Commun. Math. Phys. 113, 87–103 (1987)

Communications in Mathematical Physics

Realization of Holonomic Constraints and Freezing of High Frequency Degrees of Freedom in the Light of Classical Perturbation Theory. Part I

Giancarlo Benettin[1], Luigi Galgani[2] and Antonio Giorgilli[3]

1 Dipartimento di Fisica dell'Università di Padova Via Marzolo 8, I-35131 Padova, Italia
2 Dipartimento di Matematica dell'Università di Milano Via Saldini 50, I-20133 Milano, Italia
3 Dipartimento di Fisica dell'Università di Milano Via Celoria 16, I-20133 Milano, Italia

Abstract. The so-called problem of the realization of the holonomic constraints of classical mechanics is here revisited, in the light of Nekhoroshev-like classical perturbation theory. Precisely, if constraints are physically represented by very steep potential wells, with associated high frequency transversal vibrations, then one shows that (within suitable assumptions) the vibrational energy and the energy associated to the constrained motion are separately almost constant, for a very long time scale growing exponentially with the frequency (i.e., with the rigidity of the constraint one aims to realize). This result can also be applied to microscopic physics, providing a possible entirely classical mechanism for the "freezing" of the high-frequency degrees of freedom, in terms of non-equilibrium statistical mechanics, according to some ideas expressed by Boltzmann and Jeans at the turn of the century. In this Part I we introduce the problem and prove a first theorem concerning the realization of a single constraint (within a system of any number of degrees of freedom). The problem of the realization of many constraints will be considered in a forthcoming Part II.

1. Introduction

1.1 In this paper, and in a forthcoming second part, we will be concerned with Hamiltonian dynamical systems of the form

$$H_\omega(p, x, \pi, \xi) = h_\omega(\pi, \xi) + \hat{h}(p, x) + f(p, x, \pi, \xi), \tag{1.1}$$

where $h_\omega(\pi, \xi)$, with $(\pi, \xi) = (\pi_1, \dots, \pi_\nu, \xi_1, \dots, \xi_\nu) \in \mathbf{R}^{2\nu}$, is the Hamiltonian of a set of ν uncoupled harmonic oscillators of angular frequency $\omega = (\omega_1, \dots, \omega_\nu)$, i.e.

$$h_\omega(\pi, \xi) = \frac{1}{2} \sum_{i=1}^{\nu} (\pi_i^2 + \omega_i^2 \xi_i^2), \tag{1.2}$$

while $\hat{h}(p,x)$, $(p,x)=(p_1,\ldots,p_n,x_1,\ldots,x_n)\in G\subset\mathbf{R}^{2n}$, represents any dynamical system with n degrees of freedom, defined in a domain $G\subset\mathbf{R}^{2n}$; $f(p,x,\pi,\xi)$ is a coupling term, which is assumed to vanish for $\xi=0$.

The reason why we are interested in this class of Hamiltonians is that they naturally appear in the so-called problem of the "realization of constraints" [1 – 5] of classical mechanics, and in particular, as we explain in subsection (1.3), in one of the fundamental problems of classical statistical mechanics discussed particularly by Boltzmann [6] and Jeans [7, 8] at the turn of the century, which concerns the dynamical foundations of the principle of equipartition of energy.

As the problem of realization of constraints is not much studied in the literature, it will be recalled in some detail in the Appendix; just to give here the basic idea, let us consider the simple example of a spherical pendulum of mass m and length R. Using spherical coordinates (r,θ,φ), and denoting $x=(\theta,\varphi)$ and $p=(p_\theta,p_\varphi)$, the Hamiltonian of the constrained system is

$$\hat{h}(p,x)=\frac{p_\theta^2}{2mR^2}+\frac{p_\varphi^2}{2m(R\sin\theta)^2}+mg\,R\cos\theta. \tag{1.3}$$

