Journal of Statistical Physics, Vol. 44, Nos. 3/4, 1986

Stability of Motions near Resonances in Quasi-Integrable Hamiltonian Systems

Giancarlo Benettin[1,2] **and Giovanni Gallavotti**[1,3]

Received November 12, 1985

Nekhoroshev's theorem on the stability of motions in quasi-integrable Hamiltonian systems is revisited. At variance with the proofs already available in the literature, we explicitly consider the case of weakly perturbed harmonic oscillators; furthermore we prove the confinement of orbits in resonant regions, in the general case of nonisochronous systems, by using the elementary idea of energy conservation instead of more complicated mechanisms. An application of Nekhoroshev's theorem to the study of perturbed motions inside resonances is also provided.

KEY WORDS: Resonances; perturbation theory; KAM theory; Hamiltonian systems.

1. INTRODUCTION

The aim of this paper is to revisit Nekhoroshev's theorem[1] on the stability of motions in nearly integrable Hamiltonian Systems.

As is well known, this theorem deals with hamiltonian systems which, in action-angle variables $(\underline{A}, \underline{\phi}) = (A_1,..., A_l, \phi_1,..., \phi_l)$, have the form

$$H_\varepsilon(\underline{A}, \underline{\phi}) = h(\underline{A}) + \varepsilon f(\underline{A}, \underline{\phi}) \tag{1.1}$$

[1] Rutgers Hill Center, New Brunswick, New Jersey 08903.

[2] Partially supported by "Ministero della Pubblica Istruzione." Permanent address: Dipartimento di Fisica dell'Università di Padova e Centro Interuniversitario di Struttura della Materia, Via F. Marzolo 8–35131 Padova, Italy.

[3] Partially supported by Grant N.S.F. DMS 85-03333 and by "Ministero della Pubblica Istruzione." Permanent address: Dipartimento di Matematica, II Università di Roma, Via O. Raimondo, 00173 Rome, Italy.

0022-4715/86/0800-0293$05.00/0 © 1986 Plenum Publishing Corporation

Under quite mild assumptions (essentially a geometric property of $h(\underline{A})$ weaker than convexity, called "steepness") the theorem provides a uniform bound on the variations of action variables of the form

$$|A_j(t) - A_j(0)| \leqslant A^* \varepsilon^a \qquad j = 1,..., l$$
$$\text{for} \quad |t| \leqslant T^* \exp\left(\frac{1}{\varepsilon}\right)^b \tag{1.2}$$

A^*, a, T^*, b being positive constants.

In the present paper we present a proof of this theorem, somewhat different from the original proof by Nekhoroshev and from the more recent proof in Ref. 2. First, we consider explicitly the case of isochronous systems (i.e., weakly coupled harmonic oscillators), which do not verify the steepness hypothesis of Nekhoroshev. Actually, in our opinion, the case of isochronous systems is the simplest and transparent application of the basic ideas behind Nekhoroshev's work, because all complications of geometric nature are absent.

Then, when analyzing nonisochronous systems (we treat, as in Ref. 2, the simpler case of convex unperturbed Hamiltonians) we use as the basic tool the elementary idea of the conservation of energy in dealing with the part of the proof that we call "geometric."

This procedure simplifies the proof, at least conceptually, and in our opinion allows one to make some quantitative statements on the properties of the perturbed motion up to times exponentially long in an inverse power of ε, exhibiting in this way some use of the Nekhoroshev theorem.

The present paper is organized as follows: in Section 2 we describe precisely the class of Hamiltonians we are concerned with, and state our results for the case of weakly coupled harmonic oscillators. In Section 3 we state our main result for the anharmonic case, while in Sections 4 and 5 some corollaries are presented; in particular, the different characters of motions on different time scales, for initial data inside "resonance regions," are there analyzed. Sections 6, 7, and 8 are devoted to the proofs of the main theorems (the harmonic case, and the "analytic" and "geometric" parts of the anharmonic case respectively). Finally, Section 9 contains some lemmas, while the concluding remarks are reported in Section 10. An appendix follows, where the problem of chaotic motions (in particular, homoclinic phenomena) inside resonances is discussed on the basis of a simple example.

Some of these results have already been reported by one of us at the 1984 *Les Houches* summer school.[3]

2. INTEGRABLE SYSTEMS AND STABILITY PROBLEMS. RESULTS FOR THE HARMONIC OSCILLATORS CASE

Consider an l-degrees of freedom Hamiltonian system with Hamiltonian

$$H_\varepsilon(\underline{A}, \underline{\varphi}) = h(\underline{A}) + \varepsilon f(\underline{A}, \underline{\varphi}) \tag{2.1}$$

where $(\underline{A}, \underline{\varphi}) = (A_1, \ldots, A_l, \varphi_1, \ldots, \varphi_l)$ vary in the phase space $V_R \times \mathbf{T}^l$ with

$$\begin{aligned} V_R &= \{\underline{A} = (A_1, \ldots, A_l) \in \mathbf{R}^l;\ |A_j| \leqslant R,\ j = 1, \ldots, l\} \\ \mathbf{T}^l &= l\text{-dimensional torus} \end{aligned} \tag{2.2}$$

We suppose that h and f are real analytic on $V_R \times \mathbf{T}^l$ with analyticity parameters ρ, ξ, $0 < \rho < R$, $0 < \xi < 1$. This means that h, f are regarded as functions on $\mathbf{R}^{2l}$, periodic in φ with period 2π (h is in fact φ-independent), and furthermore, they admit a holomorphic extension to the complex domain $W(V_R; \rho, \xi)$, where for a generic subset $V \subset \mathbf{R}^l$ we denote

$$\begin{aligned} W(V; \rho, \xi) = \{\underline{A}, \underline{\varphi} \mid (\underline{A}, \underline{\varphi}) \in \mathbf{C}^{2l};\ &\mathrm{dist}(\underline{A}, V) \leqslant \rho; \\ &|\mathrm{Im}\ \varphi_j| \leqslant \xi,\ j = 1, \ldots, l\} \end{aligned} \tag{2.3}$$

with $\mathrm{dist}(\underline{A}, V) = \inf_{\underline{A}' \in V} \|\underline{A}' - \underline{A}\|$, $\|\underline{A}' - \underline{A}\| = \max_{1 \leqslant j \leqslant l} |A_j' - A_j|$. For any function $\underline{g}: V_R \times \mathbf{T}^l \to \mathbf{C}^q$, real analytic on $V_R \times \mathbf{T}^l$ with analyticity parameters ρ, ξ we define, if $W \equiv W(V_R; \rho, \xi)$:

$$\|\underline{g}\|_W = \sup_{(\underline{A}, \underline{\varphi}) \in W}\ \sup_{j = 1, \ldots, q} |g_j(\underline{A}, \underline{\varphi})| \tag{2.4}$$

For our Hamiltonian we may and shall assume, without loss of generality and to avoid introducing too many constants, that the "size" of h and of f are equal:

$$\left\| \frac{\partial h}{\partial \underline{A}} \right\|_W = \max \left(\left\| \frac{\partial f}{\partial \underline{A}} \right\|_W, \frac{1}{\rho} \left\| \frac{\partial f}{\partial \underline{\varphi}} \right\|_W \right) \equiv E < \infty \tag{2.5}$$

where $\partial \cdot / \partial \underline{A} \equiv (\partial \cdot / \partial A_1, \ldots, \partial \cdot / \partial A_l)$ and $\partial \cdot / \partial \underline{\varphi} \equiv (\partial \cdot / \partial \varphi_1, \ldots, \partial \cdot / \partial \varphi_l)$. In such a way $\varepsilon = 1$ means that the time scales associated with the "free part" h of the Hamiltonian and with the "interaction part" εf are equal. For an integer vector $\underline{v} = (v_1, \ldots, v_l)$, we shall always denote

$$|\underline{v}| = \sum_{j=1}^{l} |v_j| \tag{2.6}$$

In this paper we shall consider two extreme cases: the first is

$$h(\underline{A}) = \underline{\omega} \cdot \underline{A} \tag{2.7}$$

where the constant "angular velocities" $\underline{\omega} = (\omega_1, ..., \omega_l) \in \mathbf{R}^l$ obey the diophantine condition

$$|\underline{\omega} \cdot \underline{\nu}|^{-1} \leqslant C|\underline{\nu}|^l \qquad \text{for some } C > 0 \text{ and for all } \underline{\nu} \neq \underline{0},\ \underline{\nu} \in \mathbf{Z}^l \tag{2.8}$$

In this case the Hamiltonian (1.1) has the interpretation of a perturbation of a harmonic nonresonant oscillator system (e.g., a chain of $l+2$ particles connected by linear springs, the end particles being fixed).

The other case will be the one in which $h(\underline{A})$ is strictly convex (in particular strictly anisochronous). Precisely, we assume in this case for suitable $m, M > 0$ the conditions

$$\begin{aligned} m\|\underline{v}\|^2 &\leqslant \frac{\partial^2 h}{\partial \underline{A}\, \partial \underline{A}}(\underline{A})\, \underline{v} \cdot \underline{v} \qquad && \forall \underline{v} \in \mathbf{R}^l,\ \forall (\underline{A}) \in V_R \\ \left\| \frac{\partial^2 h}{\partial \underline{A}\, \partial \underline{A}} \underline{v} \right\|_W &\leqslant M \|\underline{v}\| && \forall \underline{v} \in \mathbf{C}^l \end{aligned} \tag{2.9}$$

For instance, a "rotators system" with inertia moments $I_j > 0$, $j = 1, ..., l$, i.e.,

$$h(\underline{A}) = \frac{1}{2} \sum_{j=1}^{l} \frac{A_j^2}{I_j} \tag{2.10}$$

verifies our assumptions.[4] One could consider intermediate cases: however, they are often reducible to (2.7) or (2.9), and we shall not discuss them here.

In the case of harmonic oscillators the behavior of the perturbed system is simply described: basically the system behaves as if no perturbation were present, up to times of order $\exp \varepsilon^{-b}$, with $b > 0$. This is made precise by the following theorem, in which $t \to (\underline{A}(t), \underline{\varphi}(t))$ denotes the time evolution for (2.1):

Proposition 1. Assume (2.7) and (2.8), and suppose ε small enough; we shall show that this can be taken to mean[5]

$$\varepsilon < \varepsilon_0(l, \xi, CE) \equiv (2^{10l+9}(l+1)^{l+1}\, \xi^{-2l-1} CE)^{-2} \tag{2.11}$$

[4] Such a system is particularly suitable for classical perturbation theory; in the special case of short-range interaction, Wayne[(4)] was able to obtain results close to Nekhoroshev's theorem, with estimates independent of the number of degrees of freedom. For an elementary illustration of classical perturbation theory on this model, see also Refs. 5 and 6.

[5] Recall, however, that we restrict ξ to be $\leqslant 1$, in order to simplify several formulas, including (2.11).

Then for all initial data $(\underline{A}_0, \underline{\varphi}_0) \in V_R \times \mathbf{T}^l$ one has

$$\|\underline{A}(t) - \underline{A}_0\| < \frac{1}{2} \rho \left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \tag{2.12}$$

for all t such that

$$|t| \leqslant T \equiv \frac{1}{\sqrt{\varepsilon}} \frac{1}{E} \left(\frac{\varepsilon}{\varepsilon_0}\right)^{-(1/\varepsilon)^b}, \qquad b = \frac{1}{4(l+1)} \tag{2.13}$$

More precisely, one can prove the following more detailed proposition which implies the above one:

Proposition 1'. Assume (2.7), (2.8), and (2.11). Then the system (2.1) is canonically conjugated via a real analytic canonical transformation $(\underline{A}, \underline{\varphi}) = \mathscr{C}_\varepsilon(\underline{A}', \underline{\varphi}')$ to the system with Hamiltonian

$$\underline{\omega} \cdot \underline{A}' + \varepsilon h_1(\underline{A}', \varepsilon) + \varepsilon \left(\frac{\varepsilon}{\varepsilon_0}\right)^{(1/\varepsilon)^b} f_\infty(\underline{A}', \underline{\varphi}', \varepsilon) \tag{2.14}$$

with

$$h_1(\underline{A}') = \langle f \rangle(\underline{A}') + \mathscr{O}(\varepsilon) \tag{2.15}$$

where $\langle . \rangle$ denotes averaging over $\underline{\varphi}$; the domains of $\mathscr{C}_\varepsilon, \mathscr{C}_\varepsilon^{-1}$ can be taken to contain the real set $V_R \times \mathbf{T}^l$, and if one denotes

$$\begin{aligned} \mathscr{C}_\varepsilon(\underline{A}', \underline{\varphi}') &= \begin{cases} \underline{A} = \underline{A}' + \underline{\Xi}(\underline{A}', \underline{\varphi}') \\ \underline{\varphi} = \underline{\varphi}' + \underline{\Delta}(\underline{A}', \underline{\varphi}') \end{cases} \\ \mathscr{C}_\varepsilon^{-1}(\underline{A}, \underline{\varphi}) &= \begin{cases} \underline{A}' = \underline{A} + \underline{\Xi}'(\underline{A}, \underline{\varphi}) \\ \underline{\varphi}' = \underline{\varphi} + \underline{\Delta}'(\underline{A}, \underline{\varphi}) \end{cases} \end{aligned} \tag{2.16}$$

then $\underline{\Xi}, \underline{\Xi}', \underline{\Delta}, \underline{\Delta}'$ are holomorphic in $W_\infty = W(V_R; \frac{1}{8}\rho, \frac{1}{4}\xi)$ where they satisfy the estimates

$$\begin{aligned} \|\underline{\Xi}\|_{W_\infty}, \|\underline{\Xi}'\|_{W_\infty} &\leqslant \frac{1}{8} \rho \left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \\ \|\underline{\Delta}\|_{W_\infty}, \|\underline{\Delta}'\|_{W_\infty} &\leqslant \frac{1}{4} \xi \left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \end{aligned} \tag{2.17}$$

and, finally, in the same domain one has

$$\left\|\frac{\partial h_1}{\partial \underline{A}}\right\|_{W_\infty} \leqslant 2E, \qquad \|f_\infty\|_{W_\infty} \leqslant \rho E \tag{2.18}$$

In particular, (2.14) implies that in perturbed systems of nonresonant harmonic oscillators (e.g., the above-mentioned simple chain) one cannot see chaotic motions before a time exponentially long compared to the typical unperturbed time scale E^{-1}. In the proof we explicitly construct h_1. The constants in (2.11), (2.17), and (2.18) are not optimal and are reported just to give an idea of their l-dependence, which should be qualitatively close to that of the estimates.

3. RESULTS FOR THE ANHARMONIC CASE

Before describing the results on the anisochronous cases obeying (2.9), we need some definitions on resonances. Indeed, at the basis of the discussion that follows is a decomposition of the action space into regions where the angular velocity $\underline{\omega}(\underline{A}) = (\partial h/\partial \underline{A})(\underline{A})$ has well-defined "resonance properties."

Given $\underline{\nu}_1,..., \underline{\nu}_r \in \mathbf{Z}^l$ and linearly independent, let $\mathcal{M} = \mathcal{M}(\underline{\nu}_1,..., \underline{\nu}_r)$ be the plane in $\mathbf{Z}^l$ generated by $\underline{\nu}_1,..., \underline{\nu}_r$; a "resonant surface" (of "order r") with the plane $\mathcal{M}$ is the surface

$$\Sigma_{\mathcal{M}} = \{\underline{A} \mid \underline{A} \in V_R; \underline{\omega}(\underline{A}) \cdot \underline{\nu}_j = 0, j = 1,..., r\} \tag{3.1}$$

Let $r > 0$, $N = \varepsilon^{-r}$, and let $\mathcal{M}$ be as above, with $\underline{\nu}_1,..., \underline{\nu}_r$ such that $|\underline{\nu}_j| \leqslant N$, $j = 1,..., r$: in this case we say that $\mathcal{M}$ admits an N-basis. Given a sequence $0 < \lambda_0 < \cdots < \lambda_l$ and a plane $\mathcal{M} \subset \mathbf{Z}^l$ admitting an N-basis, we can define the "resonant region" (or "block") $\mathcal{B}_{\mathcal{M}}$ to be a neighborhood of $\Sigma_{\mathcal{M}}$ consisting of all the $\underline{A}$'s such that, for at least one N-basis $\{\underline{\nu}_1,..., \underline{\nu}_r\}$ of $\mathcal{M}$, one has

$$|\underline{\omega}(\underline{A}) \cdot \underline{\nu}_j| < \lambda_r, \qquad j = 1,..., r \tag{3.2}$$

while at the same time it is also, for $r < l$:

$$|\underline{\omega}(\underline{A}) \cdot \underline{\nu}| \geqslant \lambda_{r+1} \qquad \forall \underline{\nu} \notin \mathcal{M}, |\underline{\nu}| \leqslant N \tag{3.3}$$

Formally, denoting $d_{\mathcal{M},N}(\underline{A}) = \inf_{N-\text{bases}} \max_{1 \leqslant j \leqslant r} |\underline{\omega}(\underline{A}) \cdot \underline{\nu}_j|$ ($d_{\mathcal{M},N}(\underline{A})$ is nothing but a convenient measure of the orthogonal component of $\underline{\omega}(\underline{A})$ to $\Sigma_{\mathcal{M}}$), we set

$$\mathcal{B}_{\mathcal{M}} = \{\underline{A} \mid \underline{A} \in V_R; |\underline{\omega}(\underline{A}) \cdot \underline{\nu}| > \lambda_{r+1}, \forall \underline{\nu} \notin \mathcal{M}, |\underline{\nu}| \leqslant N; d_{\mathcal{M},N}(\underline{A}) < \lambda_r\} \tag{3.4}$$

If $\lambda_{r+1} > 2\lambda_r$, it is also clear that the set $\mathcal{U}_{\mathcal{M}}$ defined by

$$\mathcal{U}_{\mathcal{M}} = \{\underline{A} \mid \underline{A} \in V_{R+(1/2)\rho}; |\underline{\omega}(\underline{A}) \cdot \underline{\nu}| > 2\lambda_r, \forall \underline{\nu} \notin \mathcal{M}, |\underline{\nu}| \leqslant N\} \tag{3.5}$$

contains $\mathscr{B}_{\mathscr{M}}$. Notice that all of these definitions make sense also for $r=0$ ($\mathscr{M}$, in this case, contains only the null vector; $\Sigma_{\mathscr{M}}$ is the whole space V_R, and $d_{\mathscr{M},N}$ vanishes). By construction, $\bigcup_{\mathscr{M}}\mathscr{B}_{\mathscr{M}}$ covers the set V_R.

