{T_j} g(q+\omega_j t)dt; \quad M_\infty(g) = \int_{\mathbb{T}^n} g(q)dq.$$

Then the following statement holds.

Proposition. *Assume that ω is Diophantine, more precisely:*

$$\text{(2)} \qquad \begin{array}{l} \exists \tau > n-1, \quad \gamma > 0 \quad \textit{such that} \quad \forall k \in \mathbb{Z}^n\backslash\{0\} \\ |\, \omega\cdot k \,| \geq \gamma |k|^{-\tau}, \ \textit{where} \ |k| = \sum_i |k_i|. \end{array}$$

For any $g \in \mathcal{A}_\rho$ and any δ, $0 < \delta \leqslant \rho$,

$$\text{(3)} \qquad \left\| M_j(g) - M_\infty(g) \right\|_{\rho-\delta} \leq 4^n \|g\|_\rho \delta^{-n} \exp\left[-\frac{\delta}{2}\left(\frac{\gamma}{\eta_j}\right)^{\frac{1}{\tau+1}}\right].$$

Note that the right-hand side may be estimated in terms of T_i only, using (1). To prove (3), let $g \in L^2(\mathbb{T}^n)$ with Fourier coefficients g_k, $k \in \mathbb{Z}^n$; $M_j(g)$ is the function whose only non-zero Fourier coefficients are equal to g_k, for the values of k satisfying $\omega_j \cdot k = 0$. This last relation implies that

$$\gamma |k|^{-\tau} \leq |\, \omega\cdot k \,| = \left|\, (\omega - \omega_j)\cdot k \,\right| \leq \|\omega - \omega_j\| \cdot \|k\| = \eta_j \|k\|.$$

Hence

$$\text{(4)} \qquad |k| \geq K_j \stackrel{\text{def}}{=} \left(\frac{\gamma}{\eta_j}\right)^{\frac{1}{\tau+1}},$$

where the inequality $|k| \geqslant \|k\|$ (Euclidean norm) has been used. On the other hand, the fact that $g \in \mathcal{A}_\rho$ provides the estimate

$$|g_k| \leq \|g\|_\rho e^{-2\pi\rho|k|},$$

which enables us to evaluate the tail of the Fourier series. Namely, if $K \in \mathbb{Z}_+$, one sets

$$g^{\geq K(q)} = \sum_{k, |k| \geq K} g_k e^{2\pi i (k,q)}.$$

Then (see [6], for example):

$$\text{(5)} \qquad \|g^{\geq K}\|_{\rho-\delta} \leq \|g\|_\rho \sum_{|k|\geq K} e^{-2\pi\delta|k|} \leq c_n \|g\|_\rho \delta^{-n} \exp\left(-\frac{\delta K}{2}\right);$$

one can take $c_n = 4^n$ (see [7]). Letting $K = K_j$, this implies (3), in view of (4). □

This rather elementary proposition is interesting in itself. It is however unsatisfactory because it uses an arithmetical condition of linear type (here (2), which could be generalized) whereas one would like to start from a hypothesis of simultaneous type, say $\omega \in \Omega_n(\tau)$ (see Appendix 1 for the definition). This is by no means fortuitous. Generally speaking, linear conditions which bound divisors from below allow one naturally to control the evolution of *functions*, whereas conditions from simultaneous approximation pertain more to the *trajectories* themselves. This can perhaps be thought of as an aspect of the Fourier ("wave-particle") duality. One may illustrate this further by studying directly the approximate recurrence times of a linear flow, starting from linear Diophantine conditions of type (2) as is done in [23]. Since simultaneous approximation of a vector is essentially equivalent to the distribution of these recurrence times (see Appendix 1), this is again a form of transfer principle.

In the remainder of this section we would like to pursue a purely arithmetical track; we shall prove no new result, but shall gather some facts and references which are little known and might be useful in a further exploration of some dynamical questions. We shall use some notions from algebraic number theory and accordingly refer the reader to any elementary book on this subject, for example [50]. For motivation we shall first formulate a conjecture. Let us consider a two-dimensional "standard" map, that is, a symplectic map from $\mathbb{T}^2 \times \mathbb{R}^2$ to itself defined by

$$f(\theta, r) = (\theta', r') = \left(\theta + r'(\mathrm{mod}\,\mathbb{Z}^2),\, r + \varepsilon\nabla\sigma(\theta)\right),$$
$$(\theta, r) \in \mathbb{T}^2 \times \mathbb{R}^2, \quad \sigma : \mathbb{T}^2 \to \mathbb{R} \ \text{analytic}.$$

When $\sigma = 0$ the tori $r = \omega =$ const are invariant under this transformation. Let $\mathcal{A}$ be a space of functions with prescribed analyticity widths and continuous at the boundary, with the associated sup norm $\|\cdot\|$. If $\omega \in \mathbb{R}^2$ we say that $\varepsilon_0 = \varepsilon_0(\omega)$ is the *break-up threshold* for the frequency ω if there persists an invariant torus with frequency ω (homotopic to $\mathbb{T}^2 \times \{0\} \subset \mathbb{T}^2 \times \mathbb{R}^2$) for any $\sigma \in \mathcal{A}$, $\|\sigma\| < \varepsilon_0$ and ε_0 is maximal with this property. On the basis of the one-dimensional evidence, one can expect the following.

Conjecture. *Consider the frequency vectors* $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ *such that* $(1, \omega_1, \omega_2)$ *is an integral basis (over* $\mathbb{Z}$*) of the cubic field* $\mathbb{Q}(\cos(2\pi/7))$; *then, among these, some define tori whose break-up thresholds are locally maximal in frequency — or action — space.*

We shall make this statement more precise below, so that it could be numerically supported or invalidated; it seems however very hard to prove or disprove analytically. The same assertion can be put forward if in the first component of the map f one replaces $\theta' = \theta + r'$ by $\theta' = \theta + Sr'$, where $S = \mathrm{diag}(1, -1)$ (2×2 matrix). The convexity hypothesis then no longer holds, so this lies outside the range of applicability of the results of this paper and, for example, of those of [8]. It should give rise to a more unstable situation, both from the standpoint of finite time stability and of the existence of the analogue of Aubry–Mather sets. These features could make the invariant tori more important dynamically and easier to locate numerically.

The above conjecture rests on arithmetics, not dynamics, and is linked with the search for the two-dimensional "golden vector", which we shall now summarize. Apart from being interesting for its own sake, we hope this will give an idea of how far one can go with approximation theory. We shall need the functions $\gamma_n(\alpha)$ and $\gamma^n(\alpha)$ introduced at the end of Appendix 1, and the Diophantine constant $\gamma_n = \gamma^n$. The results we shall present are too precise for simple transfer principles to hold, so the connection between linear and simultaneous approximations usually requires detailed work; there are even results which are proved in one of the two cases only.

It was long thought that the constant γ_n is determined by vectors α such that $(1, \alpha_1, ..., \alpha_n)$ form the basis of a real algebraic field of degree $n+1$. This would of course be very helpful in searching for the "worst" approximable vectors in dimension n. However, Szekeres recently presented numerical evidence strongly suggesting that this is true for $n = 2$ (the case $n = 1$ is well-known; see below) but wrong for $n = 3$ and perhaps for $n \geqslant 4$ (see [55]). In fact he performed delicate numerical computations which seem to imply that γ_3 cannot be approached using bases of quartic fields. If this turns out to be true, there is a *qualitative* difference between the cases $n = 2$ and $n > 2$; we shall henceforth restrict ourselves to the former case, which has been the object of many more studies than the higher dimensions. From the point of view of dynamical systems, one may remember that already at the level of arithmetics, little is known beyond problems with three independent frequencies, or rather two frequency ratios. This is also the lowest dimension in which Arnol'd diffusion may occur.

Before turning to the two-dimensional case, let us briefly recall some features of the well-known one-dimensional case. We need one elementary property of continued fractions: if $a = [a_1, ..., a_k, ...]$ and $b = [b_1, ..., b_k, ...]$ are the continued fraction expansions of the numbers a and b, these are called *equivalent* if, up to translation, the expansions coincide for k large enough; in other words, if there exist positive integers l and m such that $a_{l+i} = b_{m+i}$ for any positive i. One has the following elementary assertion.

Proposition (see [52], for example). *Two numbers a and b are equivalent if and only if there are integers p, q, r, s satisfying $ps-qr = \pm 1$ and*

$$b = \frac{pa+q}{ra+s}.$$

We now list the following properties (see again [52]):

i) $\gamma_1 = 1/\sqrt{5}$.

ii) γ_1 is not only an upper bound, but also a maximum, which is achieved for example by the golden number $\chi = (1+\sqrt{5})/2 = [1,1,1,\ldots](\gamma_1(\chi) = \gamma_1)$, or by $\chi' = \chi - 1$ (χ and $-\chi'$ are the roots of the equation $x^2 = x+1$).

iii) The numbers α such that $\gamma_1(\alpha) = \gamma_1$ are exactly those which are equivalent to the golden number χ.

iv) There is a gap in the Lagrange spectrum, that is, in the values of the function $\gamma_1(\alpha)$: indeed, for any number α which is *not* equivalent to χ,

$$\gamma_1(\alpha) \leq \frac{1}{\sqrt{8}} < \frac{1}{\sqrt{5}}.$$

In dimension 1 much more is known of course than what is listed above; we shall see that in dimension 2 much *less* is known (and in fact is true). First, one does not know how to prove that one may restrict consideration to cubic fields. Leaving this question aside, the following result holds.

Theorem. *Let* $\gamma_2' = \sup\{\gamma_2(\alpha)$, *where* $\alpha = (a, b) \in \mathbb{R}^2$, $(1, a, b)$ *is the basis of a cubic field*$\}$; *then* $\gamma_2' = 2/7$.

This was proved by Adams (in [1]), closing a list of works on the subject by several authors. Of course, $\gamma_2' \leqslant \gamma_2$ and equality is strongly expected to hold (and to be hard to prove); this is very well supported by numerical evidence (see [55]), in contrast to the parallel statement in dimension 3, as was mentioned above.

Let us now rephrase assertions ii) and iii) so that they can be generalized to higher dimensions. The maximum γ_1 is reached when $\chi \in \mathbb{Q}(\sqrt{5})$; it is easy to show that $\mathbb{Q}(\sqrt{5})$ is the quadratic field of *minimal discrimant*. Still more precisely, $(1, \chi)$ span over $\mathbb{Z}$ the ring of integers of this field (recall that if $d \in \mathbb{Z}$ is squarefree, the integers of $\mathbb{Q}(\sqrt{d})$ are given by $\mathbb{Z}+\dfrac{\sqrt{d}+1}{2}\,\mathbb{Z}$ if $d = 1 \pmod 4$ and $\mathbb{Z}+\sqrt{d}\,\mathbb{Z}$ if $d = 2$ or $3 \pmod 4$).