The question is then how the constraint is physically realized. The idea is that the physical device which actually produces the constraint (for example, a bar of negligible mass and length R) can be represented by a steep potential well of the form, say, $\frac{1}{2}k(r-R)^2$, with very large k. This leads to consider in place of (1.3) the complete Hamiltonian

$$\begin{aligned} H_\omega(p,x,\pi,\xi)=&\frac{\pi^2}{2m}+m\omega^2\xi^2+\frac{p_\theta^2}{2m(R+\xi)^2}\\ &+\frac{p_\varphi^2}{2m[(R+\xi)\sin\theta]^2}+mg(R+\xi)\cos\theta, \end{aligned}\tag{1.4}$$

with $\omega^2=k/m$, $\xi=r-R$, and π the corresponding momentum. One wants to study this physical model for large ω, and understand whether, and in which sense, its behavior is close to that of the corresponding ideal system (1.3). The trivial rescaling $\xi=m^{-1/2}\xi'$, $\pi=m^{1/2}\pi'$ and a Taylor expansion in ξ at $\xi=0$ for the third and the fourth terms give the Hamiltonian (1.4) the form (1.1), the accents in the new variables having been omitted; as explained in the Appendix, one is quite naturally led to consider in general Hamiltonians of the form (1.1) whenever one has a dynamical system with $n+\nu$ degrees of freedom, which reduces to n degrees of freedom via the "physical" introduction of ν holonomic constraints.

1.2 The purpose of the present paper and of the forthcoming one is thus to study the above class of Hamiltonian systems in the limit of large ω, say $\omega=\lambda\Omega$, with some fixed $\Omega=(\Omega_1,\ldots,\Omega_\nu)$ and large λ. In particular, our aim is to study the time-scale on which sensible energy exchanges among the (p,x) and the (π,ξ) degrees of freedom are possible, and to show that, within suitable assumptions, there is almost no energy exchange (more precisely, there is an energy exchange bounded by an inverse power of λ) for a time-scale growing exponentially with λ, say for

$$|t|\leq t^*e^{(\lambda/\lambda^*)^\alpha}, \tag{1.5}$$

where t^* is a certain time, while λ^* and a are suitable positive constants. To this purpose, we will make use of classical perturbation theory, treating (π, ξ) as the fast variables, which is equivalent to using an inverse power of λ as the small perturbation parameter; indeed, as is well known after Nekhoroshev's work [9] (see also refs. [10–13]), classical perturbation theory quite naturally leads to exponentially long time scales. As in the above quoted references, we will restrict ourselves to the analytic case.

Another question that can be asked in connection with a Hamiltonian of the form (1.1) concerns the relations between orbits of the complete Hamiltonian H_ω and orbits of the constrained Hamiltonian $\hat{h}$. For example, denoting $(p_\lambda(t), x_\lambda(t), \pi_\lambda(t), \xi_\lambda(t))$ a particular solution for the Hamiltonian H_ω, one can wonder whether, for initial data satisfying the constraints, one can obtain $(p_\lambda(t), x_\lambda(t)) \to (\hat{p}(t), \hat{x}(t))$ as $\lambda \to \infty$, $(\hat{p}(t), \hat{x}(t))$ being a particular solution for the Hamiltonian $\hat{h}$ with corresponding initial data. As recalled in the Appendix, a result substantially equivalent to an affirmative answer has in fact been obtained in refs. [1–5]. The point however is that the above limit is highly non-uniform in time: in fact, from the proofs reported in the above quoted references (which essentially rely on the uniqueness theorem for the solutions of ordinary differential equations) one only gets quite poor estimates of the form

$$\operatorname{dist}[((p_\lambda(t), x_\lambda(t)), (\hat{p}(t), \hat{x}(t))] \leqq K\lambda^{-1} e^{\mu t}, \quad K, \mu > 0, \tag{1.6}$$

which, although being in general optimal, loose any usefulness after a time scale of order $\mu^{-1} \log \lambda$.

Results for long time scales as in (1.5) can thus be expected only for integrals of motion of the unperturbed system, like the energy of the constrained system, and not for orbits. In fact, such a situation of rapidly lost orbits with nevertheless preservation of some integrals of motion, like $\hat{h}$, is rather natural in classical perturbation theory; for a discussion on this point, see Sect. 5 of ref. [13], or ref. [14].