Our results require, to be formulated properly, a choice of $\lambda_0,\ldots,\lambda_l$ as well as of two other parameters b and ε_c, which depend only on the functions h, f in (2.1) and on the dimension l. We are particularly interested, for rather obvious reasons, in trying to get bounds with a good l-dependence, compatibly with the methods that we use: examining our proof one can see that a rather convenient choice of the above parameters is

$$\begin{aligned}
&b=\frac{1}{8l(l+1)}\\
&\varepsilon_c=\min(\tfrac{1}{2}\varepsilon_0,\varepsilon_1)\\
&\qquad\varepsilon_0=\left[B_l^{-1}\left(\frac{E}{\rho M}\right)^2\left(\frac{m}{M}\right)^{4l}\xi^{2(l+1)}\right]^4\\
&\qquad\varepsilon_1=\left(\frac{\rho M}{2E}\right)^8\\
&\qquad B_l=2^{22l+19}l^{8l+1}\\
&\lambda_r=\lambda_r^{(0)}\varepsilon^{\sigma_r}\\
&\qquad\lambda_r^{(0)}=\left(\frac{8lM}{m}\right)^{r-l}E\leqslant E\\
&\qquad\sigma_r=\frac{1}{8}-\frac{r(r-1)}{16l(l+1)},\qquad\frac{1}{16}<\sigma_r\leqslant\frac{1}{8}
\end{aligned}\tag{3.6}$$

The above list of parameters is rather arbitrary in the choice of the various constants, but it reflects quite well the kind of dependence on h, f, l of the various constants which must be fixed in the course of the proof: if one is not interested in getting general results but only wishes to get results of the type: "there exists a constant such that ...," then the only feature to retain of the constants in (3.6) is that $\lambda_{r+1}>2\lambda_r$ and $\lambda_r\to 0$, $\lambda_r/\lambda_{r+1}\to 0$ for $\varepsilon\to 0$ (i.e., the resonant blocks become thinner and thinner the closer ε is to 0).

We can now formulate a proposition similar to Proposition 1 for the system described by (2.1), (2.9).

Proposition 2. Let $h+\varepsilon f$ verify (2.9); assume $\varepsilon<\varepsilon_c$, and define the resonant regions for $\underset{\sim}{\omega}(\underset{\sim}{A})=(\partial h/\partial\underset{\sim}{A})(\underset{\sim}{A})$ as above.

Then:

(i) For any motion $t \to (\underline{A}(t), \underline{\varphi}(t))$, with $\underline{A}(0) \in \mathscr{B}_{\mathscr{M}}$, one has

$$
\begin{aligned}
&\underline{A}(t) \in \mathscr{U}_{\mathscr{M}} \\
&\|\underline{A}(t) - \underline{A}(0)\| \leqslant (6l^2 + 4) \frac{E}{m} \varepsilon^{1/16} \qquad (3.7)\\
&\text{for} \quad |t| < T \equiv \frac{1}{E}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{-3/4} e^{(1/8)\xi(1/\varepsilon)^b}
\end{aligned}
$$

(ii) For each $\mathscr{M}$ there exists an "adapted" system of canonical coordinates $(\underline{A}', \underline{\varphi}')$ such that $(\underline{A}, \underline{\varphi}) = \mathscr{C}_{\mathscr{M},\varepsilon}(\underline{A}', \underline{\varphi}')$ with both $\mathscr{C}_{\mathscr{M},\varepsilon}, \mathscr{C}^{-1}_{\mathscr{M},\varepsilon}$ real analytic in $W_{\mathscr{M},\infty} \equiv W(\mathscr{U}_{\mathscr{M}}, \rho', \xi')$, $\rho' = (\lambda_r^{(0)}/8M)\, \varepsilon^{1/4}$, $\xi' = \frac{1}{8}\xi$, such that in the new coordinates $(\underline{A}', \underline{\varphi}')$ the Hamiltonian (2.1) takes the form

$$H'_\varepsilon(\underline{A}', \underline{\varphi}') = h(\underline{A}') + \varepsilon G(\underline{A}', \underline{\varphi}', \varepsilon) + e^{-(1/8)\xi(1/\varepsilon)^b} f_\infty(\underline{A}', \underline{\varphi}', \varepsilon) \qquad (3.8)$$

with:

(a)

$$
\begin{aligned}
G(\underline{A}', \underline{\varphi}', \varepsilon) &= \sum_{\underline{\nu} \in \mathscr{M}} G_{\underline{\nu}}(\underline{A}', \varepsilon)\, e^{i\underline{\nu}\cdot\underline{\varphi}} \qquad (3.9)\\
G_{\underline{\nu}}(\underline{A}', 0) &= f_{\underline{\nu}}(\underline{A}')
\end{aligned}
$$

where $f_{\underline{\nu}}(\underline{A}') \equiv (2\pi)^{-l} \int e^{-i\underline{\nu}\cdot\underline{\varphi}} f(\underline{A}', \underline{\varphi})\, d\underline{\varphi}$ denote the Fourier components of f;

(b)

$$|G(\underline{A}'_1, \underline{\varphi}'_1, \varepsilon) - G(\underline{A}'_2, \underline{\varphi}'_2, \varepsilon)| < lE(10\rho + \|\underline{A}'_1 - \underline{A}'_2\|) \qquad (3.10)$$

for all pairs $(\underline{A}'_1, \underline{\varphi}'_1), (\underline{A}'_2, \underline{\varphi}'_2) \in \mathscr{U}_{\mathscr{M}} \times \mathbf{T}^l$;

(c)

$$\left\|\frac{\partial f_\infty}{\partial \underline{\varphi}'}\right\|_{\mathscr{U}_{\mathscr{M}} \times \mathbf{T}^l} \leqslant \frac{\lambda_r^{(0)2}}{M} \varepsilon_0^{2\sigma_r + b} \left(\frac{\varepsilon}{\varepsilon_0}\right) < \frac{E^2}{M}\left(\frac{\varepsilon}{\varepsilon_0}\right) \qquad (3.11)$$

(iii) Finally, if the canonical maps $\mathscr{C}_{\mathscr{M},\varepsilon}, \mathscr{C}^{-1}_{\mathscr{M},\varepsilon}$ are written as (2.16), then

$$
\begin{aligned}
&\|\underline{\Xi}\|_{W_{\mathscr{M},\infty}}, \|\underline{\Xi}'\|_{W_{\mathscr{M},\infty}} \leqslant \frac{\lambda_0^{(0)}}{M} \varepsilon^{\sigma_r + b}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-4\sigma_r} < \frac{E}{M}\varepsilon^{1/8}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \qquad (3.12)\\
&\|\underline{\Delta}\|_{W_{\mathscr{M},\infty}}, \|\underline{\Delta}'\|_{W_{\mathscr{M},\infty}} \leqslant \xi\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-4\sigma_r} < \xi\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2}
\end{aligned}
$$

The meaning of Part (ii) of Proposition 2 is that, up to exponentially long times, the motions of the Hamiltonian system (2.1) are described in $\mathscr{U}_{\mathscr{M}}$ by the effective Hamiltonian

$$h(\underline{A}') + \varepsilon G(\underline{A}', \underline{\varphi}', \varepsilon) \tag{3.13}$$

Since G depends only on $r = \dim \mathscr{M}$ independent combinations of angles, it follows that (3.13) admits $l - r$ independent combinations of actions which are integrals of motion, so that the actions $\underline{A}'$ are confined to move on an r-dimensional plane $A_{\mathscr{M}}$ (actually, the plane through the initial point $\underline{A}'(0)$, parallel to $\mathscr{M}$). The convexity property of the "kinetic energy" $h(\underline{A}')$ provides, in turn, a point of minimum or maximum for $h(\underline{A}')$ which, as we shall see, is a point $\underline{A}$ where

$$\underline{\omega}(\underline{A}) \cdot \underline{v} = 0 \qquad \forall \underline{v} \in \mathscr{M} \tag{3.14}$$

Then the bound on εG given by (3.10) immediately implies the "confinement of the actions" in the sense that orbits starting inside $\mathscr{B}_{\mathscr{M}}$ cannot escape out of $\mathscr{U}_{\mathscr{M}}$ (where the canonical transformation is properly defined), in such a way that up to time T the estimate (3.7) is satisfied.

For $r = \dim \mathscr{M} = 0$, G is angle independent, and the situation is almost identical to the case of nonresonant harmonic oscillators considered in Proposition 1 and 1′.

In the next section we present some other simple corollaries of Proposition 2, which allow us to better understand its meaning.

4. SLOW AND FAST VARIABLES

Consider (3.8) with G, f_∞ verifying the properties of Proposition 2. One has then the following:

Proposition 3. There exists a canonical linear change of variables $(\underline{A}', \underline{\varphi}') \to (\underline{\tilde{A}}, \underline{\tilde{\varphi}}) = (J\underline{A}', (J^T)^{-1} \underline{\varphi}')$, where J is an integer matrix with determinant one, such that if $(\underline{\tilde{A}}, \underline{\tilde{\varphi}})$ is denoted

$$\begin{aligned} \underline{\tilde{A}} &= (S_1,..., S_r, F_1,..., F_{l-r}) \equiv (\underline{S}, \underline{F}) \\ \underline{\tilde{\varphi}} &= (\sigma_1,..., \sigma_r, \psi_1,..., \psi_{l-r}) \equiv (\underline{\sigma}, \underline{\psi}) \end{aligned} \tag{4.1}$$

where $\underline{S}, \underline{\sigma}$ stand for "slow" and $\underline{F}, \underline{\psi}$ stand for "fast," then

(i) The Hamiltonian (3.8) takes the form

$$\begin{aligned} \tilde{H}_\varepsilon(\underline{S}, \underline{F}, \underline{\sigma}, \underline{\psi}) = \tilde{h}(\underline{S}, \underline{F}) + \varepsilon \tilde{G}(\underline{S}, \underline{F}, \underline{\sigma}, \varepsilon) \\ + e^{-(1/8)\xi(1/\varepsilon)^b} \tilde{f}_\infty(\underline{S}, \underline{F}, \underline{\sigma}, \underline{\psi}, \varepsilon) \end{aligned} \tag{4.2}$$

with

$$\tilde{G}(\underline{S}, \underline{F}, \underline{\sigma}, 0) = \frac{1}{(2\pi)^{l-r}} \int f(J^{-1}(\underline{S}, \underline{F}), J^T(\underline{\sigma}, \underline{\psi}))\, d\underline{\psi} \tag{4.3}$$

(ii) The new resonance plane $\tilde{\mathcal{M}} \equiv J\mathcal{M}$ is given by

$$\tilde{\mathcal{M}} = \{\underline{v} \mid \underline{v} \in \mathbf{Z}^l,\ v_{r+1},..., v_l = 0\} \tag{4.4}$$

and correspondingly, in the new resonant block $J\mathcal{B}_{\mathcal{M}}$ the new angular velocities $\tilde{\underline{\omega}} = (\partial\tilde{h}/\partial\underline{S}, \partial\tilde{h}/\partial\underline{F})$ obey

$$\left|\frac{\partial\tilde{h}}{\partial S_j}(\underline{S}, \underline{F})\right| \equiv |\tilde{\omega}_j(\underline{S}, \underline{F})| \leqslant \lambda_r \qquad j = 1,..., r \tag{4.5}$$

The names slow and fast attributed to the $(\underline{S}, \underline{\sigma})$, $(\underline{F}, \underline{\psi})$ variables come from (4.5) which, together with (4.2), shows that $\dot{\underline{\sigma}} = \mathcal{O}(\lambda_r) \leqslant \mathcal{O}(\varepsilon^{1/16})$, $\dot{\underline{\psi}} = \mathcal{O}(1)$.

The above separation of the coordinates into fast and slow ones can be made more precise and quantitative via the following proposition (which is also a corollary of proposition 2):

Proposition 4. Denote $\underline{S}^*(\underline{F})$ the function implicitly defined in $J\mathcal{B}_{\mathcal{M}}$ by $\tilde{\omega}_j(\underline{S}^*(\underline{F}), \underline{F}) = 0$, $j = 1,..., r$. Given any motion $(\underline{S}_0, \underline{F}_0, \underline{\sigma}_0, \underline{\psi}_0) \to (\underline{S}(t), \underline{F}(t), \underline{\sigma}(t), \underline{\psi}(t))$ with $(\underline{S}_0, \underline{F}_0) \in J\mathcal{B}_{\mathcal{M}}$, introduce the following rescaling of variables:

$$\begin{aligned} \underline{S}(t) &= \underline{S}^*(\underline{F}_0) + \sqrt{\varepsilon}\, \hat{\underline{S}}(\sqrt{\varepsilon}\, t) \\ \underline{F}(t) &= \underline{F}_0 + \sqrt{\varepsilon}\, \hat{\underline{F}}(\sqrt{\varepsilon}\, t) \\ \underline{\sigma}(t) &= \hat{\underline{\sigma}}(\sqrt{\varepsilon}\, t) \\ \underline{\psi}(t) &= \hat{\underline{\psi}}(\sqrt{\varepsilon}\, t) \end{aligned} \tag{4.6}$$

Then $(\hat{\underline{S}}, \hat{\underline{F}}, \hat{\underline{\sigma}}, \hat{\underline{\psi}})$ are canonical variables, whose evolution as functions of $\theta \equiv \sqrt{\varepsilon}\, t$ is described by a Hamiltonian, parametrized by the $\underline{F}_0$, of the form

$$\begin{aligned} \hat{H}_{\underline{F}0} = \frac{1}{\sqrt{\varepsilon}} h_{\underline{F}0}(\hat{\underline{F}}, \varepsilon) &+ \left[\frac{1}{2} L_{\underline{F}0}(\hat{\underline{F}}, \varepsilon)\, \hat{\underline{S}} \cdot \hat{\underline{S}} + V_{\underline{F}0}(\hat{\underline{\sigma}}) \right. \\ &\left. + \sqrt{\varepsilon}\, V^{(1)}_{\underline{F}0}(\hat{\underline{S}}, \hat{\underline{F}}, \hat{\underline{\sigma}}, \varepsilon)\right] + e^{-(1/8)\xi(1/\varepsilon)^b}\, V^{\infty}_{\underline{F}0}(\hat{\underline{S}}, \hat{\underline{F}}, \hat{\underline{\sigma}}, \hat{\underline{\psi}}, \varepsilon) \end{aligned} \tag{4.7}$$

where the matrix $L_{\underline{F}_0}$ is the first $r \times r$ minor of the matrix

$$\tilde{L}(\underline{S}, \underline{F}) \equiv \frac{\partial \tilde{\underline{\omega}}}{\partial \tilde{\underline{A}}} = J^{-1} \frac{\partial^2 h}{\partial \underline{A}\, \partial \underline{A}} J \tag{4.8}$$

computed at $\underline{S} = \underline{S}^*(\underline{F})$, and hence has the same convexity properties of $\partial^2 h/\partial \underline{A}\, \partial \underline{A}$; while

$$V_{\underline{F}_0}(\hat{\underline{\sigma}}) = \tilde{G}(\underline{S}^*(\underline{F}_0), \underline{F}_0, \hat{\underline{\sigma}}, \underline{0}) \tag{4.9}$$

and $V^{(1)}_{\underline{F}_0}$, $V^{\infty}_{\underline{F}_0}$ are bounded in $J\mathscr{B}_{\mathscr{M}}$ as well as their derivatives, uniformly in $|\varepsilon| < \varepsilon_c$; see (3.6).

We remark that if $(\underline{S}_0, \underline{F}_0, \underline{\sigma}_0, \underline{\psi}_0) \in J\mathscr{B}_{\mathscr{M}} \times \mathbf{T}^l$, then $|\sqrt{\varepsilon}\, \hat{S}_j(0)|$, $|\sqrt{\varepsilon}\, \hat{F}_j(0)| \leqslant \mathscr{O}(\varepsilon^{1/8})$, $j = 1, ..., r$: hence initial data for (4.7) with $|\hat{S}_j(0)|$, $|\hat{F}_j(0)| \leqslant Y$ (where Y is any positive constant) are in the domain of applicability of Proposition 4, provided $\varepsilon \leqslant \hat{\varepsilon}(Y)$, where $\hat{\varepsilon}(Y)$ depends on h, f but not on ε.

Propositions 3 and 4 are simple corollaries of Proposition 2: indeed, Proposition 3 follows as a straightforward application of Lemma 5 of Section 9, which states the existence of an integer matrix J, $\det J = 1$, with the above properties; the existence of $\underline{S}^*(\underline{F})$, which is at the basis of Proposition 4, follows from convexity and the implicit function theorem; the details are left to the reader, who can also check that the detailed $\hat{\underline{S}}$-dependence of the first three terms in the r.h.s. of (4.7) follows by a Taylor expansion in $\hat{\underline{S}}$ of $\tilde{h} + \varepsilon\tilde{G}$ defined in (4.2).

One could easily provide explicit bounds for $\hat{G}$ and $\hat{f}_\infty$ in (4.2): actually, they are exactly the same as the corresponding ones of Proposition 2, apart from the change of norms due to the linear transformation which separates the slow and fast variables.

5. THE TIME SCALES OF THE PERTURBED MOTIONS. INTERMEDIATE SCALES AND CHAOTIC MOTIONS

The interest of Proposition 4 is that it allows a deeper understanding of the dynamics of our system inside an r-dimensional resonant region. In fact, form (4.7) makes evident the existence of three time scales for the motion, which are well separated for small ε:

(a) a "microscopic" time scale $t \sim (\sqrt{\varepsilon}\, E)^{-1}$ (i.e., $\theta \sim E^{-1}$), where the system is integrable and only the fast $(l-r)$ angles move;

(b) a widely extended "intermediate" time scale $(\sqrt{\varepsilon}\, E)^{-1} \ll t \ll E^{-1} e^{(1/8)\xi\varepsilon^{-b}}$ (or $E^{-1} \ll \theta \ll E^{-1} e^{(1/8)\xi\varepsilon^{-b}}$), where the fast actions $\hat{\underline{F}}$ are still "frozen," while at the same time the slow variables can have a very nontrivial dynamics in a bounded domain (actually, a domain small with ε);

(c) a "long" time scale $t \sim E^{-1} e^{(1/8)\xi\varepsilon^{-b}}$, where all of the degrees of freedom are nontrivially involved in the dynamics, and the motion is no longer local (i.e., the actions can change as much as allowed by the conservation of the total energy: a phenomenon which, when actually happening, is called "Arnold diffusion," after Arnold exhibited an example where it was present).[7,8]

Let us consider in more detail the intermediate time scale, disregarding for a moment the exponentially small coupling term of Hamiltonian (4.7). Within this approximation, the evolution of the coordinates $\hat{\underline{S}}$, $\hat{\underline{\sigma}}$ is governed by an effective Hamiltonian with only r degrees of freedom, of the form

$$H_{\text{eff}}(\hat{\underline{S}}, \hat{\underline{\sigma}}, \varepsilon) = \frac{1}{2} L_{\text{eff}} \hat{\underline{S}} \cdot \hat{\underline{S}} + V_{\text{eff}}(\hat{\underline{\sigma}}) + \mathcal{O}(\varepsilon^{1/2}) \tag{5.1}$$

The matrix L_{eff} is positive symmetric; for simplicity we can think it as diagonal: $(L_{\text{eff}})_{ij} = I_j^{-1} \delta_{ij}$. Hence

$$H_{\text{eff}}(\hat{\underline{S}}, \hat{\underline{\sigma}}) = \frac{1}{2} \sum_{j=1}^{r} \frac{\hat{S}_j^2}{I_j} + V_{\text{eff}}(\underline{\sigma}) + \mathcal{O}(\varepsilon^{1/2}) \tag{5.2}$$

This Hamiltonian represents r rotators, with inertia moments $I_1, \ldots, I_r$, coupled by an angle-dependent potential $V_{\text{eff}}(\hat{\underline{\sigma}}) + \mathcal{O}(\varepsilon^{1/2})$; thus, within the above approximation, it turns out that any system with convex unperturbed Hamiltonian essentially reduces, near an r-dimensional resonance, to a system of r coupled rotators, subject to a purely "positional" force up to $\mathcal{O}(\varepsilon^{1/2})$.