As for iii), if α is equivalent to χ, then using the proposition above we have $\alpha = (p\chi+q)/(r\chi+s)$, with $ps-qr = \pm 1$, which is equivalent to saying that $(p\chi+q, r\chi+s)$ is again an integral basis of $\mathbb{Q}(\sqrt{5})$; moreover, all the bases are of this form.

In dimension 2 one is thus led to investigate the real cubic fields of minimal discriminant. In fact (see [20] or [33]) one can restrict attention to *totally real* fields, that is, those such that the roots of a defining polynomial are all real. There is only one totally real field of minimal discriminant ($= 49$); it is $\mathbb{Q}(\xi)$, where ξ is a solution of

$$x^3 + x^2 - 2x - 1 = 0,$$

the three roots of which are $\xi = 2\cos(2\pi/7)$, $\xi' = 2\cos(4\pi/7)$ and $\xi'' = 2\cos(6\pi/7)$. Moreover, it happens that $(1, \xi, \xi^2)$ is an integral basis of $\mathbb{Q}(\xi)$ (this is far from obvious). The corresponding Diophantine constants, however, are not very close to $2/7$ (≈ 0.286); specifically, $\gamma_2(\xi, \xi^2) = \gamma^2(\xi, \xi^2) \approx 0.187$.

So in fact the upper bound $2/7$ is *not* reached in dimension 2. But more is known in this direction, which is connected with the conjecture made above. Let us start from the basis $(1, \xi, \xi^2)$, $\xi = 2\cos(2\pi/7)$, of the integers of $\mathbb{Q}(\xi)$. The other integral bases which include 1 have the form $(1, a, b)$, where

$$a = n_1 + p\xi + q\xi^2, \quad b = n_2 + r\xi + s\xi^2,$$

n_1, n_2, p, q, r, s in $\mathbb{Z}$, $ps-qr = \pm 1$; one may restrict oneself to the case $0 < a < b < 1$.

In [17] (which uses [15], [16]) Cusick describes a construction which allows one, given $\varepsilon > 0$, to determine p, q, r, s such that

$$\frac{2}{7} - \varepsilon < \gamma_2(a, b) = \gamma^2(a, b) < \frac{2}{7}$$

(the equality of the two constants is proved in [15]). This procedure is effective—and algorithmic—if one knows the continued fraction expansion of $\xi = 2\cos(2\pi/7)$ and if this satisfies some property which is generically true but which one does not know how to prove for this particular number.

As a last piece of information, one should mention *numerical* evidence which is the only two-dimensional analogue of property iv) above known to date: Szekeres carried out computations which indicate (see [55]) that if $\alpha = (a, b)$ and $(1, a, b)$ is *not* a basis (integral or not) of $\mathbb{Q}(\cos(2\pi/7))$, then $\gamma_2(\alpha) \leqslant 2/7-\delta$, where $\delta \approx 0.03$. If this is true, using the continued fraction expansion of ξ and the construction of Cusick mentioned above, it is easy to construct pairs (a, b) satisfying $\gamma_2(a, b) > 2/7-\delta$. These should correspond to local maxima of the function $\gamma_2(\alpha)$ and correspondingly to the locally most "robust" invariant tori. Of course, one could—and should—also consider *all* the integral bases of $\mathbb{Q}(\xi)$, given by the action of $Gl_3(\mathbb{Z})$ on the vector $(1, \xi, \xi^2)$.

We come to our last topic in this section, which is connected with a still very hypothetical "renormalization" theory. Again, we shall deal only with arithmetics, and in this respect simultaneous approximation is the only relevant concept. The most robust tori should correspond to the worst and most regularly (the two properties are intimately connected) approximable frequency vectors, which led to the statement of the conjecture made above. In the spirit of this article and of Greene's paper [26], which prompted the development of the renormalization "ideology" in one dimension, one may now enquire about the distribution of the closed orbits of long periods accumulating on an invariant torus. This seems out of reach for the moment, and we shall examine the much more modest problem of the location of the best approximations of a given vector. This corresponds to the distribution, in action or frequency space, of the periodic tori in the *unperturbed* integrable situation. *After* perturbation, essentially the only information we have is that some closed orbits corresponding to these tori will survive, however large the period (that is, however close we approach the unperturbed torus; see the end of Chapter IV, §4 and [8]).

Let $\alpha \in \mathbb{R}^n$, and let (α_i) be the sequence of its best approximations: $\alpha_i = p_i/q_i$, where $p_i \in \mathbb{Z}^n$, $q_i \in \mathbb{Z}_+$, (q_i) being the sequence of the periods of α. Is it possible that the sequences (q_i) and (α_i) exhibit some kind of self-similarity? The existence of a meaningful scaling transformation corresponding to a shift $i \to i+1$ on the indices is apparently subject to the existence of three quantities λ, ρ and θ, which we define as follows.

First let

$$\lambda = \lim_{i\to\infty} \frac{q_{i+1}}{q_i},$$

when this limit exists, of course. Note that we have an a priori estimate from below: $\lambda \geqslant 1+2^{-n-1}$; in two dimensions, this is improved to $\lambda \geqslant 1.270$ (>1.125, see [36]). The number λ governs the scaling of the time variable.

Second let

$$\rho = \lim_{i\to\infty} \frac{\|\alpha_i - \alpha\|}{\|\alpha_{i+1} - \alpha\|},$$

if again this exists. The number ρ governs the scaling property of space.

Of course, in the definition of λ and ρ one can generalize the shift $i \to i+1$ to the more general $i \to i+u$, $u \in \mathbb{Z}_+$ arbitrary; we have set $u = 1$ for notational simplicity. In one dimension, λ and ρ exists in particular (for some $u \in \mathbb{Z}_+$) for the quadratic irrationals. In the multidimensional case one must also take account of the angular variable.

Let

$$\theta_i = \frac{q_i\alpha - p_i}{\|q_i\alpha - p_i\|} \in S^{n-1}.$$

We define θ when $n = 2$; this could be generalized, but very little is known then. For $n = 2$, (θ_i) is a sequence of points on the circle S^1 and $\theta \in (-\pi, \pi)$ is defined as its rotation number, again if it exists. It is interesting to note that for $n = 1$ the sequence $(\theta_i) \subset S^0 = \{\pm 1\}$ is well known: it has the form $\theta_i = (-1)^i$, which simply means that the convergents of the continued fraction of an irrational number approximate the latter in turn from above and from below.

The optimistic guess is that λ, ρ and θ exist at least when $\alpha = (a, b)$ is such that $(1, a, b)$ is an integral basis of $\mathbb{Q}(\cos(2\pi/7))$, or more generally a basis (not necessarily integral) of a cubic real field. Little seems to be known in this direction; the reader interested in general results on the behaviour of best approximations is referred to [31], [32] and [37] among others. In [56] the authors construct vectors analogous to the Liouville numbers, whose best approximations can exhibit essentially any prescribed behaviour, however erratic.

There is one tool for studying approximation which we have not yet mentioned: multidimensional continued fractions. We shall say a few words about it, and this is more of a pretext to introduce references; some of them would be quite useful if one wants to study numerically the conjecture made above or related questions. We first recall that a continued fraction algorithm is a scheme which, given an n-vector, produces, as in the one-dimensional case, a string of digits from which one may construct a sequence of *rational* vectors which approximate the original vector with increasing precision. There are several requirements which may be asked from such an algorithm. They are discussed in [54], to which we refer for a simple and careful discussion of various possible schemes; the upshot is that all the requirements cannot be met simultaneously, so that, in contrast with the one-dimensional situation, there is no *optimal* algorithm.

Returning to "renormalization", the self-similarity properties of the best approximations are reflected in the periodicity properties of the multidimensional continued fraction, and it is tempting to ask which vectors correspond to periodic fractions. Here one comes across a deep arithmetical phenomenon: for vectors which are bases of number fields, all the continued fraction algorithms are linked with the search of the units of the field. By far the most favourable case arises when the group of units is monogenic (that is, generated, multiplicatively, by one element); but, by Dirichlet's theorem about the structure of the group of units of algebraic fields, this is the case for cubic fields if and only if the field is *not* totally real. In fact, the following stronger property holds: if a continued fraction algorithm yields *all* the best approximations, it can be periodic on the basis of a real number field *only* if this is quadratic or cubic and non-totally real. This result is due to Mahler, whose original article [39] we recommend, particularly to the amateur of algebraic number theory; the case of the field $\mathbb{Q}(\cos(2\pi/7))$ is treated there in detail. Note that one cannot hope to obtain *only* the best approximations; all algorithms also yield spurious approximations, which are not particularly good.

To summarize, the worst approximable vectors in dimension 2 correspond to integral bases of a totally real cubic field, whereas the vectors whose continued fractions are the most regular and the easiest to compute are associated with bases (not necessarily integral) of real, non-totally real, cubic fields, for instance $\mathbb{Q}(\sqrt[3]{m})$, $m \in \mathbb{Z}_+$; [19], [9] and [22] are devoted to the study of this last example.

One is then faced with an alternative: one can either weaken the requirements to be met by a continued fraction algorithm or modify the notion of best approximation. In the first direction, Szekeres has proposed in [54] (see also [18]) an algorithm which generalizes the classical Jacobi–Perron scheme. The last example which is examined in [54] is precisely that of the 2-fraction of (ξ, ξ^2) $(\xi = 2\cos(2\pi/7))$, of which the first 100,000 digits have been computed. This heavy computation made it possible to conjecture that there enter only 1's and 2's, and that the fraction is *almost periodic* in some precise sense; this was proved by Cusick in a long paper [18] which is entirely devoted to this fraction and is a real tour de force. Again we mention this partly because it indicates that all the arithmetical computations linked with the conjecture about robust tori are already available in the literature.

In the second direction, concerning the very definition of best approximation, one may note the elementary fact that in dimension 1 the approximations we have defined are approximations of the second kind: if $\alpha \in \mathbb{R}$, one minimizes $|q\alpha - p|$, $q \in \mathbb{Z}_+$, $p \in \mathbb{Z}$. Approximations of the first kind, which minimize

$|\alpha - p/q|$, are in some sense more natural, but they do not all appear among the convergents of α. In dimension n, approximating α means approximating the straight line in $\mathbb{R}^{n+1}$ directed along $(1, \alpha)$ by the integer lattice $\mathbf{Z}^{n+1}$; a natural way to evaluate the approxiamtion is, for example, through the Euclidean distance to this line. For all these questions, which give rise to many open problems, we refer in particular to [20], [9], [36] and [22]; again these articles examine in detail the case of cubic fields.