In this first part we will consider the simpler case $\nu = 1$ (but any n), i.e., the realization of a single constraint; all of our estimates will be independent of n, and the exponent a appearing in (1.5), which gives the correction to the pure exponential law, will be $1/2$. The same techniques could be extended to $\nu > 1$ (see the discussion in Sect. 5), but only for Diophantine frequencies, and with a heavy ν-dependence of the estimates.

1.3 On the other hand, the case $\nu > 1$ with exactly equal angular frequencies ($\nu - 1$ independent resonances, in the language of perturbation theory) is of particular importance in classical statistical mechanics. Indeed, take, as a model example, a classical diatomic gas of N identical molecules of unitary mass; assume that the free molecules vibrate elastically, with elastic constant ω^2, and interact via purely positional forces. Then, denoting by $x = (x_1, \ldots, x_{5N})$ the set of translational and rotational degrees of freedom, and by $\xi = (\xi_1, \ldots, \xi_N)$ the internal vibrational ones, it is quite evident that the Hamiltonian of the system can be given the form

$$H = \frac{1}{2} \sum_{i=1}^{N} (\pi_i^2 + \omega^2 \xi_i^2) + \hat{H}(p, x, \xi), \tag{1.7}$$

for a certain $\hat{H}$. This Hamiltonian has just the form (1.1), with $n = 5N$ and $\nu = N$, if one denotes $\hat{h}(p,x) = \hat{H}(p,x,0)$ and $f = \hat{H} - \hat{h}$; all of the frequencies are equal and f is independent of π, while $\hat{h}$ is clearly the Hamiltonian of the classical diatomic gas, with ideally rigid molecules. Thus, studying the behavior of the classical gas in the limit of high vibrational frequencies is just studying the problem of the physical realization of the constraint of rigidity of the molecules.

As is well known, such a system presents the severest of the difficulties of classical statistical mechanics: indeed, for no matter how large ω, the equipartition principle would lead to the value $C_V = \frac{7}{2}R$ for the specific heat, while ordinary temperature experiments give $C_V = \frac{5}{2}R$, as if the vibrational degrees of freedom were "frozen." A possible entirely classical way out of such and similar difficulties (including the blackbody questions) was proposed by Boltzmann and Jeans in the above quoted references [6–8], and reconsidered very recently [15] in the framework of Nekhoroshev's theorem. The idea is precisely that the time scale associated to the energy flow between the vibrational degrees of freedom and the translational and rotational ones could be so large ("years", [6] "billions of years", [7] in the words of Boltzmann and Jeans), that on the time-scale of ordinary experiments the high frequency degrees of freedom would behave as if they were really frozen. Moreover, a very relevant fact from our point of view is that in one of Jeans's papers an exponential law of the form $t = t^* e^{\alpha\omega}$, corresponding to (1.5) with $a = 1$, is explicitly proposed in connection with the diatomic gas model sketched above, t^* and α being typical microscopic times.

The second part of this work, which is published separately, treats just the case $\nu > 1$ (with the restriction that the coupling f be independent of π), with a technique that allows to consider the case described above of exactly equal angular frequencies. The results to be reported in Part II are thus intended to support theoretically the ideas of Boltzmann and Jeans. As will be better explained in Part II, a and t^* will be independent of N, while unfortunately λ^* will turn out to be proportional to N. This fact, in our opinion, cannot be avoided at a purely dynamical level, essentially because one cannot exclude those extremely unlike states where all of the energy is concentrated in a few degrees of freedom, or where a number of molecules of the order of N interact simultaneously. The reader interested in this application to statistical mechanics should perhaps consider our previous paper [16], where the ideas of Boltzmann and Jeans are revisited in greater detail, and supported by numerical experiments.