Let us make a few comments on the reduced Hamiltonian system (5.2):

(i) for $r = 1$ the system is obviously integrable (in fact, it is essentially a pendulum), independently of the nature of the perturbation in the original Hamiltonian (2.1). For $l = 2$ this is the only nontrivial possibility. (In fact, near a point with $\underline{\omega}(\underline{A}) = 0$ one can have a resonance of order 2, no matter how small ε is: this, however, essentially restricts $\underline{A}$ to a neighborhood of a point where $h(\underline{A})$ has a minimum, so that the above Proposition 4 will just tell us that forever the motion near this equilibrium point will look like the one of a nontrivial two-dimensional system. In other words, the $(\underline{A}, \underline{\varphi})$ coordinates—in particular, the action scale—were not the appropriate ones, and the system cannot really be thought of as a perturbation of an integrable system.)

(ii) for $r \geqslant 2$ the Hamiltonian system (5.2) can exhibit chaotic motions. In fact, as far as structurally stable properties are concerned, one can take $\varepsilon = 0$ in (5.2); on the other hand, V_{eff} is directly related, via (4.9) and (4.3), to the original perturbation f in (2.1). Therefore, given a resonance $\mathscr{M}$, by suitably choosing f one can give V_{eff} any preassigned form. Examples of functions V_{eff} producing, say, homoclinic phenomena (a simple form of chaotic motions), are easily constructed: one of them is explicitly produced in the Appendix. It is believed that chaotic phenomena are "generic" for (5.2) in V_{eff}, if $r \geqslant 2$ (for a numerical illustration of chaotic motions in a system of coupled rotators, see for instance Ref. 6). As far as the exponentially small term in (4.7) can be disregarded, these motions should appear as local chaotic motions for the original Hamiltonian (2.1), taking place inside resonances.

(iii) For any r and any interaction, at low energy the Hamiltonian (5.2) produces mostly ordered motions. In fact, near the point $\hat{\underline{S}} = 0$, $\hat{\underline{\sigma}} = \hat{\underline{\sigma}}_0$, where $\hat{\underline{\sigma}}_0$ is a minimum for $V_{\text{eff}}(\hat{\underline{\sigma}})$, the system is equivalent to r weakly coupled harmonic oscillators (normal modes), the energy itself measuring the coupling. Still within our approximation, in the original coordinates one should see local quasi-periodic motions taking place "deeply inside" each resonance ($\hat{\underline{S}} \sim \underline{0}$). For a numerical illustration of these ordered motions associated to resonances, see for instance Ref. 6. The apparently paradoxical situation in which chaotic and ordered motions generically coexist emerges here clearly, but this has been well known since Poincaré.[9]

The question then becomes whether it is reasonable to ignore the coupling term $\mathcal{O}(e^{-(1/8)\xi\varepsilon^{-b}})$ for the whole intermediate time scale. If the effective Hamiltonian (5.2) is integrable (thus, in particular, for $r = 1$), then the motions of the complete Hamiltonian (4.7) will certainly remain regular for the whole intermediate time scale: indeed, after an appropriate change of coordinates in the $(\hat{\underline{S}}, \hat{\underline{\sigma}})$-space, Hamiltonian (4.7) will turn in a weakly perturbed integrable Hamiltonian, with a perturbation of order $e^{-(1/8)\xi\varepsilon^{-b}}$. Apart from this case, it is not so easy to answer the above question. On one hand, it is hard to believe that, generically, the exponentially small coupling will significantly modify the qualitative properties of the dynamics. On the other hand, it is obvious that, if Hamiltonian (5.2) has a sensitive dependence on the initial conditions, then one cannot hope that the orbits of the complete Hamiltonian (4.7) with and without the coupling term (at fixed initial data) will remain close to each other for the whole intermediate time scale. Only some partial results are easily achieved: for example, in the Appendix it is shown on the basis of a simple model example that the two dynamics remain extremely close (precisely, at distance of order $e^{-(1/16)\xi\varepsilon^{-b}}$) for a time which grows as an inverse power of ε (while a naive calculation gives only a logarithmic growth). In general, however, this problem of structural stability must be considered to be basically unsolved.

6. PROOF OF PROPOSITIONS 1 AND 1′

6.1. In this section we give a quite detailed proof of Propositions 1 and 1′, using in place of the diophantine condition (2.8) the more general one

$$|\underline{\omega} \cdot \underline{\nu}|^{-1} \leqslant C|\underline{\nu}|^{\alpha} \tag{6.1}$$

where α is any positive number. In fact, as already remarked in Section 2, Proposition 1 is an obvious corollary of Proposition 1′: indeed for real variables and $|t| \leqslant T$, one has

$$\begin{aligned} \|\underline{A}(t) - \underline{A}(0)\|_{W_\infty} &\leqslant 2\|\Xi\|_{W_\infty} + 8\xi^{-1}\varepsilon\left(\frac{\varepsilon}{\varepsilon_0}\right)^{(1/\varepsilon)^b} \|f_\infty\|_{W_\infty} T \\ &\leqslant \frac{1}{4}\rho\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} + 8\xi^{-1}\rho\varepsilon^{1/2} \\ &< \frac{1}{2}\rho\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \end{aligned} \tag{6.2}$$

as follows from (2.14) and (2.18), because the *a priori* estimate, necessary to make sure that the orbit, for all $|t| \leqslant T$, does not get out of the domain of definition of $\mathscr{C}_\varepsilon^{-1}$ follows self-consistently by the same chain of inequalities in (6.2).

To prove Proposition 1′, we perform a canonical transformation $(\underline{A}, \underline{\varphi}) = \mathscr{C}_\varepsilon(\underline{A}', \underline{\varphi}')$ generated by

$$\Phi(\underline{A}', \underline{\varphi}, \varepsilon) = \sum_{k=1}^{n} \varepsilon^k \Phi_k(\underline{A}', \underline{\varphi}) \tag{6.3}$$

via the usual relations

$$\underline{A} = \underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}(\underline{A}', \underline{\varphi}, \varepsilon), \qquad \underline{\varphi}' = \underline{\varphi} + \frac{\partial \Phi}{\partial \underline{A}'}(\underline{A}', \underline{\varphi}, \varepsilon) \tag{6.4}$$

$\Phi_1, \ldots, \Phi_n$ are determined by imposing

$$\begin{aligned} &\underline{\omega} \cdot \left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}(\underline{A}', \underline{\varphi}, \varepsilon)\right) + \varepsilon f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{A}'}(\underline{A}', \underline{\varphi}, \varepsilon), \underline{\varphi}\right) \\ &\quad = \underline{\omega} \cdot \underline{A}' + \sum_{k=1}^{n} \varepsilon^k h_k'(\underline{A}') + R(\underline{A}', \underline{\varphi}, \varepsilon) \end{aligned} \tag{6.5}$$

where the last term is a remainder analytic in ε and divisible by ε^{n+1}. One thus sees that $\Phi_k(\underline{A}', \underline{\varphi})$, $k=1,..., n$, must satisfy equations of the form

$$\underline{\omega} \cdot \frac{\partial \Phi_k}{\partial \underline{\varphi}} + Q_k(\underline{A}', \underline{\varphi}) = h'_k(\underline{A}') \tag{6.6}$$

where $h'_k(\underline{A}')$ is free, while $Q_k(\underline{A}', \underline{\varphi})$ is the coefficient of ε^k in the power series development of $\varepsilon f(\underline{A}' + (\partial\Phi/\partial\underline{\varphi})(\underline{A}', \underline{\varphi}, \varepsilon), \underline{\varphi})$.

Let us write

$$f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi}\right) = \sum_{\underline{m} \in \mathbf{Z}^l} f_{\underline{m}}(\underline{A}', \underline{\varphi}) \left(\frac{\partial \Phi}{\partial \underline{\varphi}}\right)^{\underline{m}} \tag{6.7}$$

where the following compact notation has been used:

$$\begin{aligned} f_{\underline{m}}(\underline{A}', \underline{\varphi}) &= \frac{1}{\underline{m}!} \frac{\partial^{|\underline{m}|} f}{\partial \underline{A}^{\underline{m}}} (\underline{A}', \underline{\varphi}) \\ &= \frac{1}{m_1! \cdots m_l!} \frac{\partial^{m_1 + \cdots + m_l} f}{\partial A_1^{m_1} \cdots \partial A_l^{m_l}} (\underline{A}', \underline{\varphi}) \\ \left(\frac{\partial \Phi}{\partial \underline{\varphi}}\right)^{\underline{m}} &= \prod_{j=1}^{l} \left(\frac{\partial \Phi}{\partial \varphi_j}\right)^{m_j} \end{aligned} \tag{6.8}$$

From this expression one easily obtains

$$\begin{aligned} Q_1 &= f \\ Q_k &= \sum_{\substack{\underline{m} \\ 1 \leqslant |\underline{m}| \leqslant k-1}} f_{\underline{m}}(\underline{A}', \underline{\varphi}) \sum_{\{k_j^i\}_{\underline{m}, k-1}} \prod_{i=1}^{l} \prod_{j=1}^{m_i} \frac{\partial \Phi_{k_j^i}}{\partial \varphi_i} (\underline{A}', \underline{\varphi}), \qquad k \geqslant 1 \end{aligned} \tag{6.9}$$

where the latter sum is extended to the set of indices k_j^i, for $i=1,..., l$ and $j=1,..., m_i$, with the restrictions $\sum_{ij} k_j^i = k-1$ and $k_j^i \geqslant 1$. (If, for one or more values of i, m_i vanishes, the corresponding indices k_j^i are intended to be absent from the sum, while the corresponding product also disappears).

It clearly appears that Q_k depends only on $\Phi_1,..., \Phi_{k-1}$, i.e., Eq. (6.6) for $\Phi_1,..., \Phi_n$ are decoupled. Each of them is then easily solved: one takes

$$h'_k(\underline{A}') = \langle Q_k \rangle (\underline{A}') \tag{6.10}$$

where the bracket indicates averaging on $\varphi_1,..., \varphi_l$, and then makes use of Lemma 1 of Section 9, with $\mathscr{M} = \{\underline{0}\}$, $N = \infty$, and $g = Q_k - \langle Q_k \rangle$.

6.2. Let us introduce the short notation $\|\cdot\|_{\rho', \xi'} = \|\cdot\|_{W(V_R, \rho', \xi')}$. All the basic estimates are contained in the following:

Main Proposition. For any positive n, if $\delta \leqslant \xi/2n$, one has

$$\left\| \frac{\partial \Phi_k}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, \xi - k\delta} \leqslant \rho k!\, F B^k\, \delta^{-(l+\alpha+1)k}, \qquad k = 1, \ldots, n \tag{6.11}$$

with

$$F = 2^{-l-2}, \qquad B = 2^{l+2} B_0 C E \tag{6.12}$$

B_0 being the constant appearing in Lemma 1 of Section 9. Moreover,

$$\| Q_{k+1} \|_{(1/2)\rho, \xi - k\delta} \leqslant 2^{l+1} k!\, F B^k\, \delta^{-(l+\alpha+1)k} \rho E, \qquad k = 1, \ldots, n-1 \tag{6.13}$$

Proof. Inequality (6.11) is proved by induction, while (6.13) is found as a byproduct of the proof of (6.11). Let us assume (6.11) for $k = 1, \ldots, p < n$, and prove it for $k = p + 1$.

A "dimensional estimate," i.e., the use of Cauchy theorem for holomorphic functions, immediately gives, from (2.5),

$$\| f_{\underline{m}} \|_{(1/2)\rho, \xi} \leqslant \left(\frac{2}{\rho} \right)^{|\underline{m}| - 1} E \qquad \forall\, |m| \geqslant 1 \tag{6.14}$$

By inserting this expression and the recurrent hypothesis in (6.9), one gets

$$\| Q_{p+1} \|_{(1/2)\rho, \xi - p\delta} \leqslant \sum_{r=1}^{p} F^r 2^{r-1} \rho E B^p\, \delta^{-(l+\alpha+1)p} \sum_{\substack{\underline{m} \\ |\underline{m}| = r}} \sum_{\{k_j^i\}_{\underline{m}, p}} \prod_{i=1}^{l} \prod_{j=1}^{m_i} k_j^i! \tag{6.15}$$

By Lemmas 3 and 4 (see Section 9) one has then

$$\| Q_{p+1} \|_{(1/2)\rho, \xi - p\delta} \leqslant 2^{l-2} p!\, B^p\, \delta^{-(l+\alpha+1)p} \rho E \sum_{r=1}^{p} (4F)^r \tag{6.16}$$

If condition

$$4F \leqslant \frac{1}{2} \tag{6.17}$$

is satisfied, then one has

$$\| Q_{p+1} \|_{(1/2)\rho, \xi - p\delta} \leqslant 2^{l+1} p!\, F B^p\, \delta^{-(l+\alpha+1)p} \rho E \tag{6.18}$$

which coincides with (6.13) for $k = p$. From statement (iii) of Lemma 1, taking into account $\| Q_{p+1} - \langle Q_{p+1} \rangle \| \leqslant 2 \| Q_{p+1} \|$, one has

$$\left\| \frac{\partial \Phi_{p+1}}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, \xi - (p+1)\delta} \leqslant 2 B_0 C\, \delta^{-(l+\alpha+1)} \| Q_{p+1} \|_{(1/2)\rho, \xi - p\delta}$$
$$\leqslant \rho p!\, F 2^{l+2} B_0 C E B^p\, \delta^{-(l+\alpha+1)(p+1)} \tag{6.19}$$

(6.11) is then achieved, for $k = p+1$, as long as

$$B \geqslant 2^{l+2} B_0 CE \tag{6.20}$$

We must now show that (6.11) is satisfied for $p = 1$, i.e., that solving the equation

$$\underline{\omega} \cdot \frac{\partial \Phi_1}{\partial \underline{\varphi}} + f = \langle f \rangle \tag{6.21}$$

leads to the estimate

$$\left\| \frac{\partial \Phi_1}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, \xi - \delta} \leqslant \rho F B\, \delta^{-(l+\alpha+1)} \tag{6.22}$$

Actually, this estimate immediately follows from statement (ii) of Lemma 1 (see Section 9), with $g = f - \langle f \rangle$, provided

$$2B_0 CE\, \delta \leqslant FB \tag{6.23}$$

The choice we made for B, and F fits all conditions we encountered, i.e., (6.17), (6.20), and (6.23), so that the above proposition is proved.

6.3. Here we use the above proposition to obtain estimates for $\partial\Phi/\partial\underline{A}'$, $\partial\Phi/\partial\underline{\varphi}$. Precisely, we choose $\delta = \xi/2n$ and show that, if condition

$$Bn^{l+\alpha+2} \left(\frac{\xi}{2}\right)^{-(l+\alpha+1)} \varepsilon \leqslant \frac{1}{2} \tag{6.24}$$

is assumed, then one has

$$\left\| \frac{\partial \Phi}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, (1/2)\xi} \leqslant 2^{l+\alpha+2} \rho B_0 CE \xi^{-(l+\alpha+1)} n^{l+\alpha+1} \varepsilon \tag{6.25a}$$

$$\left\| \frac{\partial \Phi}{\partial \underline{A}'} \right\|_{(1/4)\rho, (1/2)\xi} \leqslant 2^{l+\alpha+2} \xi B_0 CE \xi^{-(l+\alpha)} n^{l+\alpha+1} \varepsilon \tag{6.25b}$$

Proof. Let us write

$$\begin{aligned} \left\| \frac{\partial \Phi}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, (1/2)\xi} &\leqslant \sum_{k=1}^{n} \varepsilon^k \left\| \frac{\partial \Phi_k}{\partial \underline{\varphi}} \right\|_{(1/2)\rho, (1/2)\xi} \\ &\leqslant \rho F \sum_{k=1}^{n} \varepsilon^k B^k k! \left(\frac{\xi}{2n}\right)^{-(l+\alpha+1)k} \\ &\leqslant \rho \varepsilon F B n^{l+\alpha+1} \left(\frac{\xi}{2}\right)^{-(l+\alpha+1)} \\ &\quad \times \sum_{k=0}^{n-1} \left[\varepsilon B n^{l+\alpha+2} \left(\frac{\xi}{2}\right)^{-(l+\alpha+1)} \right]^k \end{aligned} \tag{6.26}$$

Condition (6.24) directly leads to (6.25a). Concerning (6.26b), one needs a preliminary estimate for $\partial\Phi_k/\partial\underline{A}$. This can be obtained from Lemma 1 and

$$\underline{\omega}\cdot\frac{\partial}{\partial\underline{\varphi}}\left(\frac{\partial\Phi_k}{\partial\underline{A}}\right)+\frac{\partial Q_k}{\partial\underline{A}}=\left\langle\frac{\partial Q_k}{\partial\underline{A}}\right\rangle \tag{6.27}$$

which trivially follows from (6.6), (6.10). Indeed, one estimates dimensionally $\partial Q_k/\partial\underline{A}$ from (6.13), for $k\geqslant 2$, obtaining

$$\begin{aligned}\left\|\frac{\partial Q_k}{\partial\underline{A}}\right\|_{(1/4)\rho,\xi-(k-1)\delta} &\leqslant 2^{l+3}(k-1)!\,FB^{k-1}\,\delta^{-(l+\alpha+1)(k-1)}E\\ &\leqslant 2^{l+2}k!\,FB^{k-1}\,\delta^{-(l+\alpha+1)(k-1)}E\end{aligned} \tag{6.28}$$

Recalling that $Q_1=f$, and that $F=2^{-l-2}$, this expression is also good for $k=1$.

Statement (ii) of Lemma 1 (see Section 9), with $G=\partial Q_k/\partial\underline{A}'-\langle\partial Q_k/\partial\underline{A}'\rangle$, then yields

$$\left\|\frac{\partial\Phi_k}{\partial\underline{A}'}\right\|_{(1/4)\rho,\xi-k\delta}\leqslant\xi k!\,FB^k\,\delta^{-(l+\alpha+1)k}E \tag{6.29}$$

and, proceeding as above, (6.25b) also follows.

6.4. We now choose n as a function of ε in order to have $n^{l+\alpha+2}\simeq\varepsilon^{-(1/2)}$. More precisely, we set

$$\varepsilon^{-b}\leqslant n<\varepsilon^{-b}+1,\qquad b=\frac{1}{2(l+\alpha+2)} \tag{6.30}$$

Condition (6.24) is then ensured by imposing

$$\varepsilon<\varepsilon_0=(2^{3l+2\alpha+6}B_0CE\xi^{-(l+\alpha+1)})^{-2} \tag{6.31}$$

where expression (6.12) of B has been taken into account.