§2. Arnol'd diffusion

The purpose of this section is twofold: first we would like to elaborate a little on the heuristic picture of Arnol'd diffusion which emerges from the results of this paper. Then we shall develop a non-rigorous but suggestive argument in favour of showing why these results should be close to optimal, as far as the stability exponents are concerned. Throughout this section, we shall assume that the reader has a nodding acquaintance with the original note of Arnol'd ([3]). We note that the term "diffusion" is somewhat unfortunate (in [3] Arnol'd speaks of "topological instability"), because the phenomenon is probably too complex to be modelled rigorously by simple stochastic processes enjoying the Markov property; we shall however comply with the widely accepted terminology. We do not wish to discuss the difficulties linked with a rigorous treatment of Arnol'd diffusion, which are far from understood, and we shall stay at an essentially heuristic level. Let us only mention that in [21] an example is constructed which allows us to avoid the main difficulties (but not to solve them); so in this very particular case, the construction put forward in [3] has been made rigorous. Lastly, for a physical approach to the phenomenon and its physical relevance, we refer in particular to [12] and [13], where it is forcefully demonstrated why it should "look like" a diffusion process with a well-defined diffusion coefficient (see also the end of this section).

The results of Chapter III allow us to estimate the velocity of Arnol'd diffusion from above, that is to estimate from below the time needed to produce a drift of order 1 of the action variables. But our method also suggests a different picture of the phenomenon. Let us start again from the familiar Hamiltonian $H(p, q) = h(p) + \varepsilon f(p, q)$. Here we may as well assume that $h(p) = (1/2)p^2$, so that action and frequency spaces coincide; for other convex unperturbed Hamiltonians, the picture is only slightly distorted by the frequency map. The trellis formed by the rational planes in frequency space or resonance surfaces in action space naturally goes along with the use of linear approximation; it is termed "stochastic web" in the physics literature. On the other hand, simultaneous approximation is associated with the *rational lattice* formed by the rational points, properly weighted with their periods. A vivid illustration could be obtained in dimension 2 or 3 by plotting these points on a screen with a brightness or brilliance depending on the period.

The rational planes then appear as "caustics", since they are loci on which the "irrationality dimension" drops (see Lemma 3 of Chapter III).

In any case, the resonance trellis and the rational lattice are dual objects which carry essentially the same information. Using the latter, it seems that Arnol'd diffusion could be heuristically described as follows. One restricts attention to the action space (which coincides with frequency space when $h(p) = (1/2)p^2$) and uses only Theorem 1B with 1/2 in place of 1/3, that is: $\|p(0)\| \leqslant r_0\sqrt{\varepsilon}$ implies $\|p(t)\| \leqslant R_0\sqrt{\varepsilon}$ for $|t| \leqslant \mathcal{T}_0 \exp\left(\frac{\tau}{T\sqrt{\varepsilon}}\right)$ and ε small enough, where the origin $p = 0$ has been set at a point (a torus in phase space) of period T. Of course, we did not prove exactly this statement, but this section is non-rigorous anyway, and we are more interested in offering a simple picture which could perhaps be implemented numerically. Concerning the constants, one may set $m = M = E = 1$, and looking at formula (29) in Chapter II, one sees that for numerical purposes one may set for example $r_0 = 1$, $R_0 = 10$, whereas $\mathcal{T}_0$ and τ are relatively "small" constants, the exact value of which is not crucial.

Now let $p(0)$ be an arbitrary initial condition; draw a ball $B(p(0), \sqrt{\varepsilon})$ of radius $\sqrt{\varepsilon}$ around $p(0)$ ($r_0 = 1$) and look for a rational point p of minimal period T lying inside this ball. Note that this specifies a *unique* point because two points of the same period T are at least $1/T$ apart and $1/T > 2\sqrt{\varepsilon} = \operatorname{diam} B(p(0), \sqrt{\varepsilon})$. The point $p(t)$ is then a priori allowed to oscillate "randomly", with a speed of the order of $\sqrt{\varepsilon}$ inside $B(p, 10\sqrt{\varepsilon})$, the ball of radius $10\sqrt{\varepsilon}$ ($R_0 = 10$) around p, until the time $t_* = \mathcal{T}_0 \exp\left(\frac{\tau}{T\sqrt{\varepsilon}}\right)$ has elapsed (with an appropriate choice of $\mathcal{T}_0$ and τ, say of the order of 10^{-1}). However, when for some t', $0 < t' < t_*$, $p(t')$ lies within $\sqrt{\varepsilon}$ of some rational point p' of period T', one may apply the same rule with respect to p', starting at time t'. So one should compare t_* with $t' + t'_*$, where $t'_* = \mathcal{T}_0 \exp\left(\frac{\tau}{T\sqrt{\varepsilon}}\right)$, and possibly consider the ball $B(p', 10\sqrt{\varepsilon})$ for further reference (in fact, during some time, one should even consider the intersection of $B(p, 10\sqrt{\varepsilon})$and $B(p', 10\sqrt{\varepsilon})$, but since we are describing a qualitative and non-rigorous mechanism, this is probably unnecessarily elaborate). In other words, if $p(t')$ lies in the influence zone of some rational point with sufficiently small period, that is, long enough trapping time, one switches consideration to this point and starts the process afresh. It is now obvious that this will *always* happen for some $t' \leqslant t_*$, so the whole process is well-defined over any interval of time.

One could perhaps think of it as a ball (rather than just a point) trying to make its way amidst very "sticky" points (the rational points) or on the contrary a "Brownian" particle amidst sticky balls. One should however not be misled by this image and keep in mind that everything is time reversible.

A nice feature of this model is that because there is no integration to perform, it can be "speeded up", using a logarithmic timescale, with a particle moving exponentially fast trapped in a ball over a "time" of order $\frac{\tau}{T\sqrt{\varepsilon}}$.

Of course, this crude model also suffers from many defects, insufficiencies and oversimplifications, and we shall proceed to list but a few.

1. The model is so "universal" that it does not even depend on the exact form of the perturbation (!), the latter being only assumed to be "generic" in a vague sense. In fact, one works in action space only, which does not allow one to do justice to the complexity of the problem. This is in some sense tantamount to some kind of "random phase approximation", a device which is very common in physics and very difficult to justify or even express in a mathematically sensible way.

2. Related to the above is the fact that we consider the motion of the particle inside the ball as "random". Here we seemingly do not take advantage of one piece of information which is crucial in Nekhoroshev's proof; namely the oscillatory motion should, in the mean, be transverse to the resonant surfaces (see [43]). This is guaranteed by the convexity assumption, which, from that viewpoint, arises precisely as a strong transversality condition (which can be relaxed to a *weak* transversality condition, namely steepness). It is still unclear how Nekhoroshev's mechanism of "detuning", that is, drifting towards a non-resonant region, can be fitted into the picture; it does not even seem completely clear to what extent it reflects reality or only the proof method. Quite the opposite, it could somehow be built into the picture of the resonant surfaces appearing as "caustics" of the rational lattice.

3. Another piece of information which is conspicuously missing is the existence of the Kolmogorov tori (invariant tori of maximal dimension). Again these cannot be properly represented in action space since they are distorted with respect to some of the unperturbed tori (which appear as points in action space). It ought perhaps to be mentioned that the dynamical importance of the tori for systems with more than two degrees of freedom is difficult to assess. In other words, there could exist in some (many?) cases a typical "Nekhoroshev regime", that is, an interval of values of the small parameter ε for which "most" tori are destroyed but such that Nekhoroshev type estimates are still valid; that is, it may be that $0 \ll \varepsilon_K \ll \varepsilon_N \ll 1$, where, roughly speaking, ε_K and ε_N are the respective thresholds of validity for the KAM and exponential time theories.

Now Kolmogorov tori also tend to "trap" (but not "attract"; again everything is reversible) nearby trajectories over exponentially long times. This corresponds to the two-parameter Gevrey type estimates which were alluded to at the end of Chapter IV, §2. This points again to the kind of duality between very resonant and very non-resonant frequencies which was discussed above (see especially Chapter IV, §1) and also reflects the fact that "classical" perturbation theory, including estimates over exponential times, and KAM theory still do not fit quite well together. In our opinion this enhances still more the value of Kolmogorov's insight about the possible existence of invariant tori.

4. Also absent from the picture are other well-known objects, in particular the lower-dimensional invariant tori, down to the periodic orbits. All these objects, which of course can be represented only in the full phase space, form a set of zero Lebesgue measure, contrary to Kolmogorov tori, but do also correspond to some exponential time estimates in their vicinity, although these have never been worked out and some may be technically rather cumbersome to write down. Moreover, these tori go along with stable, central and unstable manifolds which also play a role, and indeed an important one, since they are the core of the original analysis of Arnol'd in [3]. At last, one should mention possible cantori (Aubry–Mather sets) about which little is known, only that one cannot draw a complete parallel with the two-dimensional situation and that convexity plays a prominent role (see the end of Chapter IV, §4).

The above remarks should have convinced the reader that our model can at best provide a crude "macroscopic" description of the phenomenon, one that may be implemented numerically. As such it may be of some value and one can hope to eventually extract some quantitative information from it, describing some features of this particular process. These could then be confronted with the results of the already existing numerical experiments on Arnol'd diffusion.

As far as a "microscopic" rigorous description of Arnol'd diffusion is concerned (or indeed a rigorous statistical description) the task looks rather formidable, and we make only one simple remark. In order to substantiate such a picture as the one described above, or even a small part of it, in a rigorous way, one would have to consider the linearly unstable closed orbits as the elementary bricks for the construction, instead of the transversally hyperbolic $(n-1)$-dimensional tori. Now suppose that $\omega \in \mathbb{R}^d$ $(1 < d \leqslant n-1)$ is Diophantine, say it satisfies

$$\exists \gamma > 0 \ \text{ such that } \ |\omega \cdot k| \geq \gamma |k|^{-d} \quad \forall k \in \mathbb{Z}^d \backslash \{0\}.$$

Then tori of dimension d and frequency ω which are transversally hyperbolic will in general exist for $0 < \varepsilon < \varepsilon_0 = \varepsilon_0(\gamma)$, and $\varepsilon_0(\gamma)$ tends to 0 together with γ. Instead, for $d = 1$, we have already mentioned (see Chapter IV, end of §4) that closed orbits of period T will in general exist for $\varepsilon < \varepsilon_0$ *independent* of T. Here T and γ play very similar roles, γ being connected with the approximate recurrence times of the linear flow directed along ω (see Appendix 1). The upshot is that generically, given ε, there will be instability regions where no torus survives, whereas this does not occur with closed orbits. This may be quite important in trying to construct transition chains in the sense of [3].