1.4 The present Part I is organized as follows: In Sect. 2 we state our main theorem, providing estimates for a canonical transformation relevant for our Hamiltonian, and deduce from it a corollary which proves the practical conservation of energy for the constrained system for long time scales. The proof of the theorem is given in Sect. 3, while a few technical lemmas are proven in Sect. 4. A conclusion follows. The problem of the realization of constraints is briefly recalled in the Appendix.

We are indebted to G. Gallavotti and J.-M. Strelcyn for useful discussions. Part of this work was done when two of us (G.B. and A.G.) were guests of the Institute of Theoretical Physics of ETH, Zurich. We are very grateful to professor J. Froehlich for his kind invitation.

2. Results for the Case $\nu = 1$

Let us consider the Hamiltonian system (1.1) for $\nu = 1$, with h_ω given by (1.2), namely

$$H_\omega = \tfrac{1}{2}(\pi^2 + \omega^2 \xi^2) + \hat{h}(p,x) + f(p,x,\pi,\xi), \tag{2.1}$$

where $(p,x) \in G \subset \mathbf{R}^{2n}$, G being a bounded domain (which should be thought of as the natural domain of definition of $\hat{h}$, say the domain inside a compact energy surface), while $(\pi,\xi) \in B \subset \mathbf{R}^2$, B being defined by a condition of the form $\frac{1}{2}(\pi^2 + \omega^2\xi^2) \leqq E_0$, with some fixed energy E_0. Both $\hat{h}$ and f are assumed to be analytic in $B \times G$; moreover, f vanishes for vanishing ξ, so that one can write $f = \xi\tilde{f}$, $\tilde{f}$ being analytic and bounded in $B \times G$.

Denote $\omega = \lambda\Omega$, where Ω is some fixed frequency (say the inverse of a typical time-scale associated to the unperturbed problem), and introduce, in place of (π,ξ) the action-angle variables (I,φ), by the usual substitution

$$\begin{aligned} \pi &= \sqrt{2I\omega}\cos\varphi = \sqrt{2\lambda\Omega I}\cos\varphi, \\ \xi &= \sqrt{2I/\omega}\sin\varphi = (\lambda\Omega)^{-1}\sqrt{2\lambda\Omega I}\sin\varphi. \end{aligned} \tag{2.2}$$

Then the Hamiltonian takes the form

$$H_\lambda = \lambda\Omega I + \hat{h}(p,x) + \lambda^{-1} f_\lambda(p,x,I,\varphi), \tag{2.3}$$

with

$$f_\lambda(p,x,I,\varphi) = \Omega^{-1}\sqrt{2\lambda\Omega I}\sin\varphi\tilde{f}(p,x,\sqrt{2\lambda\Omega I}\cos\varphi,(\lambda\Omega)^{-1}\sqrt{2\lambda\Omega I}\sin\varphi), \tag{2.4}$$

and f_λ is bounded, as follows from the assumption $\lambda\Omega I \leqq E_0$, which corresponds to $I \leqq I_0/\lambda$ with $I_0 = E_0/\Omega$, and from the fact that $\tilde{f}$ is bounded. To such a Hamiltonian we will apply classical perturbation theory, using λ^{-1} as the small parameter.

Our Hamiltonian is then defined in the real domain $D = G \times [0, I_0/\lambda] \times \mathbf{T}^1$, where $\mathbf{T}^1$ is the one-dimensional torus, or circle; this is equivalent to considering $\varphi \in \mathbf{R}$, all functions being periodic in φ with period 2π. Unfortunately, as is clear from (2.4), f_λ is not analytic as a function of I at $I = 0$. Thus, in order to use the simple apparatus of classical perturbation theory in the analytic case, we make a preliminary restriction of the domain of definition of H_λ from D to $D_0 = G \times [\rho_0/\lambda, I_0/\lambda] \times \mathbf{T}^1$, with a suitable $\rho_0 < I_0$ (as we shall see later, it will be possible to take ρ_0 proportional to $\lambda^{-\beta}$ with $0 \leqq \beta < \frac{1}{2}$, so that the restriction becomes unessential for large λ). Then, we will assume that $\hat{h}$ and f_λ are analytic and bounded in a complex domain $D_\rho \supset D_0$, defined as follows: ρ is the vector $(\rho_p, \rho_x, \rho_I, \rho_\varphi)$, while D_ρ is the union of the polydisks of radii ρ_p, ρ_I, $\lambda^{-1}\rho_I$, ρ_φ centered in the points of D_0, i.e.