Estimates (6.26a, b) yield then, in particular,

$$\left\|\frac{\partial\Phi}{\partial\underline{\varphi}}\right\|_{(1/2)\rho,(1/2)\xi}<2^{-l-5}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2}\rho \tag{6.32a}$$

$$\left\|\frac{\partial\Phi}{\partial\underline{A}'}\right\|_{(1/4)\rho,(1/2)\xi}<2^{-l-5}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2}\xi \tag{6.32b}$$

$$\max_{i,j\leqslant l}\left\|\frac{\partial^2\Phi}{\partial A_i\,\partial\varphi_j}\right\|_{(1/4)\rho,(1/2)\xi}<2^{-l-3}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2} \tag{6.32c}$$

where $n \geqslant 8$, which trivially follows from (6.30) and (6.31), has been used, while (6.32c) follows from (6.32a) by a dimensional estimate.

From (6.32) the (global) invertibility of (6.4) easily follows.[6] A canonical transformation $\mathscr{C}_\varepsilon$ is then properly defined; expressions (6.32a, b) show in particular that $\mathscr{C}_\varepsilon, \mathscr{C}_\varepsilon^{-1}$ satisfy (2.27), both being certainly defined in $W(V_R; \frac{1}{8}\rho, \frac{1}{4}\xi)$, as claimed.

6.5. The proof of Proposition 1 is now immediately concluded. First of all, from (6.10) and (6.28), one obtains, setting $W_\infty = W(V_R, \frac{1}{8}\rho, \frac{1}{4}\xi)$:

$$\begin{aligned} \left\| \frac{\partial h_1}{\partial \underline{A}'} \right\|_{W_\infty} &\leqslant \varepsilon^{-1} \sum_{k=1}^{n} \varepsilon^k \left\| \frac{\partial Q_k}{\partial \underline{A}'} \right\|_{W_\infty} \\ &\leqslant 2^{l+2} FE \sum_{k=0}^{n-1} (nB\,\delta^{-(l+\alpha+1)}\varepsilon)^k \\ &\leqslant 2E \end{aligned} \tag{6.33}$$

having used the first of (6.12), (6.24) and $\delta = \xi/2n$ in the last step. Concerning the remainder, it is estimated as follows, on the basis of its definition (2.4). By construction, $R(\underline{A}', \underline{\varphi}, \varepsilon)$ is analytic in ε divisible by ε^{n+1}, as far as $|\varepsilon| < \varepsilon_0$. It follows that

$$|R(\underline{A}', \underline{\varphi}, \varepsilon)| \leqslant \left| \frac{\varepsilon}{\varepsilon_0} \right|^{n+1} \sup_{\varepsilon' \leqslant \varepsilon_0} |R(\underline{A}', \underline{\varphi}, \varepsilon')| \tag{6.34}$$

Now, from (6.5), (6.6), and $Q_1 = f$ [see (6.9)] one obtains, for $|\varepsilon| \leqslant \varepsilon_0$,

$$\begin{aligned} |R(\underline{A}', \underline{\varphi}, \varepsilon)| &= \left| \varepsilon f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi} \right) - \sum_{k=1}^{n} \varepsilon^k Q_k(\underline{A}', \underline{\varphi}) \right| \\ &\leqslant \varepsilon_0 \left| f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi} \right) - f(\underline{A}', \underline{\varphi}) \right| + \sum_{k=2}^{n} \varepsilon_0^k |Q_k(\underline{A}', \underline{\varphi})| \end{aligned} \tag{6.34a}$$

[6] One checks that the first of (6.4) can be inverted in $\underline{A}'$ at fixed $\underline{\varphi}$, thus defining the function $\underline{\Xi}'$ of (2.16) in $W(V_R, \frac{1}{4}\rho - \frac{1}{32}\rho, \frac{1}{2}\xi)$; the result can be substituted in the second of (6.4), defining the function $\underline{\Delta}'$ of (2.16) in $W(V_R, \frac{1}{4}\rho - \frac{1}{32}\rho, \frac{1}{2}\xi)$; similarly, the second of (6.4) can be inverted, expressing $\underline{\varphi}$ as a function of $\underline{\varphi}'$ at fixed $\underline{A}'$, defining the function $\underline{\Delta}$ of (2.26) in $W(V_R, \frac{1}{4}\rho, \frac{1}{2}\xi - \frac{1}{32}\xi)$, and the result can be substituted in the first of (6.4), thus defining $\underline{\Xi}$ of (2.16) in $W(V_R, \frac{1}{4}\rho, \frac{1}{2}\xi - \frac{1}{32}\xi)$. In this way the two maps $\mathscr{C}_\varepsilon, \mathscr{C}_\varepsilon^{-1}$ of (2.16) are both well defined in domains much larger than $W(V_R, \frac{1}{8}\rho, \frac{1}{4}\xi)$, and on this latter domain one also has $\mathscr{C}_\varepsilon \mathscr{C}_\varepsilon^{-1} = \mathscr{C}_\varepsilon^{-1} \mathscr{C}_\varepsilon =$ identity, as is immediately apparent by carefully looking at the domains where $\mathscr{C}_\varepsilon$ and $\mathscr{C}_\varepsilon^{-1}$ are defined.

and thus, for $|\varepsilon| \leqslant \varepsilon_0$:

$$\|R\|_{(1/4)\rho,(1/2)\xi} \leqslant 2l \left\|\frac{\partial \Phi}{\partial \underline{\varphi}}\right\|_{(1/4)\rho,(1/2)\xi} \varepsilon_0 E + 2^{l+1} F\rho E\varepsilon_0 \sum_{k=1}^{n-1} k!\, B^k\, \delta^{-(l+\alpha+1)k} \varepsilon^k$$

$$\leqslant \left(l2^{-l-4} + \frac{1}{2}\right) \rho E\varepsilon_0 < \rho E\varepsilon_0 \tag{6.35}$$

where (6.32a) and (6.24) have been used. If we now denote $R'(\underline{A}', \underline{\varphi}') = R(\underline{A}', \underline{\varphi}' + \underline{\Delta}(\underline{A}', \underline{\varphi}'))$, from (6.34) we conclude that

$$\|R'\|_{(1/4)\rho,(1/4)\xi} \leqslant \rho E\varepsilon \left(\frac{\varepsilon}{\varepsilon_0}\right)^n \leqslant \rho E\varepsilon \left(\frac{\varepsilon}{\varepsilon_0}\right)^{(1/\varepsilon)^b} \tag{6.36}$$

which gives the second of (2.18).

The proof of Proposition 1′ is thus accomplished, as we notice that expression (6.30) for b and (6.31) for ε_0 coincide with (2.13) and (2.11) respectively, for $\alpha = l$, if the expression of B_0 entering Lemma 1 is taken into account.

7. THE ANALYTIC PART OF PROPOSITION 2

7.1. We prove here the following:

Analytic Lemma. Let the Hamiltonian

$$H_\varepsilon(\underline{A}, \underline{\varphi}) = h(\underline{A}) + \varepsilon f(\underline{A}, \underline{\varphi}) \tag{7.1}$$

be analytic in $W(V_R; \rho, \xi)$, with

$$\left\|\frac{\partial h}{\partial \underline{A}}\right\|_W = \max\left(\left\|\frac{\partial f}{\partial \underline{A}}\right\|_W, \frac{1}{\rho}\left\|\frac{\partial f}{\partial \underline{\varphi}}\right\|_W\right) = E \tag{7.2a}$$

$$\left\|\frac{\partial^2 h}{\partial \underline{A}\, \partial \underline{A}} \underline{v}\right\|_W \leqslant M \|\underline{v}\| \qquad \forall \underline{v} \in \mathbf{C}^l \tag{7.2b}$$

Let σ, τ, and β be positive constants satisfying

$$\beta + 4\sigma + 2\tau < 1 \tag{7.3}$$

and given any plane $\mathscr{M} \subset \mathbf{Z}^l$, let $\mathscr{U} \subset V_R$ be a set where the nonresonance relation

$$|\underline{\omega}(\underline{A}) \cdot \underline{\nu}|^{-1} \leqslant \frac{1}{2} C_0 \varepsilon^{-\sigma} \qquad \forall \underline{\nu} \notin \mathscr{M},\ |\underline{\nu}| \leqslant \varepsilon^{-\tau} \tag{7.4}$$

is satisfied.

Assume $C_0 E > 1$ and $\varepsilon \leqslant \max(\tilde{\varepsilon}, \varepsilon_0)$, with

$$\tilde{\varepsilon} = \left(\frac{\rho M}{2E}\right)^{1/\sigma}$$
$$\varepsilon_0 = [2^{5l+13} l B_0^2 \rho^2 C_0^4 E^2 M^2 \xi^{-2(l+1)}]^{-1/(1-4\sigma-2\tau-\beta)} \tag{7.5}$$

B_0 being the constant appearing in Lemma 1, with $\alpha = 0$, and denote

$$\tilde{\rho} = \frac{\varepsilon^{\sigma+\tau}}{MC_0} \tag{7.6}$$

$[\tilde{\rho} < \frac{1}{2}\rho$, from (7.5)]. Then there exists a canonical transformation $(\underline{A}, \underline{\varphi}) = \mathscr{C}_\varepsilon(\underline{A}', \underline{\varphi}')$, with both $\mathscr{C}_\varepsilon$ and $\mathscr{C}_\varepsilon^{-1}$ analytic in $W_\infty \equiv W(\mathscr{U}; \frac{1}{8}\tilde{\rho}, \frac{1}{4}\xi)$, which gives the Hamiltonian the form

$$H'_\varepsilon(\underline{A}', \underline{\varphi}') = h(\underline{A}') + \varepsilon G(\underline{A}', \underline{\varphi}', \varepsilon) + R(\underline{A}', \underline{\varphi}', \varepsilon) \tag{7.7}$$

where

$$G(\underline{A}', \underline{\varphi}', \varepsilon) = \sum_{\underline{\nu} \in \mathscr{M}} G_{\underline{\nu}}(\underline{A}', \varepsilon)\, e^{i\underline{\nu} \cdot \underline{\varphi}'}$$
$$G_{\underline{\nu}}(\dot{\underline{A}}, 0) = f_{\underline{\nu}}(\underline{A}) \tag{7.8}$$

$f_{\underline{\nu}}$ denoting the Fourier components of f. Moreover, given $(\underline{A}'_2, \underline{\varphi}'_2)$ and $(\underline{A}'_1, \underline{\varphi}'_1)$ belonging to W_∞, we have

$$|G(\underline{A}'_2, \underline{\varphi}'_2) - G(\underline{A}'_1, \underline{\varphi}'_1)| \leqslant lE(10\rho + \|\underline{A}'_2 - \underline{A}'_1\|) \tag{7.9}$$

while the remainder R satisfies the exponential estimate

$$\|R\|_{W_\infty} \leqslant \frac{\xi \varepsilon_0^{2\sigma+\tau}}{4C_0^2 M}\left(\frac{\varepsilon}{\varepsilon_0}\right)\left[e^{-(1/8)\xi\varepsilon^{-\tau}} + \left(\frac{\varepsilon}{\varepsilon_0}\right)^{\varepsilon^{-b}}\right]$$
$$b = \frac{\beta}{2l+3} \tag{7.10}$$

Finally, if $(\underline{A}, \underline{\varphi}) = \mathscr{C}_\varepsilon(\underline{A}', \underline{\varphi}')$, then we have

$$\|\underline{A} - \underline{A}'\|_{W_\infty} \leqslant \frac{1}{4}\tilde{\rho}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-3\sigma-2\tau-(1/2)\beta}$$
$$\|\underline{\varphi} - \underline{\varphi}'\|_{W_\infty} \leqslant \frac{1}{4}\xi\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-3\sigma-2\tau-(1/2)\beta} \tag{7.11}$$

Let us notice that the statement of this lemma essentially reduces to the analytic part (ii) of Proposition 2, with $\sigma = \sigma_r$, $C_0^{-1} = \lambda_r^{(0)}$, if the various constants are appropriately fixed. The precise choice of these constants is not made here, as it arises naturally only within the geometric part of Proposition 2. The proof of this lemma closely follows the proof of Proposition 1 (Section 6); some details are thus left to the reader, while more attention is devoted to a few new points.

7.2. First of all, to be able to make dimensional estimates, we need an extension of the nonresonance condition (7.4) to a surrounding of $\mathcal{U}$. Precisely, we need to work in the set $W(\mathcal{U}; \tilde{\rho}, \xi)$, [see (2.3) for the notation], which is characterized, for what concerns the actions, by the condition: $\operatorname{dist}(\underline{A}, \mathcal{U}) \leqslant \tilde{\rho}$. Here one immediately finds

$$|\underline{\omega}(\underline{A}) \cdot \underline{\nu}|^{-1} \leqslant C \equiv C_0 \varepsilon^{-\sigma} \qquad \forall \underline{\nu} \notin \mathcal{M}, |\underline{\nu}| \leqslant N \tag{7.12}$$

with $N = \varepsilon^{-\tau}$, as follows from

$$\begin{aligned} |\underline{\omega}(\underline{A}) \cdot \underline{\nu}| &\geqslant |\underline{\omega}(\underline{A}_0) \cdot \underline{\nu}| - |(\underline{\omega}(\underline{A}) - \underline{\omega}(\underline{A}_0)) \cdot \underline{\nu}| \\ &\geqslant 2C_0^{-1}\varepsilon^{\sigma} - MN\tilde{\rho} = C_0^{-1}\varepsilon^{\sigma} \end{aligned} \tag{7.13}$$

Let us introduce the following notation: for any function $F(\underline{A}, \underline{\varphi}) = \sum_{\underline{\nu} \in \mathbf{Z}^l} F_{\underline{\nu}}(\underline{A})\, e^{i\underline{\nu} \cdot \underline{\varphi}}$ denote

$$\Pi_{\mathcal{M}} F = \sum_{\underline{\nu} \in \mathcal{M}} f_{\underline{\nu}}(\underline{A})\, e^{i\underline{\nu} \cdot \underline{\varphi}} \tag{7.14}$$

As for Proposition 1, we generate a canonical transformation by a generating function of the form

$$\Phi(\underline{A}', \underline{\varphi}, \varepsilon) = \sum_{k=1}^{n} \varepsilon^k \Phi_k(\underline{A}', \underline{\varphi}) \tag{7.15}$$

and write

$$h\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}\right) + \varepsilon f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi}\right) = h'(\underline{A}', \underline{\varphi}', \varepsilon) + R(\underline{A}', \underline{\varphi}', \varepsilon) \tag{7.16}$$

with the following prescriptions for h' and R:

(a) $h'(\underline{A}', \underline{\varphi}', \varepsilon)$ contains only Fourier components with $\underline{\nu} \in \mathcal{M}$, i.e., $(1 - \Pi_{\mathcal{M}})\, h' = 0$.

(b) $R = R_1 + R_2$, where R_1 contains only Fourier components with $\underline{\nu} \notin \mathcal{M}$ and $|\underline{\nu}| > N$, while R_2 is analytic in ε and divisible by ε^{n+1}.

In a word, $\Phi_1, \ldots, \Phi_n$ must be chosen in order to eliminate from the Hamiltonian all the nonresonant harmonics, in the sense of relation (7.4), up to order n in ε.

Developments in ε analogous to (2.6), (2.8) give

$$h\left(\underline{A}'+\frac{\partial\Phi}{\partial\underline{\varphi}}\right)=h(\underline{A})+\sum_{k=1}^{n}\left[\underline{\omega}(\underline{A}')\cdot\frac{\partial\Phi_k}{\partial\underline{\varphi}}+X_k(\underline{A}',\underline{\varphi})\right]\varepsilon^k+\mathcal{O}(\varepsilon^{n+1}) \tag{7.17}$$

with

$$X_k=\begin{cases}0 & \text{for } k=1\\ \sum_{2\leqslant|\underline{m}|\leqslant k} h_{\underline{m}}(\underline{A}')\sum_{\{k_j^i\}_{\underline{m},k}}\prod_{i=1}^{l}\prod_{j=1}^{m_i}\frac{\partial\Phi_{k_j^i}}{\partial\varphi_i} & \text{for } 2\leqslant k\leqslant n\end{cases} \tag{7.18}$$

(for notations see Section 6.1, formula (6.9) in particular). Concerning f, we have, instead,

$$\varepsilon f\left(\underline{A}'+\frac{\partial\Phi}{\partial\underline{\varphi}},\underline{\varphi}\right)=\varepsilon f(\underline{A}',\underline{\varphi})+\sum_{k=2}^{n}\varepsilon^k Y_k(\underline{A}',\underline{\varphi})+\mathcal{O}(\varepsilon^{n+1}) \tag{7.19}$$

with

$$Y_k(\underline{A}',\underline{\varphi})=\sum_{1\leqslant|\underline{m}|\leqslant k-1} f_{\underline{m}}(\underline{A}',\underline{\varphi})\sum_{\{k_j^i\}_{\underline{m},k-1}}\prod_{i=1}^{l}\prod_{j=1}^{m_i}\frac{\partial\Phi_{k_j^i}}{\partial\varphi_i} \tag{7.20}$$

Finally, concerning h' appearing on the r.h.s. of (7.16), let us first write

$$h'(\underline{A}',\underline{\varphi}',\varepsilon)=h(\underline{A}')+\sum_{k=1}^{\infty}\varepsilon^k h_k'(\underline{A}',\underline{\varphi}') \tag{7.21}$$

Then, introducing $\underline{\varphi}$ in place of $\underline{\varphi}'$ as the independent variable, we obtain

$$\begin{aligned}h_k'\left(\underline{A}',\underline{\varphi}+\frac{\partial\Phi}{\partial\underline{A}'}\right)=h_k'(\underline{A}',\underline{\varphi})&+\sum_{r=1}^{\infty}\sum_{1\leqslant|\underline{m}|\leqslant r}h_{k,\underline{m}}'(\underline{A}',\underline{\varphi})\\&\times\sum_{\{k_j^i\}_{\underline{m},r}}\prod_{i=1}^{l}\prod_{j=1}^{m_i}\frac{\partial\Phi_{k_j^i}}{\partial A_i}+\mathcal{O}(\varepsilon^{n+1})\end{aligned} \tag{7.22}$$

We thus have

$$h'\left(\underline{A}',\underline{\varphi}+\frac{\partial\Phi}{\partial\underline{A}'}\right)=h(\underline{A}')+\sum_{k=1}^{n}\varepsilon^k[h_k'(\underline{A}',\underline{\varphi})-Z_k(\underline{A}',\underline{\varphi})] \tag{7.23}$$

with

$$Z_k(\underline{A}',\underline{\varphi})=\begin{cases}0 & \text{for } k=1\\ \sum_{r=1}^{k-1}\sum_{1\leqslant|\underline{m}|\leqslant r}h_{k-r,\underline{m}}'(\underline{A}',\underline{\varphi})\sum_{\{k_j^i\}_{\underline{m},r}}\prod_{i=1}^{l}\prod_{j=1}^{m_i}\frac{\partial\Phi_{k_j^i}}{\partial A_i} & \text{for } 2\leqslant k\leqslant n\end{cases} \tag{7.24}$$

Equation (7.16) is then equivalent to the following equations in the unknowns $\Phi_1,..., \Phi_n$, $h'_1,..., h'_n$:

$$\underline{\omega}(\underline{A}')\cdot\frac{\partial\Phi_k}{\partial\underline{\varphi}}+Q_k(\underline{A}',\underline{\varphi})=h'_k(\underline{A}',\underline{\varphi})+S_k(\underline{A}',\underline{\varphi}),\qquad k=1,...,n \tag{7.25}$$

where

$$Q_k(\underline{A}',\underline{\varphi})=\begin{cases} f(\underline{A}',\underline{\varphi}) & \text{for } k=1\\ X_k(\underline{A}',\underline{\varphi})+Y_k(\underline{A}',\underline{\varphi})-Z_k(\underline{A}',\underline{\varphi}) & \text{for } 2\leqslant k\leqslant n\end{cases} \tag{7.26}$$

The prescription is now that h'_k contains only Fourier components with $\underline{\nu}\in\mathscr{M}$, while S_k must contain Fourier components with $|\underline{\nu}|>N$. Since X_k, Y_k, Z_k depend on $\Phi_{k'}$, $h'_{k'}$ for $k'<k$, Eq. (7.25) are decoupled. To solve them one simply sets

$$\begin{aligned} h'_k(\underline{A}',\underline{\varphi})&=(\Pi_{\mathscr{M}}Q_k)(\underline{A}',\underline{\varphi})\\ S_k(\underline{A}',\underline{\varphi})&=(1-\Pi_{\mathscr{M}})\,Q_k^{>N}(\underline{A}',\underline{\varphi})\end{aligned} \tag{7.27}$$

Φ_k being then determined by Lemma 1 [see Section 9, (9.4)].