We now return to the more conventional picture of Arnol'd diffusion, using transversally hyperbolic ("whiskered") tori, with the purpose of examining the optimality of the stability exponent $a(n)$ found in Chapter III. The computation below can be read as an interpretation of certain reasonings of Chirikov, in particular in the last section of [12]. We consider the Hamiltonian

$$(1) \qquad H = \frac{1}{2}p^2 + \frac{1}{2}I^2 + \varepsilon(\cos q - 1)\big(1 + \mu F(\phi)\big),$$

where

$$(p,q) \in \mathbb{R} \times (\mathbb{R}/2\pi\mathbb{Z}), \;\; (I,\phi) \in \mathbb{R}^n \times (\mathbb{R}^n/2\pi\mathbb{Z})^n.$$

The number of dimensions is $N = n+1$ and the name of the variables has been modified in order to focus interest on a neighbourhood of the simple resonance surface $p = 0$. Moreover, ε and μ are perturbation parameters, and F is a real analytic function with analyticity width $\sigma > 0$; for convenience, we assume that it is even and write

$$F(\phi) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} f_k \cos(k \cdot \phi), \;\; \text{with} \;\; f_{-k} = f_k.$$

The Fourier coefficients satisfy the estimate $|f_k| = O(e^{-\sigma|k|})$ with $|k| = \Sigma|k_i|$; in view of possible numerical experiments, one can think of the following two examples: $f_k = e^{-\sigma|k|}$, in which case the series may be summed explicitly, and $f_k = \exp(-c\|k\|^2)$, ($\|\cdot\|$ is the Euclidean norm) which defines an entire function which is almost a theta function. We decompose H into $H = H_1 + H_2 + \varepsilon\mu\Phi(q, \varphi)$, where

$$H_1 = \frac{1}{2}p^2 + \varepsilon(\cos q - 1), \quad H_2 = \frac{1}{2}I^2.$$

When $\mu = 0$, $H_0 = H_1 + H_2$ has invariant tori of dimension $n = N-1$ defined by $p = q = 0$, $I = \omega$, $\varphi \in (\mathbb{R}/2\pi\mathbb{Z})^n$; we write $\omega = I$ only to underline that this is a frequency ($\omega = \nabla H_2$). With respect to Arnol'd's example in [3], the generalization consists in the fact that the dimension is arbitrary (so one can consider an autonomous Hamiltonian), but more crucially in that the perturbation term includes arbitrarily high harmonics. Hyperbolicity is absent when $\varepsilon = 0$, which points to the degeneracy of the problem, and we have introduced, as in [45] and [3], *two* parameters to get round a difficult singular perturbation problem (see a brief comment below); also, still as in [3], the perturbation vanishes on the tori, which are thus all conserved, a highly non-generic feature.

We shall now give in detail the computation of the Poincaré–Melnikov integrals for H, in which small divisors will arise. When $\mu \geqslant 0$ the stable $(+)$ and unstable $(-)$ manifolds of a torus are defined by equations of the form

$$H_1 = \Delta^{\pm}(q,p,I,\phi), \quad \frac{1}{2}I_j^2 = \frac{1}{2}\omega_j^2 + \Delta_j^{\pm}(q,p,I,\phi), \quad j = 1,\ldots,n.$$

$\Delta^{\pm}$ and $\Delta_j^{\pm}$ (here indices designate the components of a vector) vanish when $\mu = 0$; the classical Poincaré–Melnikov computation consists in evaluating—at least formally—these functions to the first order in μ. More precisely, let $(p(t), q(t))$ be the solution of the pendulum equation described by H_1, corresponding to the separatrix (say its upper branch) and such that $q(0) = \pi$:

$$q(t) = 4 \arctan e^{\tau}, \quad p(t) = \dot{q}(t) = \frac{2\sqrt{\varepsilon}}{\cosh \tau}, \quad \tau = \sqrt{\varepsilon} t.$$

One has

$$\frac{d}{dt}\left(\frac{1}{2} I_j^2\right) = -I_j \frac{\partial H}{\partial \phi_j} = \frac{1}{2} \mu \omega_j p^2 \frac{\partial F}{\partial \phi_j},$$

since $I_j = \omega_j$ and $H_1 = 0$ on the separatrix. The linear approximation consists in substituting the unperturbed trajectory for the perturbed one in the integration. One computes the differences $\Delta = \Delta^+ - \Delta^-$ and $\Delta_j = \Delta_j^+ - \Delta_j^-$ in the plane $q = \pi$, which we denote by $\delta = \delta H_1$ and δ_j to this approximation. These quantities represent, to the first order in μ, the distance of the projections of the stable and unstable manifolds in the planes (p, q) and (I_j, φ_j); they are functions of the initial angle $\varphi^{(0)}$ on the torus (see below), of its frequency ω, and of the parameters which describe the perturbation. We thus obtain the version of the Poincaré–Melnikov formula for this example; in particular,

$$\delta_j = \frac{1}{2} \mu \omega_j \int_{-\infty}^{+\infty} p^2 \frac{\partial F}{\partial \phi_j} \, dt. \tag{2}$$

For a harmonic $f_k \cos(k \cdot \varphi)$ of the perturbation, the contribution δ_j^k is given by

$$\delta_j^k = -\frac{1}{2} \mu f_k \omega_j k_j \int_{-\infty}^{+\infty} p^2 \sin(k \cdot \phi(t)) dt,$$

where $\varphi(t) = \varphi^{(0)} + \omega t$ describes the unperturbed trajectory. One finds that

$$\delta_j^k = -2\pi \mu f_k(\omega, k) \frac{\sin(k \cdot \phi^{(0)})}{\sinh\left(\frac{\pi}{2} \frac{\omega \cdot k}{\sqrt{\varepsilon}}\right)} \omega_j k_j. \tag{3}$$

After a summation over j, we obtain the contribution of harmonic k to δH_2, denoted as $\delta^k H_2$:

$$\delta^k H_2 = -2\pi \mu f_k(\omega, k)^2 \frac{\sin(k \cdot \phi^{(0)})}{\sinh\left(\frac{\pi}{2} \frac{\omega \cdot k}{\sqrt{\varepsilon}}\right)}. \tag{4}$$

In a similar way one can compute $\delta = \delta H_1$ and find that $\delta H_1 = -\delta H_2 (= -\Sigma_j \delta_j)$. This result is not surprising, given the decomposition of H, since the latter is invariant and $\varepsilon \mu \delta \Phi$ is negligible to the first order in μ (this term is in fact of a still higher order because Φ oscillates).

The reasoning then goes very roughly as follows (see [3]): in order to construct a "transition chain" between hyperbolic tori, one looks for heteroclinic intersections, here in the plane $q = \pi$, between the stable and unstable manifolds of two tori with respective frequencies $\omega^{(1)}$ and $\omega^{(2)}$. One must then solve the system

$$\Delta = \Delta H_1 = 0, \quad \frac{1}{2}\omega_j^{(1)^2} + \Delta_j^{(1)-} = \frac{1}{2}\omega_j^{(2)^2} + \Delta_j^{(2)+}, \quad j = 1, \ldots, n.$$

If the difference between the frequencies is small with respect to μ, the solvability of the system is equivalent, by the implicit function theorem, to that of the following linearized system, where δ and the δ_j's are computed at a common intermediate value ω lying between $\omega^{(1)}$ and $\omega^{(2)}$:

$$(5) \qquad \delta = 0, \quad \delta_j = \frac{1}{2}\omega_j^{(2)^2} - \frac{1}{2}\omega_j^{(1)^2}, \quad j = 1, \ldots, n.$$

However, in "real" problems there is a link between μ and ε, of the type $\mu = \varepsilon^p$ ($p \in \mathbb{Z}_+$) and the problem of justifying this linear computation is quite difficult; only the one-dimensional case has been investigated. To our knowledge, the most precise *formal* computations can be found in [45] (§225 et seq.), which we urge the reader to consult; in the introduction of [29] he will find a discussion of the many circumstances under which such a singular perturbation problem arises: in two words, it stems from the fact that integrable Hamiltonian systems contain no hyperbolicity, the latter being thus of the same size as the perturbation. A problem of this type but in *one* dimension was solved for the first time only quite recently, by Lazutkin and coworkers (see [38]). Here we shall simply point out this difficulty and go on with the linear computation.

Pursuing Arnol'd's reasoning, the maximal step in the transition chain, that is, $\|\omega^{(2)} - \omega^{(1)}\|$, is essentially determined, according to (5), by the size of the Poincaré–Melnikov integrals, and if one can neglect the time necessary to achieve one step, with respect to the number of steps, then the average velocity of Arnol'd diffusion, that is, the inverse of the time necessary to produce a drift of the action variables of the order of unity, will be approximately equal to that same number. The above heuristic considerations are summarized in the following "equation":

Average velocity of the diffusion $\approx$ splitting $\approx$ splitting in the variational approximation,

where the first "equality" sign is to be understood in the sense that the two quantities are of the same order of magnitude and the second in the sense of asymptotic expansions. Roughly speaking, the justification of the first is subject to the solution of hard geometrical problems, and that of the second poses hard analytical problems.

If one believes in this, one can estimate the average velocity and find, in the case of the Hamiltonian examined in [3], that it is of order $e^{-c/\sqrt{\varepsilon}}$. The situation for the Hamiltonian H above is complicated by the appearance, in formulae (3) and (4), of the expression $\omega \cdot k$, or rather the combination $(\omega \cdot k)/\sqrt{\varepsilon}$; this expresses an interlacing of hyperbolicity (heteroclinic manifolds), singular perturbation (factor $1/\sqrt{\varepsilon}$), and ellipticity ("small divisor" $\omega \cdot k$), this last ingredient being new, and of course generic. This factor may a priori assume essentially *any* value, and this seems to invalidate the variational computation à la Poincaré–Melnikov completely, even from a formal point of view. Indeed, for suitable values of the small divisor one may predict *any* value for the speed of diffusion, including one which contradicts Nekhoroshev type estimates.

Looking at the above a little more carefully, however, one sees that the factor $(\omega \cdot k)/\sqrt{\varepsilon}$ arises from the possible resonance of the frequency vector ω of the hyperbolic torus with the relative frequency $k/\sqrt{\varepsilon}$ of a harmonic in the perturbation. But recall that the frequency vector of the Hamiltonian H is really $(0, \omega) \in \mathbb{R}^N$, so a resonance of ω indicates a *double* resonance for H. Our prescription now will be to stay away as far as possible from these double resonances, which means assuming that ω is a Diophantine frequency; there are at least three good reasons for this.