$$D_\rho = \bigcup_{(p,x,I,\varphi)\in D_0} \Delta_\rho(p,x,I,\varphi), \tag{2.5}$$

with

$$\begin{aligned} \Delta_\rho(p,x,I,\varphi) = \{&(p',x',I',\varphi') \in \mathbf{C}^{2n+2}; |p'_i - p_i| \leqq \rho_p, \\ &|x'_i - x_i| \leqq \rho_x,\ i = 1,\dots,n, |I' - I| \leqq \lambda^{-1}\rho_I, |\varphi' - \varphi| \leqq \rho_\varphi\} \end{aligned} \tag{2.6}$$

(pay attention to the restriction on I, which is necessary to keep the energy $\lambda\Omega I$ finite when $\lambda \to \infty$). In particular, in agreement with a previous remark, the angle φ can be considered as defined in the strip $|\mathrm{Im}\,\varphi| \leqq \rho_\varphi$, with no restriction on $\mathrm{Re}\,\varphi$, all functions being periodic in φ with real period 2π. Although it is not strictly necessary, for practical convenience we will introduce in the following theorem the restriction $\rho_\varphi \leqq 1$. Concerning ρ_I, one clearly needs only $\rho_I \leqq \rho_0$, but the optimal choice is obviously $\rho_0 = \rho_I$. In order to make use of Cauchy inequality (see Lemma 1 of Sect. 4) we will consider restricted domains of the form $D_{\rho-\delta}$, with $\delta = (\delta_p, \delta_x, \delta_I, \delta_\varphi)$ and $\delta < \rho$, the inequality working separately on the different components. The supremum norm in any set $D_{\rho'}$, $\rho' \leqq \rho$, will be denoted by $\|\cdot\|_{\rho'}$. A relevant role in the perturbation scheme will be played by $\min(\rho_p\rho_x, \rho_I\rho_\varphi)$, which will be simply denoted by ρ^2, with a similar notation for the restriction δ.

After these preliminaries, we can now state the following

Theorem. *Consider the Hamiltonian*

$$H_\lambda = \lambda\Omega I + \hat{h}(p,x) + \lambda^{-1} f_\lambda(p,x,I,\varphi), \tag{2.7}$$

in the domain D_ρ given by (2.5), with $\rho = (\rho_p, \rho_x, \rho_I, \rho_\varphi)$ and $\rho_\varphi \leqq 1$. Assume that $\hat{h}$ and f_λ are analytic in the interior of D_ρ, and denote

$$E = \|f_\lambda\|_\rho, \quad \hat{E} = \max(\|\hat{h}\|_\rho, E, \Omega\rho^2) \tag{2.8}$$

and

$$\lambda^* = 2^{11}\rho^{-2}\Omega^{-1}\hat{E}. \tag{2.9}$$

Then there exists a real analytic canonical transformation $(p,x,I,\varphi) = \mathscr{C}_\lambda(p',x',I',\varphi')$, $\mathscr{C}_\lambda$: $D_{(1/2)\rho} \to D_\rho$, with $\mathscr{C}_\lambda(D_{(1/2)\rho}) \supset D_{(1/4)\rho}$, which satisfies the estimates

$$|p_i - p_i'| < 2^{-3}\lambda^{-1}(\lambda/\lambda^*)^{-1/2}(E/\hat{E})\rho_p \qquad (i = 1,\dots,n), \tag{2.10a}$$

$$|x_i - x_i'| < 2^{-3}\lambda^{-1}(\lambda/\lambda^*)^{-1/2}(E/\hat{E})\rho_x \qquad (i = 1,\dots,n), \tag{2.10b}$$