7.3. For any positive n and $\delta\leqslant\xi/2n$, let us introduce the following notations:

$$\eta=\frac{\tilde{\rho}}{\xi}\delta$$

$$\left.\begin{aligned}\tilde{\rho}_k&=\tilde{\rho}-k\eta\geqslant\frac{1}{2}\tilde{\rho}\\ \xi_k&=\xi-k\delta\geqslant\frac{1}{2}\xi\end{aligned}\right\}\quad k=1,...,n \tag{7.28}$$

For any $\rho'\leqslant\tilde{\rho}$, $\xi'\leqslant\xi$ we will also use the short notation

$$\|\cdot\|_{\rho',\xi'}\equiv\|\cdot\|_{W(\mathscr{U};\rho',\xi')} \tag{7.29}$$

All the basic estimates are then contained in the following:

Main Proposition. For $k=1,..., n$ one has

$$\left\|\frac{\partial\Phi_k}{\partial\underline{\varphi}}\right\|_{\tilde{\rho}_k,\xi_k}\leqslant\tilde{\rho}FB^kk!\,\delta^{-ks} \tag{7.30a}$$

$$\left\|\frac{\partial\Phi_k}{\partial\underline{A}'}\right\|_{\tilde{\rho}_k,\xi_k}\leqslant\xi FB^kk!\,\delta^{-ks} \tag{7.30b}$$

$$\|h'_{k,\underline{m}}\|_{\tilde{\rho}_k,\xi_k}\leqslant\rho EDB^{k-1}k!\,\delta^{-(k-1)s-|\underline{m}|} \tag{7.30c}$$

where the constants s, B, D, F have to satisfy several inequalities met in the course of the proof; we shall see that, for instance, the following choice will eventually meet all the necessary requirements:

$$
\begin{aligned}
s &= 2(l+1) \\
B &= 2^{l+6} l B_0^2 C_0^2 E^2 \left(\frac{\rho}{\tilde{\rho}}\right)^2 \varepsilon^{-2\sigma} \\
&= 2^{l+6} l B_0^2 \rho^2 C_0^4 E^2 M^2 \varepsilon^{-4\sigma-2\tau} \\
D &= 4l \\
F &= 2^{-l-4} B_0^{-1} C_0^{-1} E^{-1} \delta^{l+1} \varepsilon^{\sigma}
\end{aligned}
\tag{7.31}
$$

Moreover, we have

$$
\|Q_{k+1}\|_{\tilde{\rho}_k, \xi_k} \leqslant 4 l \rho E B^k (k+1)! \, \delta^{-ks} \qquad k = 1, \ldots, n-1 \tag{7.32}
$$

(7.30a–c) are proven by induction, along the same lines of the analogous proof in Section 6, while (7.32) is found as a byproduct of that proof.

Proof. We first verify (7.30a–c) for $k=1$. As $h_1' = \Pi_{\mathcal{M}} f$, by Lemma 6 one finds (after a dimensional estimate)

$$
\|h'_{1,\underline{m}}\|_{\tilde{\rho}, \xi_1} \leqslant \frac{1}{\underline{m}!} \left\| \frac{\partial^{|\underline{m}|} f}{\partial \underline{\varphi}^{\underline{m}}} \right\|_{\tilde{\rho}, \xi} \leqslant \rho E \, \delta^{-|\underline{m}|+1} \tag{7.33}
$$

so that (7.30c) is verified, for $k=1$, because $\delta < 1$ and hence

$$
D \geqslant \delta \tag{7.34}
$$

Concerning (7.30a), we use statement (9.5), (i) of Lemma 1, with $g = (1 - \Pi_{\mathcal{M}}) f$, taking into account

$$
\left\| \frac{\partial}{\partial \underline{\varphi}} (f - \Pi_{\mathcal{M}} f) \right\|_{\tilde{\rho}, \xi} \leqslant 2 \left\| \frac{\partial f}{\partial \underline{\varphi}} \right\|_{\tilde{\rho}, \xi} \leqslant 2 \rho E \tag{7.35}
$$

here the factor 2 arises because $\Pi_{\mathcal{M}} f$ can be expressed as a suitable average of f, see (9.15). It follows that

$$
\left\| \frac{\partial \Phi_1}{\partial \underline{\varphi}} \right\|_{\tilde{\rho}, \xi_1} \leqslant 2 B_0 C \rho E \, \delta^{-1} \tag{7.36}
$$

which implies (7.30a) for $k=1$, provided [as our choices (7.31) imply]

$$
F B \, \delta^{-s} \geqslant 2 B_0 C E \frac{\rho}{\tilde{\rho}} \delta^{-l} \tag{7.37}
$$

Finally, concerning (7.30b), we need a preliminary estimate of $\|f-\Pi_{\mathcal{M}} f\|_{\tilde{\rho},\xi}$. By Lemma 6, it is not difficult to obtain

$$\|f-\Pi_{\mathcal{M}} f\|_{\tilde{\rho},\xi} \leqslant 2l \sup_{\underline{\varphi}_1,\underline{\varphi}_2} \|\underline{\varphi}_1-\underline{\varphi}_2\| \left\|\frac{\partial f}{\partial \underline{\varphi}}\right\|_{\tilde{\rho},\xi} \tag{7.38}$$

where the sup is in $0 \leqslant \operatorname{Re} \varphi_j \leqslant \pi$, $|\operatorname{Im} \varphi_j| \leqslant \xi$, $j=1,..., l$. It follows that

$$\|f-\Pi_{\mathcal{M}} f\|_{\tilde{\rho},\xi} \leqslant 2l \sqrt{\pi^2+\xi^2}\, \rho E < 2^3 l \rho E \tag{7.39}$$

since $\xi \leqslant 1$, and consequently, by statement (9.5), (iv) of Lemma 1:

$$\left\|\frac{\partial \Phi_1}{\partial \underline{A}'}\right\|_{\tilde{\rho},\xi} \leqslant 2^3 l B_0 C \rho E\, \delta^{-l} \eta^{-1} \tag{7.40}$$

which gives (7.30b) for $k=1$, if we impose [see also (7.28a) to eliminate η]:

$$FB\, \delta^{-s} \geqslant 2^3 l B_0 C E \frac{\rho}{\tilde{\rho}}\, \delta^{-l-1} \tag{7.41}$$

Now we assume (7.30a–c) for $k \leqslant p < n$, and prove them for $k=p+1$. Using the induction hypothesis, dimensional estimates like (6.14), Lemmas 3 and 4, and imposing in the various steps the following constraints

$$\begin{gathered} 8F\frac{\tilde{\rho}}{\rho} < 1 \\ 4F\xi\, \delta^{-1} < 1 \\ 4FB\, \delta^{-s+1} \geqslant D\left(\frac{\tilde{\rho}}{\rho}\right)^2 \end{gathered} \tag{7.42}$$

one finds after some computations

$$\|X_{p+1}\|_{\tilde{\rho}_p,\xi_p} \leqslant 2^{l+3} \tilde{\rho}^2 \rho^{-1} E F^2 B^{p+1} (p+1)!\, \delta^{-(p+1)s} \tag{7.43a}$$

$$\|Y_{p+1}\|_{\tilde{\rho}_p,\xi_p} \leqslant 2^{l+1} \tilde{\rho} E F B^p p!\, \delta^{-ps} \tag{7.43b}$$

$$\|Z_{p+1}\|_{\tilde{\rho}_p,\xi_p} \leqslant 2^{l+1} \rho E D F B^p (p+1)!\, \delta^{-ps-1} \tag{7.43c}$$

$$\|Q_{p+1}\|_{\tilde{\rho}_p,\xi_p} \leqslant 2^{l+4} \tilde{\rho}^2 \rho^{-1} E F^2 B^{p+1} (p+1)!\, \delta^{-(p+1)s} \tag{7.43d}$$

[according to (7.26), (7.43d) is simply the sum of (7.43a–c), taking also into account (7.42c)]. Lemma 1 now can be applied to estimate Φ_k and h'_k, which are known to satisfy (7.25); one can see that (7.30a–c) for $k=p+1$ are obtained, if one further assumes

$$F^2 B\,\delta^{-s} \leqslant 2^{-l-4} D\left(\frac{\rho}{\tilde{\rho}}\right)^2 2^{l+5} B_0 CEF \frac{\tilde{\rho}}{\rho}\,\delta^{-l-1} \tag{7.44}$$

The Main Proposition is thus proven, if the assumptions (7.34), (7.37), (7.41), (7.42), and (7.44) are satisfied. A consistent and convenient choice of constants s, B, D, F is given by (7.31); by means of these expressions, (7.43d) gives in particular (7.32), as claimed.

7.4. We now fix $\delta=\xi/2n$, so that we have $\xi_n=\frac{1}{2}\xi$, $\tilde{\rho}_n=\frac{1}{2}\tilde{\rho}=\frac{1}{2}M^{-1}C_0^{-1}\varepsilon^{\sigma+\tau}$, and look for global estimates on the canonical transformation.

From (7.30a, b), in order to be able to estimate the sum (7.15) over k, we need to impose $nB\,\delta^{-s}\varepsilon \leqslant \frac{1}{2}$, i.e., by (7.6), (7.31):

$$2^{s+l+7} l B_0^2 \rho^2 C_0^4 E^2 M^2 \xi^{-s} n^{s+1} \varepsilon^{1-4\sigma-2\tau} < 1 \tag{7.45}$$

As in Section 6, we satisfy this condition by choosing n to be a suitable power of ε: precisely, we set

$$\varepsilon^{-b} \leqslant n < \varepsilon^{-b}+1 \qquad b=\frac{\beta}{s+1} \tag{7.46}$$

β being any number between zero and one. After this choice, condition (7.45) takes the form

$$2^{2s+l+8} l B_0^2 \rho^2 C_0^4 E^2 M^2 \xi^{-s} \varepsilon^{1-4\sigma-2\tau-\beta} \leqslant 1 \tag{7.47}$$

or [see (7.31) for the value of s] $\varepsilon \leqslant \varepsilon_0$, with

$$\varepsilon_0 = \left(2^{5l+13} l B_0^2 \rho^2 C_0^4 E^2 M^2 \xi^{-2(l+1)}\right)^{-1/(1-4\sigma-2\tau-\beta)} \tag{7.48}$$

Once this condition is satisfied, one easily gets, in particular, the following estimates:

$$\left\|\frac{\partial \Phi}{\partial \underline{\varphi}}\right\|_{(1/2)\tilde{\rho},(1/2)\xi} \leqslant 2\tilde{\rho} FB\,\delta^{-s}\varepsilon \leqslant K\frac{\tilde{\rho}}{C_0 E}\,\varepsilon_0^{\sigma}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-3\sigma-2\tau-(1/2)\beta} \tag{7.49a}$$

$$\left\|\frac{\partial \Phi}{\partial \underline{A}'}\right\|_{(1/2)\tilde{\rho},(1/2)\xi} \leqslant 2\xi FB\,\delta^{-s}\varepsilon \leqslant K\frac{\xi}{C_0 E}\,\varepsilon_0^{\sigma}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-3\sigma-2\tau-(1/2)\beta} \tag{7.49b}$$

$$\max_{i,j}\left\|\frac{\partial^2 \Phi}{\partial A_i\,\partial \varphi_j}\right\|_{(1/4)\tilde{\rho},(1/2)\xi} \leqslant 8FB\,\delta^{-s}\varepsilon \leqslant K\frac{4}{C_0 E}\,\varepsilon_0^{\sigma}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1-3\sigma-2\tau-(1/2)\beta} \tag{7.49c}$$

with

$$K < 2^{-3l-8}\xi^l B_0^{-1} \tag{7.50}$$

(7.49c) implies in particular the invertibility of the canonical transformation, as far as $\varepsilon \leqslant \varepsilon_0$, because $\sum_{k=1}^{l} |\partial^2\Phi/\partial\varphi_k\, \partial \underline{A}'_j| \leqslant 4Kl < \frac{1}{8}$, and one can simply apply a simple implicit function theorem, of the type of the one in Ref. 10, for holomorphic function problems. In particular, the domain of definition of $\mathscr{C}_\varepsilon, \mathscr{C}_\varepsilon^{-1}$ contains $W_\infty(\mathscr{U}; \frac{1}{8}\tilde{\rho}, \frac{1}{4}\xi)$, and in W_∞ they satisfy the inequalities (7.11), as claimed.

Let us now study $h'(\underline{A}', \varphi', \varepsilon) = h(\underline{A}') + \varepsilon G(\underline{A}', \varphi', \varepsilon)$, where

$$G(\underline{A}', \varphi', \varepsilon) = \sum_{k=1}^{n} \varepsilon^{k-1} h'_k(\underline{A}', \varphi', \varepsilon) \tag{7.51}$$

According to (7.27), we have

$$G(\underline{A}', \varphi', \varepsilon) = (\Pi_{\mathscr{M}} f)(\underline{A}', \varphi') + \sum_{k=2}^{n} \varepsilon^{k-1} (\Pi_{\mathscr{M}} Q_k)(\underline{A}', \varphi') \tag{7.52}$$

Now, given $(\underline{A}'_1, \varphi'_1)$ and $(\underline{A}'_2, \varphi'_2)$ both in W_∞, it follows that

$$\begin{aligned} |G(\underline{A}'_2, \varphi'_2, \varepsilon) - G(\underline{A}'_1, \varphi'_1, \varepsilon)| &\leqslant l(\|\underline{A}'_2 - \underline{A}'_1\| + \rho\, \|\varphi'_2 - \varphi'_1\|)\, E \\ &\quad + 2 \sum_{k=2}^{n} \varepsilon^{k-1} \|Q_k\|_{(1/2)\tilde{\rho}, (1/2)\xi} \\ &\leqslant lE(10\rho + \|\underline{A}'_2 - \underline{A}'_1\|) \end{aligned} \tag{7.53}$$

where expression (7.32) for Q_k has been used, together with condition (7.45).

7.5. We complete here the proof of the analytic lemma, by estimating the remainder R. By construction, this term has the form $R_1 + R_2$, where

$$R_1 = \sum_{k=1}^{\infty} \varepsilon^k (Q_k - \Pi_{\mathscr{M}} Q_k)^{>N} \tag{7.54}$$

while R_2 is known to be analytic in ε and divisible by ε^{n+1}.

From Lemma 2, with $g = Q_k - \Pi_{\mathscr{M}} Q_k$, $\frac{1}{2}\xi$ in place of ξ and $\frac{1}{4}\xi$ in place of δ, one easily finds

$$\begin{aligned} \|R_1\|_{(1/2)\tilde{\rho}, (1/4)\xi} &\leqslant 2^{4l+4} l E \rho \xi^{-l} e^{-(1/8)\xi N} \varepsilon \\ &\leqslant \frac{1}{4} \frac{\xi}{C_0^2 M} \varepsilon_0^{2\sigma+\tau} \left(\frac{\varepsilon}{\varepsilon_0}\right) e^{-(1/8)\xi N} \end{aligned} \tag{7.55}$$

Concerning R_2, we proceed as in the proof of Proposition 1 (Section 6.4), i.e., we use the holomorphy in ε, for ε complex and $|\varepsilon| < \varepsilon_0$, and the maximum principle, and write

$$|R_2(\underline{A}', \underline{\varphi}', \varepsilon)| \leqslant \left|\frac{\varepsilon}{\varepsilon_0}\right|^{n+1} \sup_{|\varepsilon'| \leqslant \varepsilon_0} |R_2(\underline{A}', \underline{\varphi}', \varepsilon')| \tag{7.56}$$

On the other hand, R_2 is defined by

$$R_2(\underline{A}', \underline{\varphi}', \varepsilon) = h\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}\right) + \varepsilon f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi}\right)$$
$$- h'\left(\underline{A}', \underline{\varphi} + \frac{\partial \Phi}{\partial \underline{A}'}, \varepsilon\right) - R_1(\underline{A}', \underline{\varphi}, \varepsilon) \tag{7.57}$$

and by the choice of Φ we made [see (7.25)–(7.27)] it follows that

$$R_2(\underline{A}', \underline{\varphi}', \varepsilon) = h\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}\right) - h(\underline{A}') - \underline{\omega}(\underline{A}') \cdot \frac{\partial \Phi}{\partial \underline{\varphi}}$$
$$+ \varepsilon f\left(\underline{A}' + \frac{\partial \Phi}{\partial \underline{\varphi}}, \underline{\varphi}\right) - \varepsilon f(\underline{A}', \underline{\varphi})$$
$$+ \varepsilon G(\underline{A}', \underline{\varphi}) - \varepsilon G\left(\underline{A}', \underline{\varphi} + \frac{\partial \Phi}{\partial \underline{A}'}\right)$$
$$- \sum_{k=2}^{n} \varepsilon^k Q_k(\underline{A}', \underline{\varphi}) \tag{7.58}$$

Each of these lines can be easily estimated; recalling that, for $|\mathrm{Im}\ \varphi_j'| < \frac{1}{4}\xi$ one certainly has $|\mathrm{Im}\ \varphi_j| < \frac{1}{2}\xi$, $j = 1, \ldots, l$, one obtains

$$\|R_2\|_{(1/2)\bar{\rho}, (1/4)\xi} \leqslant 2E \left\|\frac{\partial \Phi}{\partial \underline{\varphi}}\right\|_{(1/2)\bar{\rho}, (1/2)\xi} + \varepsilon_0 E \left\|\frac{\partial \Phi}{\partial \underline{\varphi}}\right\|_{(1/2)\bar{\rho}, (1/2)\xi}$$
$$+ 16 l \rho E \varepsilon_0 + 4 l \rho E \varepsilon_0$$
$$\leqslant \frac{1}{4} \frac{\xi}{C_0^2 M} \varepsilon_0^{2\sigma + \tau} \tag{7.59}$$

Consequently, from (7.56), one has

$$\|R_2\|_{(1/2)\bar{\rho}, (1/4)\xi} \leqslant \frac{1}{4} \frac{\xi}{C_0^2 M} \varepsilon_0^{2\sigma + \tau} \left(\frac{\varepsilon}{\varepsilon_0}\right)^{n+1} \tag{7.60}$$

Estimates (7.55) and (7.60), with n satisfying (7.46), give (7.10).