First, in general, the only transversally hyperbolic tori to arise will be precisely those corresponding to Diophantine frequencies. Here, of course, this is a rather bad argument since *all* the frequencies coexist.

Second, the variational argument above is adapted to single resonances. If one stays on an r-fold resonance ($1 \leqslant r \leqslant N-1$), that is, if one examines transversally hyperbolic tori of dimension $d = N-r$, one must in principle compute an $r \times r$ matrix of functions.

Third, by the result of this paper, if we want to maximize the velocity of Arnol'd diffusion, we should stay away from resonances of high multiplicities. Indeed, if two points in action space with the same unperturbed energy lie on, say, two simply resonant surfaces which intersect on the unperturbed energy surface, to go from one to the other necessitates a passage through a double resonance, and this causes the first stability exponent to increase (by Corollaries 2 and 3), that is, the velocity of the drift to decrease.

With this in mind, let us look back at formula (3). Up to now, the only condition has been $\delta = \delta H_1 = 0$, which constrains the vector I (or ω) to move on a sphere ($\|I\| = \text{const}$). Now we set $\sin(k \cdot \varphi^{(0)}) = \pm 1$, because we maximize the splitting over the initial angle (note that when k tends to infinity, the condition $|\sin(k \cdot \varphi^{(0)})| \geqslant \text{const}$ partitions the torus into thinner and thinner strips). We add the condition that ω is Diophantine, that is, it satisfies the familiar inequalities

$$\exists \tau \geq n-1,\ \gamma > 0 \quad \text{such that} \quad |\omega \cdot k| \geq \gamma |k|^{-\tau}, \quad \forall k \in \mathbb{Z}^n \backslash \{0\}.$$

We can at last perform the following estimate; look at the harmonics k such that

$$\sqrt{\varepsilon} \ll |\omega \cdot k| \approx \frac{\gamma}{|k|^\tau} \ll 1;$$

assume that the amplitude $|f_k|$ is of order $e^{-\sigma|k|}$. According to formula (3), this harmonic will contribute to the splitting of the stable and unstable manifolds a term

$$\delta_j^k \approx \mu|k|^{-\tau} e^{-\sigma|k|} \exp\left[-\frac{\rho}{\sqrt{\varepsilon}|k|^\tau}\right],$$

where $\rho = \pi\gamma/2$. Let us now extremize (maximize) this quantity with respect to k, with ε held fixed. According to the considerations above, this will yield the maximal possible step in a transition chain and, roughly speaking, the maximal average velocity. One immediately finds that for $\varepsilon \ll 1$ one must have $|k| \approx \left(\frac{\rho\tau}{\sigma}\right)^{\frac{1}{\tau+1}} \varepsilon^{-\frac{1}{2(\tau+1)}}$, so

$$\delta_j^k \approx \mu\varepsilon^{\frac{\tau}{2(\tau+1)}} \exp\left(-\Xi\varepsilon^{-\frac{1}{2(\tau+1)}}\right) \tag{6}$$

for some constant Ξ (which can be computed easily). By far the most important feature of this formula is the exponent $\dfrac{1}{2(\tau+1)}$ of ε in the iterated exponential. In particular, setting $\tau = n$, one finds that the velocity decreases like $\exp(-c\varepsilon^{-\frac{1}{2N}})$ $(N = n+1)$, which is almost identical to the estimate *from above* found in Chapter III. To tell the truth, one could even set $\tau = n-1$ or rather use any $\tau > n-1$ to define a set of frequencies of relative measure 1, and this would result in a slight overestimate of the velocity, with respect to stability results; this is not really troublesome, in view of the many assumptions to which the above computation is subject.

In the same perspective, if one would try to perform a similar computation for r-fold resonances, we have already mentioned that one would need to compute an $r \times r$ matrix, and it is conceivable that from the point of view of the exponent this would result in a replacement of n by $d = N-r$ above, because of the Diophantine conditions on the frequency; we would recover in this way the stability exponents of Corollary 2. The fact is that stability exponents are very rough, hence very robust indicators which in particular are insensitive to many algebraic operations. In more intuitive terms, one can say that the more the dimension of the hyperbolic torus, the more important can be the resonance of the perturbation with the linear flow the torus carries, even if this resonance is assumed to be minimal, given the dimension.

We have already said that the reasoning above is closely connected with estimates that were performed by Chirikov in a rather different language. This led him to a prediction of the "diffusion coefficient", with an exponent again equal to $1/(2N)$, N being the number of independent frequencies. Although they are dimensionally different, the "diffusion coefficient" (which cannot really be defined in a rigorous way) and the "mean velocity" essentially point to the same thing. Chirikov's prediction has been checked numerically and the agreement seems good, given the great difficulty of the experiments (see [13]). Thanks to the results of the present paper, one can say that the stability results which estimate the speed of Arnol'd diffusion from above and the heuristic reasonings which provide rough estimates for it essentially touch each other, at least as far as the exponents in the iterated exponentials are concerned.

APPENDIX 1

SOME DIOPHANTINE APPROXIMATION

Here we briefly recall, for the convenience of the reader, some simple results from approximation theory, referring to the specialized monographs (especially [11] and [52]) and articles for more details and many more results.

Let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$; two important and related problems consist in studying the *simultaneous* and *linear* approximations of α. In the first case, one is interested in the sequence of numbers $\|q\alpha\|_{\mathbb{Z}}$, $q \in \mathbb{Z}$ (or $\mathbb{N}$), in the second, one considers the value of $|\alpha \cdot k|$, $k \in \mathbb{Z}^n$. We recall that we use the notation

$$\|\alpha\|_{\mathbb{Z}} = \inf_{\zeta \in \mathbb{Z}^n} \|\alpha - \zeta\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the sup norm (largest component). A more general problem is to consider the size of p linear forms on $\mathbb{R}^q$, and this viewpoint is technically very useful, even for proving results on simultaneous or linear approximations, but we shall not enter into this.

It is obviously equivalent, from the standpoint of simultaneous approximation, to approximate $\alpha \in \mathbb{R}^n$ or $(1, \alpha) \in \mathbb{R}^{n+1}$; this is why the simultaneous approximation of α corresponds to *inhomogeneous* linear approximation. One considers the values of $|k_0 + \alpha \cdot k|$, $k_0 \in \mathbb{Z}$, $k = (k_1, \ldots, k_n) \in \mathbb{Z}^n$, that is, the values of $\|\alpha \cdot k\|_{\mathbb{Z}}$. Also, to obtain good transfer properties, one should use $|k|_\infty = \sup_i |k_i|$ to estimate the size of k. One thus defines

$$\Omega_n(\tau,\gamma) = \left\{\alpha \in \mathbb{R}^n,\ \forall q \in \mathbb{N}\backslash\{0\},\ \|q\alpha\|_{\mathbb{Z}} \geq \left(\frac{\gamma}{q}\right)^{\frac{1}{n}(1+\tau)}\right\},$$

$$\Omega_n(\tau) = \bigcup_{\gamma>0} \Omega_n(\tau,\gamma);$$

$$\Omega^n(\tau,\gamma) = \left\{\alpha \in \mathbb{R}^n,\ \forall k \in \mathbb{Z}^n\backslash\{0\},\ \|\ \alpha\cdot k\ \|_{\mathbb{Z}} \geq \gamma |k|_\infty^{-n(1+\tau)}\right\},$$

$$\Omega^n(\tau) = \bigcup_{\gamma>0} \Omega^n(\tau,\gamma);$$

The inclusion properties

$$\Omega_n(\tau_1,\gamma) \subset \Omega_n(\tau_2,\gamma) \quad \text{for} \quad \tau_1 < \tau_2,$$
$$\Omega_n(\tau,\gamma_1) \subset \Omega_n(\tau,\gamma_2) \quad \text{for} \quad \gamma_1 > \gamma_2$$

are checked by inspection, along with the analogous properties for the sets $\Omega^n(\tau, \gamma)$.

We shall first present a few simple properties of the sets $\Omega_n(\tau, \gamma)$ and then state a transfer theorem which in particular allows one to translate these properties for $\Omega^n(\tau, \gamma)$.

A vector is said to be *badly approximable* if it belongs to $\Omega_n(0)$. It is not obvious that this set is not empty, but one may in fact exhibit explicit examples: any vector α such that $(1, \alpha_1, ..., \alpha_n)$ constitutes a basis of an algebraic field (of degree $n+1$) over $\mathbb{Q}$ is badly approximable; this simple proposition is the multidimensional version of Liouville's theorem about the approximation of algebraic numbers.

On the other hand, a vector is said to be *very well approximable* if it does not belong to the set $\Omega_n(\tau)$ for sufficiently small $\tau > 0$, which is the same as requiring the existence of $\varepsilon > 0$ such that the inequality $\|q\alpha\|_{\mathbb{Z}} \leqslant q^{-1/n-\varepsilon}$ has an *infinite* number of integer solutions. With these definitions, the following proposition holds.

Proposition. *Almost all vectors are neither badly nor very well approximable:*

$$\operatorname{mes} \Omega_n(0) = \operatorname{mes}\big(\mathbb{R}^n - \bigcap_{\tau>0} \Omega_n(\tau)\big) = 0.$$

This means that the estimate in Dirichlet's theorem is almost nowhere optimal (because mes $\Omega_n(0) = 0$), but that the exponent $1/n$ of $q^{-1/n}$ can be improved almost nowhere. This proposition is a special case of a basic result in the metric theory of approximation, which may be stated as follows.

Theorem (Khinchin). *Let $\varphi : \mathbb{N} \to \mathbb{R}^+$ be monotonically decreasing. If the sum $\sum_{q\geq 0}(\varphi(q))^n$ converges (respectively, diverges), the inequality $\|q\alpha\|_{\mathbb{Z}} \leqslant \varphi(q)$ possesses, for almost all α, a finite (respectively, infinite) number of integer solutions.*

This theorem may then be applied with $\varphi(q) = q^{-1/n}$ or even $\varphi(q) = (q \log q)^{-1/n}$ ($q > 1$, divergent case) or on the contrary $\varphi(q) = q^{-1/n-\varepsilon}$ ($\varepsilon > 0$) or $\varphi(q) = (q \log^2 q)^{-1/n}$ (convergent case). Note that the statement in the theorem is easy to prove in the convergent case, much less so in the divergent case. There is a more precise version in [52] (due to Schmidt). Note also that although $\Omega_n(0)$ has zero Lebesgue measure, its Hausdorff dimension is equal to n.