$$|I - I'| < 2^{-5}\lambda^{-1}(\lambda/\lambda^*)^{-1}(E/\hat{E})\rho_I, \tag{2.10c}$$

$$|\varphi - \varphi'| < 2^{-3}(\lambda/\lambda^*)^{-1/2}(E/\hat{E})\rho_\varphi, \tag{2.10d}$$

and gives the Hamiltonian $H_\lambda' \equiv H_\lambda \circ \mathscr{C}_\lambda$ the form

$$H_\lambda'(p',x',I',\varphi') = \lambda\Omega I' + \hat{h}(p',x') + \lambda^{-1}g_\lambda(p',x',I') + \lambda^{-1}e^{-(\lambda/\lambda^*)^{1/2}} f_\lambda'(p',x',I',\varphi'), \tag{2.11}$$

where g_λ and f_λ' satisfy the estimates

$$\|g_\lambda\|_{(1/2)\rho} < 2E, \quad \|f_\lambda'\|_{(1/2)\rho} < 8E. \tag{2.12}$$

The proof is deferred to the next section. From this theorem one immediately deduces the following

Corollary. *For any orbit $(p(t), x(t), I(t), \varphi(t))$ of the Hamiltonian system (2.7), with initial datum in D_0, denoting by $(p'(t), x'(t), I'(t), \varphi'(t))$ the corresponding orbit $\mathscr{C}_\lambda^{-1}(p(t), x(t), I(t), \varphi(t))$, one has*

$$|\lambda\Omega I'(t) - \lambda\Omega I'(0)| < \tfrac{1}{2}\left(\frac{\lambda}{\lambda^*}\right)^{-1} E, \tag{2.13a}$$

$$|\hat{h}(p'(t), x'(t)) - \hat{h}(p'(0), x'(0))| < \left(\frac{\lambda}{\lambda^*}\right)^{-1} E, \tag{2.13b}$$

$$|\lambda I(t) - \lambda I(0)| < \frac{3}{4}\left(\frac{\lambda}{\lambda^*}\right)^{-1} E, \tag{2.13c}$$

$$|\hat{h}(p(t), x(t)) - \hat{h}(p(0), x(0))| < \left(\frac{\lambda}{\lambda^*}\right)^{-1} E, \tag{2.13d}$$

for

$$|t| \leqq \min(T, T_0), \tag{2.14}$$

where

$$T = 2^{-5}\Omega^{-1}\left(\frac{\lambda}{\lambda^*}\right)^{-1} e^{(\lambda/\lambda^*)^{1/2}} \rho_\varphi, \tag{2.15}$$

while T_0 is the (possibly infinite) escape time of $(p(t), x(t), I(t), \varphi(t))$ from D_0. T_0 is actually greater than T if the initial datum is further restricted according to

$$\rho_I + \tfrac{3}{4}\left(\frac{\lambda}{\lambda^*}\right)^{-1} E/\Omega \leqq \lambda I(0) \leqq I_0 - \tfrac{3}{4}\left(\frac{\lambda}{\lambda^*}\right)^{-1} E/\Omega,$$

$$|\hat{h}(p(0), x(0))| \leqq \hat{E}_0 - \left(\frac{\lambda}{\lambda^*}\right)^{-1} E, \tag{2.16}$$

$\hat{E}_0$ being the minimum of $\hat{h}$ on the border of G.

Indeed, one has $\dot{I}' = -\lambda^{-1} e^{(\lambda/\lambda^*)^{1/2}} \partial f'_\lambda/\partial\varphi'$, and one can estimate $\partial f'_\lambda/\partial\varphi'$ for real φ' by Cauchy inequality (see Lemma 1 of Sect. 4). So one obtains (2.13a) for t satisfying (2.14). The estimates (2.13b–d) are consequences of (2.13a), taking into account the estimate (2.10c) on I and the conservation of energy to pass from estimates on $\lambda\Omega I$ to estimates on $\hat{h}$. The last statement concerning T_0 is also completely trivial.

Remarks.