The analytic lemma is thus proven.

8. THE GEOMETRIC PART OF PROPOSITION 2

8.1. We conclude here the proof of Proposition 2; our aim is to use the analytic lemma of the previous section, together with the convexity property of the unperturbed Hailtonian $h(\underline{A})$, in order to obtain a good confinement of motions in the action space.

Given any resonance plane $\mathcal{M} \subset \mathbf{Z}^l$, dim $\mathcal{M} = r$, denote by $\Lambda_{\mathcal{M}}(\underline{A}) \subset \mathbf{R}^l$ the r-dimensional plane parallel to $\mathcal{M}$ through $\underline{A}$. Consider then any real initial datum $(\underline{A}(0), \underline{\varphi}(0)) \in \mathscr{B}_{\mathcal{M}}$. As long as $\underline{A}(t)$ remains in $\mathscr{U}_{\mathcal{M}}$ (this question will be discussed later), we can use the analytic lemma, with $C_0^{-1} = \lambda_r^{(0)}$ and $\sigma = \sigma_r$, obtaining Hamiltonian (7.7) for the new variables $\underline{A}'$, $\underline{\varphi}'$. The form of this Hamiltonian gives us a basic information: indeed, from the expression (7.8) of G, it follows that $\partial G/\partial \underline{\varphi}'$ is a linear combination of vectors $\underline{\nu} \in \mathcal{M}$, and consequently

$$\begin{aligned} \underline{A}'(t) &= \underline{A}'(0) - \int_0^t \left[\varepsilon \frac{\partial G}{\partial \underline{\varphi}'} (\underline{A}'(t'), \underline{\varphi}'(t'), \varepsilon) + \frac{\partial R}{\partial \underline{\varphi}'} (\underline{A}'(t'), \underline{\varphi}'(t'), \varepsilon) \right] dt' \\ &= \underline{A}''(t) - \int_0^t \frac{\partial R}{\partial \underline{\varphi}'} (\underline{A}'(t'), \underline{\varphi}'(t'), \varepsilon)\, dt', \qquad \underline{A}''(t) \in \Lambda_{\mathcal{M}}(\underline{A}'(0)) \end{aligned} \tag{8.1}$$

so that

$$\begin{aligned} \operatorname{dist}(\underline{A}'(t), \Lambda_{\mathcal{M}}(\underline{A}'(0)) &\leqslant |t| \left\| \frac{\partial R}{\partial \underline{\varphi}'} \right\| \leqslant |t| \frac{\lambda_\tau^{(0)2} \varepsilon_0^{2\sigma_r + \tau}}{M} \left(\frac{\varepsilon}{\varepsilon_0} \right) \chi \\ \chi &\equiv \left[e^{-(1/8)\xi \varepsilon^{-\tau}} + \left(\frac{\varepsilon}{\varepsilon_0} \right)^{\varepsilon^{-b}} \right] \end{aligned} \tag{8.2}$$

where $\partial R/\partial \underline{\varphi}'$ has been dimensionally estimated from (7.10), for real values of the variables.

In the particular case $r = 0$, one obtains $\partial G/\partial \underline{\varphi}' = 0$, and one has simply

$$\|\underline{A}'(t) - \underline{A}'(0)\| \leqslant |t| \frac{\lambda_r^{(0)2} \varepsilon_0^{2\sigma_r + \tau}}{M} \left(\frac{\varepsilon}{\varepsilon_0} \right) \chi \tag{8.3}$$

For $r \geqslant 1$, expression (8.2) shows that the motion $\underline{A}'(t)$ is essentially "flat" on the plane $\Lambda_{\mathcal{M}}(\underline{A}'(0))$, up to times exponentially long in ε^{-1}. A good control on $\|\underline{A}'(t) - \underline{A}'(0)\|$ is then achieved by the conservation of energy, combined with the convexity property (2.9) of h. The idea is quite simple: as $\underline{\omega}(\underline{A}') \cdot \underline{\nu} = 0$ for $\underline{A}' \in \Sigma_{\mathcal{M}}$ and $\underline{\nu}$ parallel to $\mathcal{M}$, from (2.9) it follows that the "kinetic energy" $h(\underline{A}')$, restricted to $\Lambda_{\mathcal{M}}(\underline{A}'(0))$, has a quadratic minimum in the point of intersection of $\Sigma_{\mathscr{M}}$ and $\Lambda_{\mathscr{M}}$; then, taking into account the bound (7.9) for G, it follows that $\underline{A}'(t)$ is confined in a convenient vicinity of such point.

Formally, we can proceed as follows: by a Taylor expansion around $\underline{A}'(0)$, we obtain

$$h(\underline{A}'(t)) = h(\underline{A}'(0)) + \underline{\omega}(\underline{A}'(0)) \cdot (\underline{A}'(t) - \underline{A}'(0)) + \frac{1}{2}\left(\frac{\partial^2 h}{\partial \underline{A}\, \partial \underline{A}}\right)_{\underline{A}^*(t)} (\underline{A}'(t) - \underline{A}'(0))(\underline{A}'(t) - \underline{A}'(0)) \tag{8.4}$$

where $\underline{A}^*(t)$ is a suitable point of the segment joining $\underline{A}'(0)$ and $\underline{A}'(t)$. From (2.9) it then follows that

$$\tfrac{1}{2} m \|\underline{A}'(t) - \underline{A}'(0)\|^2 \leqslant |h(\underline{A}'(t)) - h(\underline{A}'(0))| + |\underline{\omega}(\underline{A}'(0)) \cdot (\underline{A}'(t) - A'(0))| \tag{8.5}$$

Denote $a = \|\underline{A}'(t) - \underline{A}'(0)\|$; from conservation of energy it follows immediately that

$$\begin{aligned} &|h(\underline{A}'(t)) - h(\underline{A}'(0))| \\ &\quad \leqslant \varepsilon |G(\underline{A}'(t), \underline{\varphi}'(t)) - G(\underline{A}'(0), \underline{\varphi}'(0))| + 2\|R\|_{W_\infty} \\ &\quad \leqslant lE(10\rho + a)\,\varepsilon + \frac{\xi \lambda_r^{(0)^2} \varepsilon_0^{2\sigma_r + \tau}}{2M} \left(\frac{\varepsilon}{\varepsilon_0}\right) \chi \end{aligned} \tag{8.6}$$

Concerning the last term of (8.5), let us first notice that, as $\underline{A}(0) \in \mathscr{B}_{\mathscr{M}}$, then for an N-basis $\{\underline{y}_1, ..., \underline{y}_r\}$ of $\mathscr{M}$ one has

$$\begin{aligned} |\underline{\omega}(\underline{A}'(0)) \cdot \underline{y}_j| &\leqslant |\underline{\omega}(\underline{A}(0)) \cdot \underline{y}_j| + MN\|\underline{A}'(0) - \underline{A}(0)\| \\ &\leqslant 2\lambda_r^{(0)} \varepsilon^{\sigma_r} \end{aligned} \tag{8.7}$$

as follows trivially from (2.9), (3.4), (7.11).

Denote by $\tilde{\underline{\omega}}$ the orthogonal projection of $\underline{\omega}(\underline{A}'(0))$ on $\mathscr{M}$; from (8.3) and the geometric Lemma 7 of Section 9, we have $\|\tilde{\underline{\omega}}\| \leqslant 2rN^{r-1} \lambda_r^{(0)} \varepsilon^{\sigma_r}$, and thus

$$\begin{aligned} &|\underline{\omega}(\underline{A}'(0)) \cdot (\underline{A}'(t) - \underline{A}'(0))| \\ &\quad \leqslant l \|\tilde{\underline{\omega}}\| \, \|\underline{A}'(t) - \underline{A}'(0)\| + l \|\underline{\omega}(\underline{A}'(0))\| \text{ dist}(\underline{A}'(t), \Lambda_{\mathscr{M}}(\underline{A}'(0))) \\ &\quad \leqslant 2l^2 \lambda_r^{(0)} \varepsilon^{\sigma_r - (r-1)\tau} a + lE|t| \frac{\lambda_r^{(0)^2} \varepsilon_0^{2\sigma_r + \tau}}{M} \left(\frac{\varepsilon}{\varepsilon_0}\right) \chi \end{aligned} \tag{8.8}$$

From (8.5) it then follows that

$$\tfrac{1}{2}ma^2 \leqslant [lE\varepsilon + 2l^2\lambda_r^{(0)}\varepsilon^{\sigma_r-(r-1)\tau}]\, a + 10lE\rho\varepsilon + \left(\frac{\xi}{2} + lE|t|\right)\frac{\lambda_r^{(0)2}}{M}\varepsilon_0^{2\sigma_r+\tau}\left(\frac{\varepsilon}{\varepsilon_0}\right)\chi \tag{8.9}$$

and if t is bounded by

$$|t| \leqslant 2E^{-1}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{-(1-2\sigma_r)}\left[e^{-(1/8)\xi\varepsilon^{-r}} + \left(\frac{\varepsilon}{\varepsilon_0}\right)^{\varepsilon^{-b}}\right]^{-1} \tag{8.10}$$

we easily obtain

$$\begin{aligned}
&\tfrac{1}{2}ma^2 \leqslant Xa + Y\\
&a \leqslant \frac{X}{m}[1 + \sqrt{1 + 2mY/X^2}]\\
&X \equiv (2l^2 + 1)\,\lambda_r^{(0)}\varepsilon^{\sigma_r-(r-1)\tau}\\
&Y \equiv 10lE\rho\varepsilon + 3l\frac{\lambda_r^{(0)2}}{M}\varepsilon_0^{\tau}\varepsilon^{2\sigma_r} < 4l\frac{\lambda_r^{(0)2}}{M}\varepsilon_0^{\tau}\varepsilon^{2\sigma_r}
\end{aligned} \tag{8.11}$$

From (8.11), with $r \geqslant 1$, it easily follows that $mY/X^2 < 1$, and thus

$$a < \frac{3X}{m} = (6l^2 + 3)\frac{\lambda_r^{(0)}}{m}\varepsilon^{\sigma_r-(r-1)\tau} \tag{8.12}$$

However, from (8.3) we easily check that this inequality is also true for $r = 0$, if $|t|$ is bounded by (8.10).

Finally, if we take into account $\|\underline{A}' - \underline{A}\| \leqslant \frac{1}{4}(\lambda_r^{(0)}/m)\,\varepsilon^{\sigma_r+\tau}$, which trivially follows from (7.11), we immediately obtain, for the old variables,

$$\|\underline{A}(t) - \underline{A}(0)\| < (6l^2 + 4)\frac{\lambda_r^{(0)}}{m}\varepsilon^{\sigma_r-(r-1)\tau} \tag{8.13}$$

8.2. As remarked above, all of these considerations are valid as far as the analytic lemma can be applied, i.e., as far as $\underline{A}(t) \in \mathscr{U}_{\mathscr{M}}$. The problem is thus to guarantee, for t satisfying (8.10), condition (7.4), i.e.,

$$|\underline{\omega}(\underline{A}(t)) \cdot \underline{\nu}| \geqslant 2\lambda_r^{(0)}\varepsilon^{\sigma_r}, \qquad \forall \underline{\nu} \notin \mathscr{M},\ |\underline{\nu}| \leqslant N \tag{8.14}$$

This will be done by imposing some quite strict conditions on the constants $\lambda_0^{(0)},\ldots, \lambda_l^{(0)}$, $\sigma_0,\ldots, \sigma_l$, which up to now were free and independent of each other, as we always worked separately in the different domains $\mathscr{U}_{\mathscr{M}}$.

Clearly, this is nothing but a consistency problem: indeed, to guarantee (8.14) it is sufficient to impose that this condition cannot be violated within a distance D_r from $\underline{A}(0)$, D_r being given by the r.h.s. of (8.13).

To this purpose, consider the partition of V_R into resonant blocks, introduced in Section 3, and suppose $\underline{A}(0) \in \mathscr{B}_{\mathscr{M}}$, dim $\mathscr{M} = r$. Using (3.3), we obtain for all $\underline{v} \notin \mathscr{M}$, $|\underline{v}| \leqslant N$,

$$|\underline{\omega}(\underline{A}(t)) \cdot \underline{v}| > |\underline{\omega}(\underline{A}(0)) \cdot \underline{v}| - MND_r$$

$$\geqslant \lambda_{r+1}^{(0)} \varepsilon^{\sigma_{r+1}} - (6l^2 + 4) \frac{M}{m} \lambda_r^{(0)} \varepsilon^{\sigma_r - r\tau} \tag{8.15}$$

(8.14) is then achieved, as long as one has

$$\begin{aligned} &\sigma_{r+1} \leqslant \sigma_r - r\tau \\ &\lambda_{r+1}^{(0)} \geqslant \left(2 + (6l^2 + 4) \frac{M}{m}\right) \lambda_0^{(0)} \qquad r = 0, \ldots, l-1 \end{aligned} \tag{8.16}$$

A simple choice is

$$\begin{aligned} &\sigma_r = \sigma_0 - \frac{r(r-1)}{2} \tau \\ &\lambda_r^{(0)} = \left(\frac{8l^2 M}{m}\right)^r \lambda_0^{(0)} \\ &\tau < \frac{2}{l(l-1)} \sigma_0 \end{aligned} \tag{8.17}$$

We need then a choice for τ, σ_0, and $\lambda_0^{(0)}$, as well as for the constant β we left free in the analytic part. The only limitations we have, beside (8.17), are

$$\begin{aligned} &\left(\frac{8l^2 M}{m}\right)^l \lambda_0^{(0)} E^{-1} \leqslant 1 \\ &4\sigma_0 + 2\tau + \beta < 1 \end{aligned} \tag{8.18}$$

which assure both condition $C_0 E > 1$ and condition (7.3) entering the analytic lemma. A simple choice is

$$\begin{aligned} &\lambda_0^{(0)} = \left(\frac{m}{8l^2 M}\right)^l E \\ &\tau = \frac{1}{l(l+1)} \sigma_0 \\ &\beta = \frac{2l+3}{l(l+1)} \sigma_0 \\ &\sigma_0 < \frac{1}{6} \end{aligned} \tag{8.19}$$

which gives $b = \tau$, and $4\sigma_0 + 2\tau + \beta < 6\sigma_0 < 1$ for any $l \geqslant 2$.

After this choice, (8.10) can be replaced by

$$|t| \leqslant T \equiv E^{-1} \left(\frac{\varepsilon}{\varepsilon_0}\right)^{-(1-2\sigma_0)} e^{(1/8)\xi\varepsilon^{-b}} \tag{8.20}$$

where $\varepsilon \leqslant \frac{1}{2}\varepsilon_0$ has also been used.

Estimate (8.13) gives then in particular for any r

$$\|\underline{A}(t) - \underline{A}(0)\| < (6l^2 + 4)\frac{E}{m}\varepsilon^{(1/2)\sigma_0} \tag{8.21}$$

Finally, for any r one has

$$\varepsilon_0 \geqslant \left[B_l^{-1} \left(\frac{E}{\rho M}\right)^2 \left(\frac{m}{M}\right)^{4l} \xi^{-2(l+1)} \right]^{1/(1-6\sigma_0)} \tag{8.22}$$

with

$$B_l = 2^{18l+13} l^{8l+1} B_0^2 < 2^{22l+19} l^{8l+1} \tag{8.23}$$

Now, for $\sigma_0 = \frac{1}{8}$, as in (3.6), inequalities (8.20) and (8.21) give (3.7), while the estimates of Part (ii) of Proposition 2 are contained in the analytic lemma of Section 7. The proof of Proposition 2 is thus concluded.

9. LEMMAS

In this section we recall a few elementary lemmas, which are used throughout the paper.

Lemma 1. Let $\mathscr{M}$ be an r-dimensional subspace of $\mathbf{Z}^l$. Suppose that the function

$$g(\underline{A}, \underline{\varphi}) = \sum_{\underline{\nu} \notin \mathscr{M}} g_{\underline{\nu}}(\underline{A})\, e^{i\underline{\nu}\cdot\underline{\varphi}} \tag{9.1}$$

is analytic for $(\underline{A}, \underline{\varphi}) \in W(\mathscr{U}; \rho, \xi)$, $\mathscr{U}$ being an open subset of $\mathbf{R}^l$; let $\underline{\omega}(\underline{A})$ satisfy for any $\underline{A}$ the diophantine condition

$$|\underline{\omega}(\underline{A}) \cdot \underline{\nu}|^{-1} < C|\underline{\nu}|^{\alpha} \qquad \forall \underline{\nu} \notin \mathscr{M},\ |\underline{\nu}| \leqslant N \tag{9.2}$$

(N could also be infinite), and denote

$$g^{\leqslant N} = \sum_{\substack{\underline{\nu} \notin \mathscr{M} \\ |\underline{\nu}| \leqslant N}} g_{\underline{\nu}}(\underline{A})\, e^{i\underline{\nu}\cdot\underline{\varphi}} \tag{9.3}$$

Then the equation for $\Phi(\underline{A}, \underline{\varphi})$

$$\underline{\omega}(\underline{A}) \cdot \frac{\partial \Phi}{\partial \underline{\varphi}} (\underline{A}, \underline{\varphi}) + g^{\leqslant N}(\underline{A}, \underline{\varphi}) = 0 \tag{9.4}$$

can be solved, and for any positive $d < \xi$ and $\eta < \rho$ one has the estimates

$$\begin{aligned}
&\text{(i)} && \|\Phi\|_{\rho, \xi - \delta} < \|g\|_{\rho, \xi} B_0 C \delta^{-l-\alpha} \\
&\text{(ii)} && \left\| \frac{\partial \Phi}{\partial \underline{\varphi}} \right\|_{\rho, \xi - \delta} < \left\| \frac{\partial g}{\partial \underline{\varphi}} \right\|_{\rho, \xi} B_0 C \delta^{-l-\alpha} \\
&\text{(iii)} && \left\| \frac{\partial \Phi}{\partial \underline{\varphi}} \right\|_{\rho, \xi - \delta} < \|g\|_{\rho, \xi} B_0 C \delta^{-l-\alpha-1} \\
&\text{(iv)} && \left\| \frac{\partial \Phi}{\partial \underline{A}'} \right\|_{\rho - \eta, \xi - \delta} < \|g\|_{\rho, \xi} B_0 C \delta^{-l-\alpha} \eta^{-1}
\end{aligned} \tag{9.5}$$

The constant B_0 depends on l and α, and can be taken to be

$$B_0 = 2^{2l+\alpha+1} e^{\alpha+1} (\alpha+1)^{\alpha+1} < 2^{2l+3\alpha+3} (\alpha+1)^{\alpha+1} \tag{9.6}$$

The proof is based on elementary dimensional estimates; for essentially the same estimates, see, for example, Ref. 10. [Notice that, for (9.1), the average of g over $\underline{\varphi}$ is assumed to vanish, even for $r = 0$].