We may now introduce the positive valued functions τ_* and τ^* on $\mathbb{R}^n$ defined as

$$\tau_*(\alpha) = \inf\{\tau \geq 0 \quad \text{such that} \quad \alpha \in \Omega_n(\tau)\},$$
$$\tau^*(\alpha) = \inf\{\tau \geq 0 \quad \text{such that} \quad \alpha \in \Omega^n(\tau)\}.$$

In general, of course, these lower bounds are not reached, that is, α does not necessarily belong to $\Omega_n(\tau_*(\alpha))$ (or $\Omega^n(\tau^*(\alpha))$. One may restate the definition as

$$\tau_*(\alpha) = \inf\{\tau \geq 0 \quad \text{such that} \quad \|q\alpha\|_{\mathbb{Z}} \leq q^{-\frac{1}{n}(1+\tau)} \quad \text{has a } \textit{finite} \text{ number of integer solutions}\}.$$

A similar statement holds for τ^*. One has then the following transfer theorem.

Theorem. i) *For any* $\alpha \in \mathbb{R}^n$

$$n\tau^*(\alpha) \geq \tau_*(\alpha) \geq \frac{\tau^*(\alpha)}{(n-1)\tau^*(\alpha)+n}.$$

ii) $\Omega_n(0) = \Omega^n(0)$.

The metric theorem stated above implies that $\tau_*(\alpha) = 0$ almost everywhere, hence $\tau^*(\alpha) = 0$ almost everywhere. This theorem says in particular that the notion of being badly or very well approximable does not depend on the kind of approximation.

One advantage of simultaneous approximation is that one need only consider a *sequence* of numbers (indexed by the positive integers). In fact, one may often restrict oneself to the *best approximations*. Let us first introduce the corresponding *periods* (this last name is not standard terminology, but we do not know of any widely used term).

Definition. For any vector $\alpha \in \mathbb{R}^n$, its *periods* $(q_i)_{i \geqslant 0}$ are the positive integers such that

$$q_0 = 1 \text{ and } \forall q \in \mathbb{N},\ q < q_{i+1} \Rightarrow \|q\alpha\|_{\mathbb{Z}} \geq \|q_i\alpha\|_{\mathbb{Z}}.$$

One of the components of α may be rescaled to unity, so one may assume that $\alpha = (1, \alpha')$, $\alpha' \in \mathbb{R}^{n-1}$. For $n = 2$, the q_i's are the denominators of the convergents of the continued fraction of α'. In any dimension one may define integer vectors p_i such that

$$\|q_i\alpha\|_{\mathbb{Z}} = \|q_i\alpha - p_i\|_\infty.$$

The *rational vectors* $\alpha_i = p_i/q_i$, $p_i \in \mathbb{Z}^n$, $q_i \in \mathbb{N}$, with periods q_i, are called the *best* (Dirichlet) *approximations* of α.

Turning to dynamics for a short while, one may note that the periods are linked with the approximate recurrence times on the torus. In fact, let $\omega = (1, \omega') \in \mathbb{R}^n$ with associated periods (T_i), and let F^t be the linear flow on the n-dimensional torus along ω:

$$\phi \in \mathbb{T}^n \to F^t(\phi) = \phi + \omega t \in \mathbb{T}^n.$$

Consider $C(\delta)$, $0 < \delta < 1/2$, the cube centred at the origin with sides along the coordinate axes, of length 2δ (viewing $\mathbb{T}^n$ as $\mathbb{R}^n/\mathbb{Z}^n$); finally, let $T(\delta)$ be the return time into $C(\delta)$ for a trajectory starting at the origin. Then, apart from some trivial exceptions,

$$T(\delta) = T_i + O(\delta), \text{ where } i \text{ is the smallest index such that } \|T_i\omega\|_{\mathbb{Z}} \leq \delta.$$

This is in fact simply a dynamical restatement of the definition of the periods. In the present article, when studying canonical perturbation theory, we thus substitute the use of the periods for that of the small divisors, or the *recurrence times* for the *resonances*. Dirichlet's theorem asserts that $T(\delta) = O(\delta^{-n+1})$, which is almost obvious.

Coming back to arithmetics and $\alpha \in \mathbb{R}^n$, Dirichlet's theorem is also seen to be equivalent to the sequence of inequalities

$$\|q_i\alpha\|_{\mathbb{Z}} \leq q_{i+1}^{-\frac{1}{n}}, \quad i \geq 0.$$

We may now introduce yet another type of Diophantine conditions by defining

$$\Omega(\tau,\gamma) = \{\alpha \in \mathbb{R}^n, \forall_i \geq 0 \quad q_{i+1} \leq \gamma q_i^{1+\tau}\},$$
$$\Omega(\tau) = \bigcup_{\gamma>0} \Omega(\tau,\gamma).$$

We thus require that the sequence of the periods does not grow too fast. Imposing a *polynomial* bound is largely arbitrary, as is the case for the other Diophantine conditions. It is important to point out that the sets $\Omega(\tau, \gamma)$ do not really describe the rate of approximation, but rather its regularity. In particular, if $\alpha = (0, \alpha')$ with $0 \in \mathbb{R}^r$, $\alpha' \in \mathbb{R}^d$ $(d+r = n)$, then $\alpha \in \Omega(\tau, \gamma)$ if and only if $\alpha' \in \Omega(\tau, \gamma)$. This is why in some sense this set is independent of n. In other words, when a vector is resonant, a trajectory of the corresponding linear flow on the n-dimensional torus is not dense; rather, the torus is foliated into tori of lower dimension over which the trajectories are dense. Belonging to $\Omega(\tau, \gamma)$ or not depends only on the motion on these lower-dimensional tori. It is conceivable that this kind of arithmetical conditions turns out to be the most useful or natural under various circumstances.

Finally there are some straightforward inclusion relations which relate these and the previously defined sets. In particular,

$$\Omega_n(\tau,\gamma) \subset \Omega(\tau,\gamma^{-(1+\tau)}), \qquad \text{hence} \qquad \Omega_n(\tau) \subset \Omega(\tau).$$

To prove this, let $\alpha \in \Omega_n(\tau, \gamma)$; then, in particular, for any i

$$\left(\frac{\gamma}{q_i}\right)^{\frac{1}{n}(1+\tau)} \leq \|q_i\alpha\|_{\mathbb{Z}} \leq q_{i+1}^{-\frac{1}{n}},$$

which implies the above inclusion (as we have already observed, inequality on the right is equivalent to Dirichlet's theorem). We thus also obtain

$$\text{mes}(\mathbb{R}^n \setminus \cap_{\tau>0}\,\Omega(\tau)) = 0,$$

but in fact much more is true; as was noted above this statement holds with $\mathbb{R}^n$ replaced by any rational plane (the "resonant" planes of dynamics).

As a final topic, let us examine the set of badly approximable vectors a little more closely. We use the notation $\Omega_n = \Omega_n(0) = \Omega^n(0)$, $\Omega = \Omega(0)$. One has $\Omega_n \subset \Omega$, which means that the sequence of the periods of badly approximable vectors increases at most geometrically. More precisely, $\Omega_n(0, \gamma) \subset \Omega(0, \gamma^{-1})$ $(\gamma > 0)$, that is, when $\alpha \in \Omega_n(0, \gamma)$, the corresponding periods (q_i) satisfy $q_i \leqslant \gamma^{-i}$ $(q_0 = 1)$.

In the other direction, for *any* $\alpha \in \mathbb{R}^n$ this sequence increases *at least* geometrically (when it is finite, that is, when at least one of the components of the vector is irrational). If fact, for any α

$$\underline{\lim}_{i \to \infty} (q_i)^{\frac{1}{i}} \geq g_n > 0,$$

with an explicit estimate $g_n \geqslant 1+2^{-n-1}$ (see [36]). By analogy with the case of dimension 2 (or 1 depending on the terminology), one could say that badly approximable vectors are of constant type.

For these vectors one can define $\gamma_n(\alpha)$ and $\gamma^n(\alpha)$ as

$$\gamma_n(\alpha) = \underline{\lim}_{q \to \infty} q \|q\alpha\|^n_{\mathbb{Z}}, \quad \gamma^n(\alpha) = \underline{\lim}_{|k| \to \infty} |k|^n_\infty \| \ \alpha \cdot k \ \|_{\mathbb{Z}}.$$

True, these functions can be defined for any vector, but they vanish if it is not badly approximable. Equivalently:

$$\gamma_n(\alpha) = \inf \{\gamma > 0 \ \text{ such that } \ \|q\alpha\|_{\mathbb{Z}} \leq \left(\frac{\gamma}{q}\right)^{\frac{1}{n}}$$

$$\text{has an infinite number of integer solutions}\},$$

with a similar definition of $\gamma^n(\alpha)$.

The n-dimensional *Diophantine constants* are defined as

$$\gamma_n = \sup\{\gamma_n(\alpha), \alpha \in \mathbb{R}^n\}; \quad \gamma^n = \sup\{\gamma^n(\alpha), \alpha \in \mathbb{R}^n\}.$$

The following transfer theorem holds.

Theorem (Davenport; see [20]).

$$\gamma_n = \gamma^n.$$

This is a more subtle result than the transfer theorem quoted above, and its proof uses the characterization of the Diophantine constants in terms of geometry of numbers. Except in dimension 1, $\gamma_n(\alpha)$ and $\gamma^n(\alpha)$ are different in general. We may also define γ_n as

$$\gamma_n = \inf\{\gamma > 0 \ \text{ such that for any } \ \alpha \in \mathbb{R}^n, \ \|q\alpha\|_{\mathbb{Z}} \leq \left(\frac{\gamma}{q}\right)^{\frac{1}{n}}$$

$$\text{has an infinite number of integer solutions}\}.$$

From this definition and Dirichlet's theorem it follows that $\gamma_n \leqslant 1$, which may be improved to $\gamma_n \leqslant (n/(n+1))^n$ (see [11] or [52]), which again may be improved still further. Considerations from the geometry of numbers allow us to show that, in the other direction, $\gamma_n \geqslant 1/\Delta_{n+1}$, where Δ_{n+1} stands for the modulus of the smallest discriminant of a real algebraic field of degree $n+1$ (Furtwängler, 1928); one may also improve on this result (see [30] and [33]). Here, in Chapter IV, §5, some results about the two-dimensional Diophantine constant are discussed, in connection with dynamics.