1. The estimates (2.10) and (2.12) given in the theorem are global, i.e., refer to the whole of the phase space, but with very minor modifications in the proof one easily obtains more detailed local estimates. Precisely, denote $\tilde{\Delta}_\rho(p, x, I) = \bigcup_{\varphi\in T^1} \Delta_\rho(p, x, I, \varphi)$, and

$$E_\rho(p, x, I) = \sup_{(\tilde{p}, \tilde{x}, \tilde{I}, \tilde{\varphi})\in\tilde{\Delta}_\rho(p,x,I)} |f_\lambda(p, x, I, \varphi)|; \tag{2.17}$$

then, for $(p^0, x^0, I^0) \in G \times [\rho_0/\lambda, I_0/\lambda]$ and $(p, x, I) \in \tilde{\Delta}_{1/2\rho}(p^0, x^0, I^0)$ one gets the analogs of (2.10a–d), with $E_\rho(p^0, x^0, I^0)$ in place of E, and similarly one gets

$$|g_\lambda(p', x', I')| < 2E_\rho(p^0, x^0, I^0), \quad |f'_\lambda(p', x', I', \varphi)| < 8E_\rho(p^0, x^0, I^0) \tag{2.18}$$

as the analog of (2.12). We did not insert this local form of the estimates in the statement and in the proof of the theorem merely to simplify the notation. Some consequences of these local estimates will be pointed out in the conclusions.

2. As already remarked, one can also take the constants ρ_p, ρ_x, ρ_I, ρ_φ appearing

in the statement of the above theorem to be dependent on λ. In particular, it could be convenient to take $\rho = (\lambda^{-\alpha/2}\tilde{\rho}_p, \lambda^{-\alpha/2}\tilde{\rho}_x, \lambda^{-\alpha}\tilde{\rho}_I, \tilde{\rho}_\varphi)$, with some given $(\tilde{\rho}_p, \tilde{\rho}_x, \tilde{\rho}_I, \tilde{\rho}_\varphi)$ independent of λ, which gives $\rho^2 = \lambda^{-\alpha}\tilde{\rho}^2$, $\lambda^* = 2^{11}\hat{E}\tilde{\rho}^2\lambda^\alpha$, and thus (assuming $T_0 \geqq T$)

$$|\hat{h}(t) - \hat{h}(0)| < \left(\frac{\lambda}{\tilde{\lambda}}\right)^{1-\alpha} E, \tag{2.19}$$

for

$$|t| \leqq \tilde{T}_\alpha = 4\Omega^{-1}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{1-\alpha} e^{(\lambda/\tilde{\lambda})^{(1-\alpha)/2}}, \tag{2.20}$$

with $\tilde{\lambda} = (2^{11}E\tilde{\rho}^2)^{1/(1-\alpha)}$.

The advantage of this reformulation of the result is that if one takes, as is possible, $\rho_0 = \rho_I$, the unnatural restriction of the domain from D to D_0 disappears for large λ; moreover, as a byproduct, a very small amount of analyticity of $\hat{h}$ is needed.

3. More precisely, with $\rho_0 = \rho_I$ the domain $[\rho_0/\lambda, I_0/\lambda] \times \mathbf{T}^1$ of the I, φ variables corresponds, for the original variables π, ξ, to the annulus

$$\Omega\tilde{\rho}_I\lambda^{-\alpha} \leqq \tfrac{1}{2}(\pi^2 + \lambda^2\Omega^2\xi^2) \leqq \Omega I_0. \tag{2.21}$$

Although perturbation theory does not work directly in the "hole" $\frac{1}{2}(\pi^2 + \lambda^2\Omega^2\rho^2) < \Omega\tilde{\rho}_I\lambda^{-\alpha}$, it is quite clear that orbits with initial datum $I(0)$ in the hole cannot escape from a small neighborhood of the hole for $|t| \leqq \tilde{T}_\alpha$. Taking for example $\alpha = \frac{1}{2}$, one obtains that the oscillation of $\hat{h}$ is bounded as in (2.19), for $|t| \leqq \tilde{T}_{1/2}$ and any initial datum, including those inside the hole.