Lemma 2. If $g(\underline{A}, \underline{\varphi})$ is analytic in $W(U; \rho, \xi)$, then the quantity

$$g^{>N}(\underline{A}, \underline{\varphi}) \equiv \sum_{|\underline{\nu}| > N} g_{\underline{\nu}}(\underline{A}) \, e^{i\underline{\nu} \cdot \underline{\varphi}} \tag{9.7}$$

satisfies the estimate

$$\begin{aligned}
\|g^{>N}\|_{\rho, \xi - \delta} &\leqslant \left(\frac{1 + e^{-(1/2)\delta}}{1 - e^{-(1/2)\delta}} \right)^l e^{-(1/2)\delta N} \|g\|_{\rho, \xi} && \forall \delta > 0 \\
&\leqslant 2^{2l} \delta^{-l} e^{-(1/2)\delta N} \|g\|_{\rho, \xi} && 0 < \delta \leqslant \tfrac{2}{3}
\end{aligned} \tag{9.8}$$

The proof is based on the elementary property of analytic functions of having exponentially decreasing Fourier coefficients; see, for example, Ref. 10.

Lemma 3. For any integers k, s, with $1 \leqslant s \leqslant k$, one has

$$\sum_{\substack{k_1, \ldots, k_s \geqslant 1 \\ k_1 + \cdots + k_s = k}} \prod_{i=1}^{s} k_i! \leqslant k! \tag{9.9}$$

The proof is made by induction, and left to the reader.

Lemma 4. The number $J(l, p)$ of integer l-tuples $\underline{m} = (m_1,..., m_l) \in \mathbf{Z}^l$, with $m_i \geqslant 0$ and $\sum_{i=1}^{l} m_i = p$, is

$$J(l, p) = \frac{(l-1+p)!}{(l-1)!\, p!} < 2^{l+p-1} \tag{9.10}$$

The result follows directly from counting.

Lemma 5. Let $\mathscr{M}$ be an r-dimensional subspace of $\mathbf{Z}^l$, generated by the integer vectors $\underline{v}_1,..., \underline{v}_r$. Then there exists an $l \times l$ matrix J, with $\det J = 1$, and integer entries, such that $\underline{v}'_i \equiv J\underline{v}_i$, $i = 1,..., r$, satisfy $(\underline{v}'_i)_j = 0$ for $j = i+1,..., l$, so that one has in particular

$$\mathscr{M}' \equiv JM = \{\underline{v}' \mid \underline{v}' \in \mathbf{Z}^l;\ v'_{r+1},..., v'_l = 0\} \tag{9.11}$$

Proof. For $l = 2$, $r = 1$, let $\underline{\mu} \in \mathbf{Z}^2$ be parallel to $\underline{v}_1$, μ_1 and μ_2 having no common divisor. One can always find integers m and n such that $m\mu_1 + n\mu_2$ equals 1 (or any other preassigned integer); one can then take

$$J = \begin{pmatrix} m & n \\ -\mu_2 & \mu_1 \end{pmatrix} \tag{9.12}$$

For general l and $r = 1$ one clearly obtains $(\underline{v}'_1)_j = 0$ for $j = 2,..., l$ by solving $l-1$ times the two-dimensional problem in the subspaces $(1, j)$, $j = 2,..., l$.

Suppose now one has already found $\hat{J}$ which gives $(\underline{v}'_i)_j = 0$ for $i = 1,..., r-1$, $j = i+1,..., l$. One can then choose J of the form

$$J = \begin{pmatrix} I & 0 \\ 0 & \tilde{J} \end{pmatrix} \hat{J} \tag{9.13}$$

I being the identity on the first $r-1$ components; indeed, such a matrix leaves unchanged $\underline{v}_1,..., \underline{v}_{r-1}$, while, as shown above, one can always find a $(l-r+1) \times (l-r+1)$ matrix $\tilde{J}$ with determinant one, such that the $(l-r+1)$-dimensional vector $\tilde{J}((\underline{v}_r)_r,..., (\underline{v}_r)_l)$ has, as required, all the components but the first which vanish.

Lemma 6. Let $\mathscr{M}$ be an r-dimensional subspace of $\mathbf{Z}^l$; given $f(\underline{\varphi}) = \sum_{\underline{v} \in \mathbf{Z}^l} f_{\underline{v}} e^{i\underline{v} \cdot \underline{\varphi}}$, denote

$$\Pi_{\mathscr{M}} f = \sum_{\underline{v} \in \mathscr{M}} f_{\underline{v}} e^{i\underline{v} \cdot \underline{\varphi}} \tag{9.14}$$

Then we have

$$\Pi_{\mathscr{M}} f = (2\pi)^{r-l} \int f(J^T \underline{\varphi}')\, d\varphi'_{r+1} \cdots d\varphi'_l \tag{9.15}$$

J being the integer matrix introduced in Lemma 5.

Proof. Let $\varphi = J^T \varphi'$, $f'(\varphi') = f(J^T \varphi')$. As $\det J = 1$, J is an invertible mapping of $\mathbf{T}^l$ into itself; it follows that

$$\begin{aligned}\Pi_{\mathcal{M}} f &= \sum_{\underline{\nu}' \in \mathcal{M}'} f_{J^{-1}\underline{\nu}'} e^{iJ^{-1}\underline{\nu}' \cdot \underline{\varphi}} = \sum_{\underline{\nu}' \in \mathcal{M}'} f_{J^{-1}\underline{\nu}'} e^{i\underline{\nu}' \cdot (J^T)^{-1}\underline{\varphi}} \\ &= \sum_{\underline{\nu}' \in \mathcal{M}'} f'_{\underline{\nu}'} e^{i\underline{\nu}' \cdot \underline{\varphi}'} = (2\pi)^{r-l} \int f'(\underline{\varphi}')\, d\varphi'_{r+1} \cdots d\varphi'_l \\ &= (2\pi)^{r-l} \int f(J^T \underline{\varphi}')\, d\varphi'_{r+1} \cdots d\varphi'_l \end{aligned} \tag{9.16}$$

as claimed.

Lemma 7 ("Geometric Lemma"). Let $\underline{v}_1,..., \underline{v}_r \in \mathbf{R}^l$ be linearly independent, and consider any vector $\underline{w} \in \mathbf{R}^l$ belonging to the subspace generated by $\underline{v}_1,..., \underline{v}_r$, satisfying

$$|\underline{w} \cdot \underline{v}_j| \leqslant \lambda_j, \qquad j = 1,..., r \tag{9.17}$$

Then we have

$$\|w\|_e \leqslant \frac{\prod_{j=1}^r \|\underline{v}_j\|_e}{\mathrm{Vol}(\underline{v}_1,..., \underline{v}_r)} \sum_{j=1}^r \frac{\lambda_j}{\|\underline{v}_j\|_e} \tag{9.18}$$

where $\|\cdot\|_e$ denotes the euclidean norm, and $\mathrm{Vol}(\underline{v}_1,..., \underline{v}_r)$ is the euclidean volume of the parallelepiped generated by $\underline{v}_1,..., \underline{v}_r$. In particular, if $\underline{v}_1,..., \underline{v}_r$ are integer vectors satisfying $|\underline{v}_j| \leqslant N$, $j = 1,..., r$, and for $\lambda_1 = \cdots = \lambda_r = \lambda$, one has

$$\|\underline{w}\|_e \leqslant r N^{r-1} \lambda \tag{9.19}$$

Statement (9.18), which is trivial for $r = 1$, is easily proven by induction for $r > 1$. The idea is to introduce the decomposition $\underline{w} = \underline{w}' + \underline{w}''$, $\underline{v}_{r+1} = \underline{v}' + \underline{v}''$, with $\underline{w}', \underline{v}'$ belonging to the linear subspace generated by $\underline{v}_1,..., \underline{v}_r$, and $\underline{w}'', \underline{v}''$ perpendicular to it. For a detailed proof of essentially the same statement, see Ref. 2.

10. CONCLUDING REMARKS, AND COMPARISON WITH KAM THEOREM

It is perhaps worth pointing out that one needs very little in the proof of Proposition 2, if one is willing to give up the determination of most constants. All one needs to know is that something like the analytic lemma of Section 7 is true, in the sense that, given a resonance $\mathcal{M}$ and a layer $\mathcal{U}_{\mathcal{M}}$ of points $\underline{A}$ within ε^{a_1} of the resonant surface $\Sigma_{\mathcal{M}}$, $a_1 < \frac{1}{2}$, then one can build a canonical transformation $\mathscr{C}_\varepsilon$ in $\mathscr{U}_{\mathcal{M}} \times \mathbf{T}^l$, close to the identity within ε^{a_2}, $a_2 > a_1$, such that in the new coordinates the Hamiltonian takes the form (7.7). Then, by the arguments used in connection with Propositions 3 and 4, if a_1, a_2 are fixed *a priori*, then data closer to the resonant surface $\Sigma_{\mathcal{M}}$ than ε^{a_3}, with $a_1 < a_3$ (i.e., data well inside the resonance), will never get out of $\mathscr{U}_{\mathcal{M}}$ before an exponentially long time. Indeed, if $a_1 < a_3 < a_2$ can be chosen *a priori* uniformly in $\mathcal{M}$, and the exponential estimate of the remainder in the analytic lemma is also uniform in $\mathcal{M}$ (with the only restriction that ε is small enough), then the arguments leading to Propositions 3 and 4 work, and one obtains the qualitative picture of the motion one is looking for.

However, it is interesting and necessary for applications to have a true control on the dependence of the various quantities involved on the interaction parameters E, ρ, ξ, l, m, M. We have here presented a derivation of these results, in which some care has been devoted to this question, although most numerical constants can be certainly improved. The reader should realize that most of the inequalities are quite natural, or even obvious, if one gives up the determination of the constants in terms of E, ρ, ξ, l, m, M.

Let us compare more closely Nekhoroshev's theorem with the well-known KAM theorem.[11-13]

Both theorems deal with Hamiltonians of the form (2.1), i.e.,

$$H_\varepsilon(\underline{A}, \underline{\varphi}) = h(\underline{A}) + \varepsilon f(\underline{A}, \underline{\varphi}) \tag{10.1}$$

which are assumed to be regular in a suitable domain $V \times \mathbf{T}^l$. Both theorems are based on the construction of canonical transformations of the form $(\underline{A}, \underline{\varphi}) = \mathscr{C}_\varepsilon^{(n)}(\underline{A}', \underline{\varphi}')$, defined in a convenient domain $V_\varepsilon^{(n)} \times \mathbf{T}^l$, which give the new Hamiltonian the form

$$H_\varepsilon^{(n)}(\underline{A}', \underline{\varphi}') = h_\varepsilon^{(n)}(\underline{A}', \underline{\varphi}') + \varepsilon^{n+1} f^{(n+1)}(\underline{A}', \underline{\varphi}', \varepsilon) \tag{10.2}$$

where $h_\varepsilon^{(n)}$ must be either $\underline{\varphi}'$-independent (KAM), or possibly dependent on $\underline{\varphi}'$ in a restricted way (Nekhoroshev), while $f^{(n+1)}$, or better its derivatives, are required to be conveniently bounded in $V_\varepsilon^{(n)} \times \mathbf{T}^l$ for small ε. Finally, both the KAM theorem and Nekhoroshev's theorem are concerned with the problem of taking the limit $n \to \infty$: however, this limit is conceived in substantially different ways.

On the one hand, the KAM theorem looks for truly asymptotic results for $t \to \infty$, at small but fixed ε; this requires taking the limit $n \to \infty$ at fixed $\varepsilon \neq 0$. As is well known, the existence of the limit can be proven, but a serious sacrifice is necessary for what concerns the domain $V_\varepsilon^\infty \times \mathbf{T}^l$ where the final canonical transformation is defined (more precisely, according to Refs 14 and 15, the domain where the old and final Hamiltonians are conjugated by the canonical transformation one is constructing). Indeed the set V_ε^∞ that one is able to construct, although large in measure for small ε, has empty interior. This difficulty, as is well known, arises from the necessity of taking care, for any fixed n, of all possible resonances of $\underline{\omega}(\underline{A})$ with all integer vectors $\underline{v}$, up to an ultraviolet cut-off $N \to \infty$ for $n \to \infty$, so that, in this limit, resonances become dense.

On the other hand, the purpose of Nekhoroshev's theorem is to work in the whole phase space. As we have seen, such a result can be achieved, if one accepts taking the limit $n \to \infty$ together with $\varepsilon \to 0$. In fact, one can say that the essence of Nekhoroshev's theorem is to show that n can be consistently chosen to grow as a power of ε, say $n \sim \varepsilon^{-\beta}$; indeed, replacing this expression in (10.2) directly gives an exponential estimate for the remainder.

Concerning the geometric construction entering Nekhoroshev's theorem (our Proposition 2), this clearly corresponds to the procedure of the KAM theorem of eliminating resonances from the action space. Moreover, as we have seen in Section 9.2, a basic element of the geometric part of Nekhoroshev's theorem is the fact that resonances are, so to speak, well separated from each other. Clearly, this is the counterpart of the basic fact of the KAM theorem that the resonant set to be eliminated from the action space has small measure.

To this purpose, let us consider again the condition defining, in Proposition 2, the nonresonant region ($r=0$), i.e., from (3.3) and (3.6):

$$|\underline{\omega}(\underline{A}) \cdot \underline{v}| > \lambda_1^{(0)} \varepsilon^{\sigma_1} \qquad \forall \underline{v} \neq \underline{0},\ |\underline{v}| \leqslant N \tag{10.3}$$

This condition could be generalized to

$$|\underline{\omega}(\underline{A}) \cdot \underline{v}| > \lambda_1^{(0)} \varepsilon^{\sigma_1} |\underline{v}|^{-\alpha} \qquad \forall \underline{v} \neq 0,\ |\underline{v}| \leqslant N \tag{10.4}$$

which also contains the diophantine condition entering the KAM theorem in the limit $N \to \infty$ at fixed ε. Now, it could be easily seen that the set of angular velocities which do not satisfy the above condition has a measure bounded by an expression of the form

$$\varepsilon^{\sigma_1}(c_1 + c_2 N^{l-1-\alpha}) \tag{10.5}$$

c_1, c_2 being suitable constants. For $N \to \infty$ at fixed ε (KAM) one finds a measure small with ε, as long as $\alpha > l-1$. Instead, if one takes, as we did, $N = \varepsilon^{-\tau}$, the condition becomes

$$\sigma_1 - \tau(l - 1 - \alpha) > 0 \tag{10.6}$$

This condition is easily satisfied, even at $\alpha = 0$, if $\tau < \sigma_1/(l-1)$; this explains why in Proposition 2 the KAM-like diophantine condition was not necessary.

APPENDIX. HOMOCLINIC POINTS FOR A SYSTEM WITH TWO DEGREES OF FREEDOM; AN EXAMPLE

A.1. Here we provide an example of Hamiltonian system with two degrees of freedom, of the form (5.2) with $\varepsilon = 0$, which can be shown to have a homoclinic point.

Denote $(\hat{\underline{S}}, \hat{\underline{\sigma}}) \equiv (a, b, \varphi, \psi)$, and assume for simplicity (forgetting here dimensional correctness) $I_1 = I_2 = 1$. The Hamiltonian is then written

$$H(a, b, \varphi, \psi) = \frac{a^2}{2} + \frac{b^2}{2} + V(\varphi, \psi) \tag{A.1}$$

Let us take

$$V(\varphi, \psi) = -\cos\varphi + \mu\cos(\varphi - \psi) \tag{A.2}$$

which gives, for small μ, a pendulum weakly couled to a free rotator. At $\mu = 0$, the pendulum admits a separatrix, while the rotator regularly turns, with a period T depending on the initial datum. The Poincaré map $(a, \varphi) \mapsto \Psi_\mu(a, \varphi)$, corresponding to the section of the flow $\psi = 0$, $\dot{\psi} > 0$ at any fixed energy E, coincides, for $\mu = 0$, with the time-T-map of the pendulum, so that $\underline{F}_0(a, \varphi)$ admits (for $E > 1$) two hyperbolic fixed points $z_0^\pm \equiv (0, \pm\pi)$, connected by a separatrix ξ_0. For small μ, $\underline{F}_\mu$ will also have two hyperbolic fixed points $\underline{z}_\mu^-$ and $\underline{z}_\mu^+$, with an unstable manifold ξ_μ^- and respectively a stable manifold ξ_μ^+, which replace ξ_0. By means of Melnikov's method,[16] it is not difficult to show that, for small μ, ξ_μ^- and ξ_μ^+ intersect transversally.

For this purpose, let us consider a particular solution $(A(t), B(t), \Phi(t), \Psi(t))$ of Hamiltonian (A.1) for $\mu = 0$, which corresponds to a movement of the pendulum on the separatrix, at fixed total energy E. Such solution is easily checked to be

$$\begin{aligned} \Phi(t) &= \pi - 4\arctan e^{-t} \\ A(t) &= \dot{\Phi}(t) = \frac{4}{e^t + e^{-t}} \\ \Psi(t) &= \omega t \\ B(t) &= \omega \end{aligned} \tag{A.3}$$

with $\omega = \sqrt{2(E-1)}$. For any t, consider the line orthogonal to ξ_0 at $(A(t), \Phi(t))$, oriented, for example, toward the exterior, and introduce an euclidean coordinate s on it, with origin on the separatrix. Let $s^{\pm}(t)$ be the s-coordinates of the intersection of this line with $\xi_\mu^{\pm}$, for small μ. Our main result is contained in the following:

Proposition A1. The coordinates $s^{\pm}(t)$ are given at fixed t by

$$\begin{aligned} s^+(t) &= +\mu[A(t)^2 + (\sin \Phi(t))^2]^{-1/2} \int_0^\infty A(t+\tau) \\ &\quad \times \sin(\Phi(t+\tau) - \Psi(\tau))\, d\tau + \mathcal{O}(\mu^2) \\ s^-(t) &= -\mu[A(t)^2 + (\sin \Phi(t))^2]^{-1/2} \int_{-\infty}^0 A(t+\tau) \\ &\quad \times \sin(\Phi(t+\tau) - \Psi(\tau))\, d\tau + \mathcal{O}(\mu^2) \end{aligned} \tag{A.4}$$

The immediate consequence of this proposition is that ξ_μ^+ and ξ_μ^- intersect transversally, whenever the function

$$\alpha(t) \equiv \int_{-\infty}^{\infty} \frac{\sin(\pi - 4 \arctan e^{-\tau} - \omega(\tau - t))}{e^\tau + e^{-\tau}}\, d\tau \tag{A.5}$$

vanishes with nonvanishing derivative. As $\pi - 4 \arctan e^{-\tau} - \omega\tau$ is an odd function of τ, the above integral vanishes at $t_n = n\pi/\omega$, n integer; for these values of t the derivative of α turns out to be

$$\alpha'(t) = (-1)^n\, \omega \int_{-\infty}^{\infty} \frac{\cos(\pi - 4 \arctan e^{-\tau} - \omega\tau)}{e^\tau + e^{-\tau}}\, d\tau \tag{A.6}$$

This integral is holomorphic in ω, and thus for generic ω (for generic E) different from zero, as claimed.