APPENDIX 2

GEVREY ASYMPTOTIC EXPANSIONS

In this Appendix we shall first recall a few definitions and standard properties about Gevrey properties of formal series and the functions asymptotic to them, referring once and for all to [49] (among other sources, of course). One motivation is the hope that further investigations along these lines will result in a deeper understanding of the divergence of the series which occur in canonical perturbation theory. We start with the following definition.

Definition 1. A formal series $\Sigma a_n z^n \in \mathbb{C}[[z]]$ is said to be *Gevrey of order k* (k a real positive number) if there are constants $C > 0$ and $M > 0$ such that

$$\text{(1)} \qquad \text{for any } n \geq 0 \quad |a_n| \leq CM^n (n!)^{\frac{1}{k}}.$$

The algebra of these series, which is stable under derivation, is denoted by $\mathbb{C}[[z]]_k$. Of course, $\mathbb{C}[[z]]_k \subset \mathbb{C}[[z]]_{k'}$ if $k > k'$, and $\mathbb{C}[[z]]_\infty = \mathbb{C}\{z\}$, the algebra of convergent power series.

Definition 2. Let $f(z)$ be a function defined and analytic over an open sector S:

$$S = \{z \in \mathbb{C}, \quad \theta_0 < \arg z < \theta_1, \quad 0 < |z| < R\},$$

and let $\widehat{f} = \sum a_n z^n$ be a formal series. By definition, f admits $\widehat{f}$ as an *asymptotic expansion* at 0 over S if for any proper subsector S' of S (defined by inequalities $\theta_0 < \theta_0' < \arg z < \theta_1' < \theta_1$) and any $N \geqslant 0$

$$f(z) = \sum_{n=0}^{N} a_n z^n + z^N R_N(z),$$

where R_N is defined in S' and $R_N(z) \to 0$ when $z \to 0$ in S'.

Definition 3. A function f analytic in a sector S is said to be *Gevrey of order k* if it admits an asymptotic expansion $\widehat{f} = \sum a_n z^n$ at the origin, and for any proper subsector S' of S there are constants C' and M' (possibly dependent on S') such that

$$\text{(2)} \qquad \sup_{z \in S'} \left| \frac{f^{(n)}(z)}{n!} \right| \leq C' M'^n (n!)^{\frac{1}{k}}.$$

The algebra, stable under derivation, of these functions is denoted by $G_k(S)$. Of course, if $f \in G_k(S)$, then $f^{(n)}(0) = n!\, a_n$ and $\widehat{f} \in \mathbb{C}[[z]]_k$; moreover, by (2) and Taylor's formula,

$$\text{(3)} \qquad \forall n > 0, \quad \left| f(z) - \sum_{p=0}^{n-1} a_p z^p \right| \leq C' (M'|z|)^n (n!)^{\frac{1}{k}}.$$

Conversely, using Cauchy's formula, it is easy to prove that $f \in G_k(S)$ if (3) holds true over any proper subsector S' of S (and possibly different constants C' and M'). On the other hand, (3) implies the following proposition.

***Proposition* 1.** *Let $f \in G_k(S)$, $\widehat{f} = \sum a_n z^n$ its asymptotic expansion at the origin, and S' a proper subsector of S. There exist positive constants A, B, C such that*

$$\forall z \in S' \left| f(z) - \sum_{n \leq A|z|^{-k}} a_n z^n \right| \leq Be^{-\frac{C}{|z|^k}}. \tag{4}$$

To prove this, just minimize the right-hand side of (3) over n for fixed $|z|$, using Stirling's formula to estimate $n!$. The upshot is that using a least term cut-off, Gevrey asymptotic expansions naturally lead to exponentially small remainders with exponent k for functions in $G_k(S)$. Unfortunately, there is no simple converse to this proposition, that is, inequalities (4) are strictly weaker than (3), although they are essentially equivalent for "well-behaved" functions. This is the case in particular when all the terms of the expansion vanish: specifically, we say that $f \in G_k(S)$ is *flat* at the origin over S if $\widehat{f} = 0$, that is, if $a_n = 0$ for all $n \geqslant 0$. Then the following proposition holds.

***Proposition* 2.** *A function $f \in G_k(S)$ is flat if and only if it is exponentially decreasing, that is, if for any proper subsector S' there are positive constants B and C such that*

$$\forall z \in S' \left| f(z) \right| \leq Be^{-\frac{C}{|z|^k}}. \tag{5}$$

The function $\exp(-z^k)$ belongs to $G_k(S)$, where

$$S = \left\{ z \in \mathbb{C}, \quad -\frac{\pi}{2k} < \arg z < \frac{\pi}{2k} \right\}$$

(when $k < 1/2$ one should think of S as a region of the universal covering of $\mathbb{C}/\{0\}$). Moreover $\exp(-z^{-k})$ is flat at 0 over S, and by Proposition 2 it is some sense as large as possible there. More generally, one may consider the function $\exp[-(\lambda/z)^k]$, which is flat over a sector bisected by the ray directed along $\lambda \in \mathbb{C}$.

The last feature of Gevrey functions that we shall mention is a very important quasi-analytic property, which can be viewed as a version of the Phragmén–Lindelöf principle.

***Proposition* 3.** *Assume that $f \in G_k(S)$ is flat at 0 over S and angle(S) $> \pi/k$; then $f = 0$.*

Here angle(S) $= |\theta_1 - \theta_0|$ denotes the aperture of the sector. The function $\exp(-z^{-k})$ is thus flat over a sector which is as large as possible.

With the above definitions and properties in mind, let us return to canonical perturbation theory. The overall idea would be to clarify the Gevrey properties, if any, of the various series which appear classically in normal form theory. For instance, as in §2 of Chapter IV, one can look at the perturbation of harmonic oscillators (see (5) there) or at the elliptic fixed point problem (see (7), (8) there), adding a Diophantine condition ((6) there) on the frequency. One then builds the normalizing series, and by controlling its growth one can prove exponential estimates. This is a quantitative version of Birkhoff's original construction, leading to what we termed Gevrey type estimates. Why? Roughly speaking, one in fact proves inequalities as in (4) above, that is, a least term cut-off procedure yields an exponentially small remainder. Here the role of z is played of course by ε, and one confines attention to *real* values of this parameter, working with series which live in, say, $\mathbb{C}\{p, q\}[[\varepsilon]]$, that is, they are formal series in the perturbation parameter whose terms are analytic functions of the phase variables.

The message is simply that it might be interesting to dig a little deeper, exploring the analyticity properties in ε and perhaps proving inequalities as in (3) above, which, as we said, are the true signature of Gevrey functions and imply (4) but not conversely. The index k of the Gevrey spaces (for the time being at the level of (4)) and the exponent τ of the Diophantine condition are closely related. Indeed, one obtains more or less $k = 1/(2\tau)$ for the harmonic oscillator problem (or $k = 1/\tau$ if $\sqrt{\varepsilon}$ is taken to be the small parameter) and $k = 1/\tau$ in the elliptic fixed-point problem. In turn, $\tau \geqslant n-1$, where n is the number of dimensions, which is thus related to the divergence of the series. Note also that in normal form theory one deals with series in two essentially different ways, using either majorant series or an iterative procedure. The first method is better suited to prove Gevrey properties; unfortunately it is available only in comparatively simple situations.

In this paper we have produced exponentially small remainders in a very different way, without any arithmetical condition, but using the non-linearity or anharmonicity and approximation, instead of the usual resonant normal forms. We find roughly $a = 1/(2n)$ for the first stability exponent, so $a = k(\tau)$ if one sets $\tau = n$ in Gevrey type estimates. In other words, things happen as if the *non-linear* estimate (Nekhoroshev type) for the quasi-convex situation more or less coincides with the *linear* estimate (Gevrey type) in the best possible case, that is, for the worst approximable frequencies. This is certainly not obvious and not fortuitous, and requires perhaps further investigation. Is the stability exponent linked with some Gevrey index? Of what functions or series exactly? The fact is that in our proof we have bypassed the construction of the series completely and restricted ourselves to the treatment of a *one* frequency problem, supplemented with approximation. In this respect, one should write $a = 1/(2n) = 1/2 \times 1 \times 1/n$. The factor 1/2 comes simply from the fact that $\varepsilon^{1/2}$ and not ε is the natural small parameter; $1/n$ comes from Dirichlet's theorem, which corresponds to $\tau \geqslant n-1$ in terms of linear approximation; lastly, "1" comes from the 1 frequency problem and, simple as it looks, this is perhaps yet another interesting path to follow.

In its barest version, one starts from the equation

$$\frac{dx}{dt} = \varepsilon f(x,t) \tag{6}$$

with a scalar unknown x (one could take a vector as well) and a function f periodic in the time variable t, say of period 1. The function f has to be analytic in x and at least Lipschitz in t. The goal is to conjugate (6) with an autonomous problem

$$\frac{dy}{dt} = \varepsilon g(y, \varepsilon), \tag{7}$$

which is "integrable" (in the scalar case); this is effected via a change of variable

$$x = y + \varepsilon u(y, t, \varepsilon). \tag{8}$$

This is always possible *formally*, that is, there exists $\widehat{u} \in \mathbb{C}\{y, t\}[[\varepsilon]]$, which maps (6) into (7) with some $\widehat{g} \in \mathbb{C}\{y, t\}[[\varepsilon]]$. Moreover, one can obtain an exponentially small remainder by a least term cut-off procedure: there are *functions* $u(y, t, \varepsilon)$ and $g(y, \varepsilon)$ obtained as truncations of $\widehat{u}$ and $\widehat{g}$ such that (8) transforms (6) into

$$\frac{dy}{dt} = \varepsilon g(y, \varepsilon) + O(e^{-\frac{c}{\varepsilon}}),$$

where the estimate is for *real* ε only. This result was obtained explicitly by Neishtadt in [42]. It is in fact simpler than the iterative lemma in Chapter II above; there we had to deal with the facts that the problem is canonical, that there are other degrees of freedom, and that the angle one seeks to eliminate corresponds to a variable frequency. It resulted altogether in a loss in the exponent; for instance in Theorem 1B we find 1/3 instead of 1/2 (beware of the correspondence $\varepsilon \leftrightarrow \sqrt{\varepsilon}$ in the above setting). We believe that it would be interesting to investigate the Gevrey 1 properties of problem (6), (7), (8), which expresses a simple form of the occurrence of divergence *without* small divisors. It is quite possible that summability and resurgence (in the sense of Ecalle) will prove useful in this context.