Proof. To prove the proposition, let us introduce the compact notations $\underline{x} = (a, \varphi)$, $\underline{y} = (b, \psi)$, and write the Hamilton equations of motion corresponding to Hamiltonian (A.1), with V given by (A.2), in the form

$$\begin{aligned} \dot{\underline{x}} &= \underline{f}(\underline{x}) + \mu \underline{h}(\underline{x}, \underline{y}) \\ \dot{\underline{y}} &= \underline{g}(\underline{y}) + \mu \underline{k}(\underline{x}, \underline{y}) \end{aligned} \tag{A.7}$$

Four our Hamiltonian system it is $\underline{f}(\underline{x}) = (-\sin \varphi, a)$, $\underline{g}(\underline{y}) = (0, b)$, $\underline{h}(\underline{x}, \underline{y}) = -\underline{k}(\underline{x}, \underline{y}) = (\sin(\varphi - \psi), 0)$. However, it turns out to be somehow simpler to work within a more general context, assuming only that, for

$\mu=0$, the subsystem $\underline{x}$ has a separatrix ξ_0, connecting two hyperbolic fixed points $\underline{z}_0^{\pm}$, with a particular motion $\underline{X}(t)\to\underline{z}_0^{\pm}$ for $t\to\pm\infty$, while at the same time the subsystem $\underline{y}$ admits a periodic orbit $\underline{Y}(t)$. Let E be the total energy of the system in these conditions, and consider as before the Poincaré map $\underline{x}\mapsto\underline{F}_\mu(\underline{x})$, obtained by keeping fixed E and imposing any condition on $\underline{y}$, which assures transversality with the above periodic orbit at $\mu=0$; $\underline{Y}(0)$ can always be taken on the section.

Proposition A1 is then a particular case of the following Proposition A2, where $s^{\pm}$ are defined as above, while J is the matrix $\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$:

Proposition A2. Within the above assumptions, one has

$$s^{\pm}(t)=\mu\frac{1}{\|\underline{f}(\underline{X}(t))\|}\int_0^{\pm\infty} J\underline{f}(\underline{X}(t+\tau))\cdot\underline{h}(\underline{X}(t+\tau),\underline{Y}(\tau))\,d\tau+\mathcal{O}(\mu^2) \tag{A.8}$$

where the dot denotes the euclidean scalar product and $\|\cdot\|$ the euclidean norm.

Proof. Consider an initial datum $\underline{x}^{\pm}(\mu)$ for $\underline{x}$,

$$\underline{x}^{\pm}(\mu)=\underline{x}_0+\mu\underline{x}_1^{\pm}+\cdots \tag{A.9}$$

with $\underline{x}_0\in\xi_0$, i.e., $\underline{x}_0=\underline{X}(t_0)$, and $\underline{x}_1^{\pm}$ perpendicular to ξ_0 at $\underline{x}_0$, at distance $s_1^{\pm}$, i.e.,

$$\underline{x}_1^{\pm}=s_1^{\pm}\frac{J\underline{f}(\underline{x}_0)}{\|\underline{f}(\underline{x}_0)\|} \tag{A.10}$$

The corresponding initial datum $\underline{y}^{\pm}(\mu)=\underline{y}_0+\mathcal{O}(\mu)$ for $\underline{y}$ is imposed by the choice of the Poincaré section; for $\mu=0$, it will be $\underline{y}^{\pm}(0)=\underline{y}_0=\underline{Y}(0)$.

We must now impose that $\underline{x}^{\pm}(\mu)$ belongs to $\xi_\mu^{\pm}$; as we will see, this condition, to first order in μ, directly leads to (A.8). For this purpose, denote by $(\underline{\Theta}^{\pm}(t,\mu),\underline{\Xi}^{\pm}(t,\mu))$ the solution to (A.7) corresponding to the initial datum $(\underline{x}^{\pm}(\mu),\underline{y}^{\pm}(\mu))$, and write

$$\begin{aligned}\underline{\Theta}^{\pm}(t,\mu)&=\underline{\Theta}_0(t)+\mu\underline{\Theta}_1^{\pm}(t)+\cdots\\ \underline{\Xi}^{\pm}(t,\mu)&=\underline{\Xi}_0(t)+\cdots\end{aligned} \tag{A.11}$$

Obviously, $\underline{\Theta}_0$ must be a motion on the separatrix, corresponding to initial datum $\underline{x}_0$, i.e., $\underline{\Theta}_0(t)=\underline{X}(t+t_0)$, while (for the choice of the section) one has $\underline{\Xi}_0(t)=\underline{Y}(t)$. Concerning $\underline{\Theta}_1^{\pm}(t)$, it must be a solution of the linearized equation

$$\dot{\underline{\vartheta}}=L(\underline{\Theta}_0(t))\,\underline{\vartheta}+\underline{h}(\underline{\Theta}_0(t),\underline{\Xi}_0(t)) \tag{A.12}$$

with initial datum $\underline{x}_1^{\pm}$, where L is the matrix $(\partial\underline{f}/\partial\underline{x})$.

Denote by $\underset{\sim}{\vartheta} \mapsto \Gamma(t, \underset{\sim}{x}_0)\, \underset{\sim}{\vartheta}$ the solution of the homogeneous equation associated to (A.12) (which is just the linearized equation of the unperturbed $\underset{\sim}{x}$ problem, on the separatrix); the solution to (A.12) with the above initial datum can then be written in the form

$$\underset{\sim}{\Theta}_1^{\pm}(t) = \Gamma(t, \underset{\sim}{x}_0) \left[\underset{\sim}{x}_1^{\pm} + \int_0^t \Gamma^{-1}(\tau, \underset{\sim}{x}_0)\, h(\underset{\sim}{\Theta}_0(\tau), \underset{\sim}{\Xi}_0(\tau))\, d\tau \right] \quad \text{(A.13)}$$

In order that this quantity remain finite at $t \to \pm\infty$, it is necessary that the vector in the square bracket in (A.13), for $t \to \pm\infty$, becomes tangent to the separatrix at $\underset{\sim}{x}_0$ (moreover, with a well-defined speed), as follows from the hyperbolicity of $\underset{\sim}{z}_0^{\pm}$. This gives, after a projection in the direction orthogonal to ξ_0 at $\underset{\sim}{x}_0$,

$$s_1^{\pm} = -\frac{1}{\|\underset{\sim}{f}(\underset{\sim}{x}_0)\|} \int_0^{\pm\infty} J\underset{\sim}{f}(\underset{\sim}{X}(t_0)) \cdot \Gamma^{-1}(\tau, \underset{\sim}{X}(t_0))\, h(\underset{\sim}{X}(t_0+\tau), \underset{\sim}{Y}(\tau))\, d\tau \quad \text{(A.14)}$$

Recalling now that Γ is symplectic, so that $(\Gamma^{-1})^T = -J\Gamma J$, and that one has

$$\Gamma(\tau, \underset{\sim}{X}(t_0))\, \underset{\sim}{f}(\underset{\sim}{X}(t_0)) = \underset{\sim}{f}(\underset{\sim}{X}(t_0+\tau)) \quad \text{(A.15)}$$

(A.14) immediately leads to (A.8), as claimed.

A.2. One can use the above example to illustrate in a more quantitative way the statement that in the intermediate time scale chaotic motions become, in principle, observable, as hinted in Section 5.

Consider the system

$$\tfrac{1}{2}(A_1^2 + A_2^2 + A_3^2) + \varepsilon(-\cos\varphi_1 + \mu\cos(\varphi_1 - \varphi_2)) + \varepsilon^2 \mathcal{f}(\underset{\sim}{A}, \underset{\sim}{\varphi}) \quad \text{(A.16)}$$

near the resonance $\underset{\sim}{\nu}_1 = (1, 0, 0)$, $\underset{\sim}{\nu}_2 = (0, 1, 0)$, say, and with initial data close to $\underset{\sim}{A} = (0, 0, 1)$. After making the change of variables described in Section 4, Propositions 3 and 4, and (4.3), (4.7), (4.9), the motion will be described in the slow $(\underset{\sim}{S}, \underset{\sim}{\sigma})$-variables $[\underset{\sim}{S} = (S_1, S_2),\ \underset{\sim}{\sigma} = (\sigma_1, \sigma_2)]$ and fast (F, ψ)-variables by a Hamiltonian

$$\frac{1}{\sqrt{\varepsilon}} h(F, \varepsilon) + \{\tfrac{1}{2}(S_1^2 + S_2^2) - \cos\sigma_1 + \mu\cos(\sigma_1 - \sigma_2)\} + \sqrt{\varepsilon}\, V^{(1)}(\underset{\sim}{S}, F, \underset{\sim}{\sigma}, \varepsilon) + e^{-(1/8)\xi\varepsilon^{-b}} V^{\infty}(\underset{\sim}{S}, F, \underset{\sim}{\sigma}, \psi, \varepsilon) \quad \text{(A.17)}$$

where $h(F, \varepsilon) = F + \frac{1}{2}\sqrt{\varepsilon}\, F^2$. The description of the motion by (A.17) also incorporates an inessential time scale change, so that one unit of time corresponds to $1/\sqrt{\varepsilon}$ units of the "physical" (i.e., old) time; the description holds for initial data of the form (S_1, S_2, F) with $|F|, |S_i| \leqslant 1$, say, up to a time $\mathcal{O}(e^{(1/8)\xi\varepsilon^{-b}})$.

We suppose μ small so that the system with two degrees of freedom in curly brackets has a homoclinic point as discussed above, and we shall use the fact that near such homoclinic point one can prove by rather standard techniques[17] that there is a "large" set $\mathscr{R}$ of initial conditions which produce chaotic motions, in the sense that if one observes them every time they cross, say, the section $\sigma_2 = 0$, $\dot{\sigma}_2 > 0$, [i.e., via the iterates of the above-introduced Poincaré map $F_\mu(S_1, \sigma_1)$], then in a suitable set of coordinates the points of $\mathscr{R}$ can be represented by sequences of symbols, on which the dynamics acts as a simple shift, which can be randomly prescribed. In other words, $\mathscr{R}$ is homeomorphic to a space $\{A, B, C \cdots\}^2$ of sequences of symbols, and in such "coordinates" the action of the Poincaré map is just a shift.

If we now look at the evolution including the $V^{(1)}$ term, i.e., if we consider the Hamiltonian

$$\frac{1}{\sqrt{\varepsilon}} h(F, \varepsilon) + \{\tfrac{1}{2}(S_1^2 + S_2^2) - \cos \sigma_1 + \mu \cos(\sigma_1 - \sigma_2)\} + \sqrt{\varepsilon}\, V^{(1)}(\underline{S}, F, \underline{\sigma}, \varepsilon) \qquad \text{(A.18)}$$

then the basic property of the existence of chaotic motions, for small ε, is not changed, because F is a constant of motion, so that the system in fact remains two-dimensional, and because the homoclinic points will persist under the perturbation, being structurally stable.

Let us next look at the true evolution including V^∞ too. We want to estimate for how long the motions described by (A.17) and (A.18) stay close to each other, within an error, say, $\mathcal{O}(e^{-(1/16)\xi\varepsilon^{-b}})$: up to this time it is clear that the motions of (A.17) can be "confused" with those of (A.18) which, as discussed above, are possibly chaotic.

An abstract picture of our problem could be the following. Write $\underline{x} = (\underline{S}, \underline{\sigma})$ and fix once and for all the initial datum $(\underline{S}_0, F_0, \underline{\sigma}_0, \psi_0)$. Then we want to compare the solutions of the two equations

$$\dot{\underline{x}} = g(\underline{x}, F) + e^{-(1/8)\xi\varepsilon^{-b}} g_\infty(\underline{x}, F, \psi) \qquad \underline{x}(0) = \underline{x}_0 \qquad \text{(A.19a)}$$

$$\dot{\underline{y}} = g(\underline{y}, F_0) \qquad \underline{x}(0) = \underline{x}_0 \qquad \text{(A.19b)}$$

where F, ψ are imagined as known functions of time (to be computed from the complete set of Hamilton equations corresponding to (A.17) and from initial data).

The theory of previous sections provides us immediately with a bound γ on $|\partial g_i/\partial x_j|$, $|\partial g_i/\partial F|$ uniform in ε for $\varepsilon \to 0$: hence, naively, we should expect a divergence of the solutions of (A.19a) from those of (A.19b), at the rate $e^{\gamma t}$, which would not allow us to reach any time scale larger than $\mathcal{O}(\log \varepsilon)$. However, (A.19a) has the special feature that the ψ-dependence does not appear in $\underline{g}$, but "only" in the remainder. Therefore we can repeat the argument leading to the $e^{\gamma t}$ growth, making use of this fact and of the estimate

$$|F - F_0| \leqslant |t|\, e^{-(1/8)\xi\varepsilon^{-b}} \mathcal{O}(1) \tag{A.20}$$

followed by integration from

$$\dot{F} = -e^{-(1/8)\xi\varepsilon^{-b}} \frac{\partial V^{\infty}}{\partial \psi}, \qquad \left|\frac{\partial V^{\infty}}{\partial \psi}\right| \leqslant \mathcal{O}(1) \tag{A.21}$$

Denoting $\underline{\delta} = (\underline{x} - \underline{y})$, we can write

$$\begin{aligned}\dot{\underline{\delta}} = \underline{g}(\underline{x}, F_0) - \underline{g}(\underline{y}, F_0) + (\underline{g}(\underline{x}, F) - \underline{g}(\underline{x}, F_0)) - (\underline{g}(\underline{y}, F) - \underline{g}(\underline{y}, F_0))\\ + e^{-(1/8)\xi\varepsilon^{-b}} \underline{g}_{\infty}(\underline{x}, F, \psi)\end{aligned} \tag{A.22}$$

and by integration, using $\underline{\delta}(0) = \underline{0}$ and setting $d = |\underline{\delta}|$, we find

$$\begin{aligned}\delta(t) &\leqslant \int_0^t \gamma\delta(\tau)\, d\tau + \int_0^t [e^{-(1/8)\xi\varepsilon^{-b}} |\underline{g}_{\infty}(\underline{x}(\tau), F(\tau), \psi(\tau))| + \gamma |F(\tau) - F_0|]\, d\tau\\ &\leqslant \int_0^t \gamma\delta(\tau)\, d\tau + \int_0^t e^{-(1/8)\xi\varepsilon^{-b}} \mathcal{O}(1)\, d\tau\end{aligned} \tag{A.23}$$

i.e.,

$$\delta(t) \leqslant \frac{1}{\gamma} e^{\gamma t - (1/8)\xi\varepsilon^{-b}} \mathcal{O}(1) \tag{A.24}$$

This shows that the error remains $\mathcal{O}(e^{-(1/16)\xi\varepsilon^{-b}})$ up to a time t_0 such that $\gamma t_0 < \frac{1}{16}\xi\varepsilon^{-b}$, i.e., a time much longer than $\log \varepsilon^{-1}$. So we can see chaotic motions, and even compute their initial conditions, with great accuracy up to times of order ε^{-b} times larger than their natural time scale, in spite of the fact that in the above bounds we proceed as if the reference system had sensitive dependence on initial conditions bounded by a positive ε-independent Lyapunov exponent γ!

After a time $\mathcal{O}(\varepsilon^{-b})$ has elapsed (in the original physical units this is a time of order $\varepsilon^{-b-1/2}$) the motion is still described by (A.17), as discussed in this paper, but we can no longer estimate the effects of neglecting the last term: since the evolution of (F, ψ) is not independent of that of $(\underline{S}, \underline{\sigma})$, and since the set $\mathcal{R}$ of chaotic motions is large but does not exhaust the whole phase space (e.g., one constructs $\mathscr{R}$ as a set of zero measure, and it is not even known whether it could be enlarged to a set of positive measure), it is conceivable that the motion of $(\underline{S}, \underline{\sigma})$ will be systematically outside $\mathscr{R}$, thus destroying chaos on scales larger than ε^{-b}. Although the latter event seems unlikely under "general situations," unfortunately the present techniques do not allow us to prove a weak generality statement.

ACKNOWLEDGMENT

We are grateful to A. Celletti and C. Liverani for stimulating discussions and remarks. We are also very grateful to C. E. Wayne for a series of detailed comments and remarks.

Most of this work was done while the authors were at Rutgers University, Hill Center. One of us (G. G.) has been partly supported by NSF grant No. DMS-85-033-33.

REFERENCES

1. N. N. Nekhoroshev, *Usp. Mat. Nauk* **32**:5 (1977) [*Russ. Math. Surv.* **32**:1 (1977)]; *Tr. Sem. Petrows.* No. 5, 5 (1979).
2. G. Benettin, L. Galgani, and A. Giorgilli, *Celest. Mech.* **37**:1 (1985).
3. G. Gallavotti, lectures given at the 1984 *Les Houches* Summer School, to be published.
4. C. E. Wayne, *Commun. Math. Phys.* **103**:351 (1986); **104**:21 (1986).
5. G. Benettin, L. Galgani, and A. Giorgilli, *Nuovo Cimento B.* **89**:89 (1985).
6. G. Benettin, L. Galgani, and A. Giorgilli, *Nuovo Cimento B.* **89**:103 (1985).
7. V. I. Arnold, *Dokl. Akad. Nauk SSSR* **156**:9 (1964) [*Sov. Math. Dokl.* **6**:581 (1964)].
8. V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
9. H. Poincaré, *Les Méthodes Nouvelles de la Méchanique Céleste*, Vol. 3, Chapter 33 (Gauthier-Villars, Paris, 1899).
10. G. Gallavotti, *The Elements of Mechanics* (Springer Verlag, Berlin 1983).
11. A. N. Kolmogorov, *Dokl. Akad. Nauk. SSSR* **98**:527 (1954); English translation in *Lecture Notes in Physics*, G. Casati and G. Ford, eds., No. 93 (Springer Verlag, Berlin, 1979).
12. V. I. Arnold, *Usp. Mat. Nauk* **18**:13 (1963) [*Russ. Math. Surv.* **18**:9 (1963)]; *Usp. Mat. Nauk* **18**:91 (1963) [*Russ. Math. Surv.* **18**:85 (1963)].
13. J. Moser, *Nachr. Akad. Wiss. Gottingen Math. Phys. Kl.* **2**:1 (1962).
14. J. Poeschel, *Commun. Pure Appl. Math.* **35**:653 (1982); *Celest. Mech.* **28**:133 (1982).
15. L. Chierchia and G. Gallavotti, *Nuovo Cimento B* **67**:277 (1982).
16. V. K. Melnikov, *Trans. Moscow Math. Soc.* **12**:1 (1963).
17. J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton University Press, Princeton, 1973).