Added in proof. Since this paper was written, a technical improvement of the iterative lemma enabled P. Lochak, A. Neishtadt and L. Niederman to reach the value $a = \dfrac{1}{2n}$ (the proof appears in the Proceedings of the 1991 Conference on Dynamical Systems of the Euler Institute in St Petersburg, published by Birkhaüser). On the other hand, in a recent preprint (to appear in Math. Zeitschrift), Pöschel found this same value for a, by improving in the convex case the original method of Nekhoroshev.

References

[1] W.W. Adams, The best two-dimensional Diophantine approximation constant for cubic irrationals, Pacific J. Math. **91** (1980), 29–30. MR **82i**:10038.

[2] V.I. Arnol'd, Small denominators and problems of stability of motion in classical and celestial mechanics, Uspekhi Mat. Nauk **18**:6 (1963), 91–192. MR **30** # 943.
= Russian Math. Surveys **18**:6 (1963), 91–191.

[3] ———, Instability of dynamical systems with several degrees of freedom, Dokl. Akad. Nauk **156** (1964), 9–12. MR **29** # 329.
= Soviet Math. Dokl. **5** (1964), 581–585.

[4] ———, *Matematicheskie metody klassicheskoi mekhaniki,* 3rd ed., Nauka, Moscow 1989.
English translation of 1st ed.: Mathematical methods of classical mechanics, Springer-Verlag, Berlin 1978.

[5] G.E. Backus, Critique of "The resonant structure of the solar system", by A.M. Molchanov, Icarus **11** (1969), 88–92.

[6] G. Benettin, L. Galgani, and A. Giorgilli, A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems, Celestial Mech. **37** (1985), 1–25. MR **87i**:58051.

[7] ——— and G. Gallavotti, Stability of motions near resonances in quasi-integrable Hamiltonian systems, J. Statist. Phys. **44** (1986), 293–338. MR **88h**:58042.

[8] D. Bernstein and A. Katok, Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonians, Invent. Math. **88** (1987), 225–241. MR **88i**:58048.

[9] A. Brentjes, A two-dimensional continued fraction algorithm for best approximations with an application in cubic number fields, J. Reine Angew. Math. **326** (1981), 18–44. MR **83b**:10037.

[10] H.W. Broer, G.B. Huitema, F. Takens, and B.L. Braaksma, Unfoldings and bifurcations of quasi-periodic tori, Mem. Amer. Math. Soc. **83** (1990), no. 421. MR **91e**:58156.

[11] J.W.S. Cassels, An introduction to Diophantine approximation, Cambridge Tracts in Mathematics and Mathematical Physics 45, Cambridge Univ. Press, New York 1957. MR **19**–396.

[12] B.V. Chirikov, A universal instability of many-dimensional oscillator systems, Phys. Rep. **52** (1979), 263–279. MR **80h**:70022.

[13] ——— and V.V. Vecheslavov, KAM integrability, in: Analysis etc., Academic Press, Boston, MA 1990, pp. 219–236. MR **91h**:58039.

[14] C.C. Conley and E. Zehnder, The Birkhoff–Lewis fixed point theorem and a conjecture of V.I. Arnol'd, Invent. Math. **73** (1983), 33–49. MR **91h**:58039.

[15] T.W. Cusick, Formulas for some Diophantine approximation constants, Math. Ann. **197** (1972), 182–188. MR **47** # 142.

[16] ———, Formulas for some Diophantine approximation constants. II, Acta Arith. **26** (1974/75), 117–128. MR **50** # 7041.

[17] ———, The two-dimensional Diophantine approximation constant, Monatsh. Math. **78** (1974), 297–304. MR **50** # 7042.

[18] ———, The Szekeres multidimensional continued fraction, Math. Comp. **31** (1977), 280–317. MR **55** # 2775.

[19] ———, Best Diophantine approximations for ternary linear forms, J. Reine Angew. Math. **315** (1980), 40–52.

[20] H. Davenport, On a theorem of Furtwängler, J. London Math. Soc. **30** (1955), 186–195. MR **16**–803.

[21] R. Douardy, Stabilité ou instabilité des points fixes elliptiques, Ann. Sci. École Norm. Sup. (4) **21** (1988), 1–46. MR **89m**:58113.

[22] E. Dubois and G. Rhin, Meilleures approximations d'une forme linéaire cubique, Acta Arith. **40** (1982), 197–208. MR **83f**:10039.

[23] H.S. Dumas, Ergodization rates for linear flow on the torus, J. Dynamics Differential Equations **3** (1991), 593–610. MR **92g**:58109.

[24] A. Giorgilli, Rigorous results on the power expansions for the integrals of a Hamiltonian system near an elliptic equilibrium point, Ann. Inst. H. Poincaré Phys. Théor. **48** (1988), 423–439. MR **89k**:58087.

[25] ———, 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. Differential Equations **77** (1989), 167–198. MR **90c**:70026.

[26] J.M. Greene, A method for determining a stochastic transition, J. Math. Phys. **20** (1979), 1183–1201.

[27] M. Hénon, A comment on "The resonant structure of the solar system", by A.M. Molchanov, Icarus **11** (1969), 93–94.

[28] M.R. Herman, Inégalités "a priori" pour des tores Lagrangiens invariants par des diffémorphismes symplectiques, Publ. Math. IHES **70** (1989), 47–101.

[29] P.J. Holmes, J.E. Marsden, and J. Scheurle, Exponentially small splittings of separatrices with applications to KAM theory and degenerate bifurcations, in: Contemp. Math. 81, Amer. Math. Soc., Providence, RI 1988, pp. 213–244. MR **90b**:58217.

[30] Yu.S. Il'yashenko. A steepness test for analytic functions, Uspekhi Mat. Nauk **41**:1 (1986), 193–194. MR **87i**:58155.
= Russian Math. Surveys **41**:1 (1986), 229–230.

[31] W.B. Jurkat, W. Kratz, and A. Peyerimhoff, Explicit representations of Dirichlet approximations, Math. Ann. **228** (1977), 11–25. MR **57** # 5913.

[32] ———, ———, and ———, On best two-dimensional Dirichlet approximations and their algorithmic calculation, Math. Ann. **244** (1979), 1–32.

[33] S. Krass, Estimates for n-dimensional Diophantine approximation constants for $n \geqslant 4$, J. Number Theory **20** (1985), 172–176. MR **86j**:11070.

[34] S.B. Kuksin, On the interpolation of an analytic symplectomorphism by a Hamiltonian flow, to appear in Russian J. Pure Appl. Math.

[35] ———, Perturation theory for quasi-periodic solutions of infinite-dimensional Hamiltonian systems and its application to the Korteweg-de Vries equation, Mat. Sb. **136** (1988), 396–412. MR **89m**:58096.
= Math. USSR-Sb. **64** (1989), 397–413.

[36] J.C. Lagarias, Best simultaneous Dirichlet approximations. I. Growth rates of best approximation denominators, Trans. Amer. Math. Soc. **272** (1982), 545–554. MR **84d**:10039a.

[37] ———, Best simultaneous Dirichlet approximations. II. Behaviour of consecutive best approximations, Pacific J. Math. **102** (1982), 61–88. MR **84d**:10039b.

[38] V.F. Lazutkin, I.G. Schachmannski, and M.B. Tabanov, Splitting of separatrices for standard and semistandard mappings, Phys. D **40** (1989), 235–248. MR **91b**:58175.

[39] K. Mahler, Über die Annäherung algebraischen Zahlen durch periodische Algorithmen, Acta Math. **68** (1937), 109–144.

[40] A.M. Molchanov, The resonant structure of the solar system, Icarus **8** (1968), 203–215.

[41] ———, Resonances in complex systems. A reply to critiques, Icarus **11** (1969), 95–103.

[42] A.I. Neishtadt, On the separation of motions in systems with a rapidly rotating phase, Prikl. Mat. Mekh. **48** (1984), 197–204. MR **86j**:34043.
= J. Appl. Math. Mech. **48** (1984), 133–139.

[43] N.N. Nekhoroshev, An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems, Uspekhi Mat. Nauk **32**:6 (1977), 5–66. MR **58** # 18570.
= Russian Math. Surveys **32**:6 (1977), 1–65.

[44] ———, An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. II, Trudy Sem. Petrovsk. No. 5 (1979), 5–50. MR **80m**:58017.

[45] H. Poincaré, Les méthodes nouvelles de la mécanique céleste, Gauthier-Villars, Paris 1892, 1899.

[46] J. Pöschel, Integrability of Hamiltonian systems on Cantor sets, Comm. Pure Appl. Math. **35** (1982), 653–696. MR **84d**:58039.

[47] ———, On elliptic lower-dimensional tori in Hamiltonian systems, Math. Z. **202** (1989), 559–608. MR **91a**:58065.

[48] ———, Small divisors with spatial structure in infinite-dimensional Hamiltonian systems, Comm. Math. Phys. **127** (1990), 351–393. MR **91i**:58047.

[49] J.P. Ramis, Théorèmes d'indices Gevrey pour les équations différentielles ordinaires, Mem. Amer. Math. Soc. **48** (1984), no. 296. MR **86e**:34021.

[50] P. Samuel, Théorie algébrique des nombres, Hermann, Paris 1967. MR **35** # 6643.

[51] J. Scheurle, Quasi-periodic solutions of the plane three-body problem near Euler's orbits, Celestial Mech. **28** (1982), 141–151. MR **84g**:70008.

[52] W.M. Schmidt, Diophantine approximation, Lecture Notes in Math. **785** (1980). MR **81j**:10038.

[53] M.B. Sevryuk, Lower-dimensional tori in reversible systems, Chaos **1** (1991), 160–167. MR **93c**:58184.

[54] G. Szekeres, Multidimensional continued fractions, Ann. Univ. Sci. Budapest Eötvös Sect. Math. **13** (1970), 113–140. MR **47** # 2753.

[55] ———, Search for the three-dimensional approxiamtion constant, London Math. Soc. Lecture Note Ser. **109** (1986), 139–146. MR **88b**:11041.

[56] ——— and V.T. Sós, Rational approximation vectors, Acta Arith. **49** (1988), 255–261. MR **89c**:11106.

[57] D.V. Treshchëv, The mechanism of destruction of resonance tori of Hamiltonian systems, Mat. Sb. **180** (1989), 1325–1346. MR **91i**:58124.
= Math. USSR-Sb. **68** (1991), 181–203.

[58] G.M. Zaslavskii, R.Z. Sagdeev, D.A. Usikov, and A.A. Chernikov, Minimal chaos, Stochastic webs, and structures of quasicrystal symmetry, Uspekhi Fiz. Nauk **156** (1988), 193–251. MR **90f**:58124.
= Soviet Phys. Uspekhi **31** (1989), 887–915.

École Normale Supérieure, Paris

Received by the Editors 20 April 1992