3. Proof of the Theorem

The theorem is a direct consequence of the following

Iterative Lemma. *Let the Hamiltonian*

$$H_\lambda^{(k)} = \lambda\Omega I + \hat{h}(p,x) + \lambda^{-1}g_k(p,x,I,\lambda) + \lambda^{-1-k}f_k(p,x,I,\varphi,\lambda) \tag{3.1}$$

be analytic in $D_{\rho-k\delta}$, *with* $\rho = (\rho_p, \rho_x, \rho_I, \rho_\varphi)$, $\rho_\varphi \leqq 1$, *and* $\delta < \rho/(k+1)$; *assume*

$$\|\hat{h}\|_{\rho-k\delta} \leqq \hat{E}, \tag{3.2a}$$

$$\|g_k\|_{\rho-k\delta} \begin{cases} \leqq E\sum_{l=0}^{k-1}(\Lambda/\lambda)^l & \text{if } k>0 \\ = 0 & \text{if } k=0 \end{cases}, \tag{3.2b}$$

$$\|f_k\|_{\rho-k\delta} \leqq \Lambda^k E, \tag{3.2c}$$

with $\hat{E} \geqq E$ *and* $\Lambda = 2^8\delta^{-2}\Omega^{-1}\hat{E}$; *assume also*

$$\lambda \geqq 2\Lambda. \tag{3.3}$$

Then there exists a canonical transformation $(p,x,I,\varphi) = \mathscr{C}_\lambda^{(k)}(p',x',I',\varphi')$, $\mathscr{C}_\lambda^{(k)}$:

$D_{\rho-(k+1)\delta} \to D_{\rho-k\delta}$, *bounded by*

$$
\begin{aligned}
|p_i - p_i'| &\leqq 2^{-4}\lambda^{-1}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_p,\\
|x_i - x_i'| &\leqq 2^{-4}\lambda^{-1}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_x,\\
|I - I'| &\leqq 2^{-7}\lambda^{-1}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_I\delta_\varphi,\\
|\varphi - \varphi'| &\leqq 2^{-4}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_\varphi,
\end{aligned}
\tag{3.4}
$$

such that the new Hamiltonian $H_\lambda^{(k+1)} \equiv H^{(k)} \circ \mathscr{C}_\lambda^{(k)}$ *has the form (3.1), and satisfies (3.2), with* $k+1$ *in place of* k.

Proof of the Iterative Lemma. Given $H_\lambda^{(k)}$ of the form (3.1), we generate the canonical transformation $\mathscr{C}_\lambda^{(k)}$ as the time-one solution of an auxiliary Hamiltonian problem, with a suitably chosen Hamiltonian χ: $D_{\rho-k\delta} \to C$.

Denoting $\bar{f}_k = (2\pi)^{-1} \int_0^{2\pi} f_k d\varphi$, so that $\|\bar{f}_k\|_{\rho-k\delta} \leqq \|f\|_{\rho-k\delta} \leqq \Lambda^k E$, our choice of χ is

$$
\chi = \lambda^{-2-k}\Omega^{-1}\int_0^\varphi (f_k - \bar{f}_k)d\varphi. \tag{3.5}
$$

The reason of this choice is the following: first of all, using $\delta_\varphi < \rho_\varphi \leqq 1$, from the above expression for χ one finds

$$
\|\chi\|_{\rho-k\delta} \leqq \lambda^{-2-k}\Omega^{-1}\sqrt{\pi^2+\delta_\varphi^2}\,\|f_k - \bar{f}_k\|_{\rho-k\delta} \leqq 8E\lambda^{-2}\Omega^{-1}(\Lambda/\lambda)^k; \tag{3.6}
$$

recalling then (3.3) and the expression of Λ, one deduces $\|\chi\|_{\rho-k\delta} \leqq 2^{-6}(\Lambda/\lambda)^k\lambda^{-1}\delta^2 < \frac{1}{4}\lambda^{-1}\delta^2$: thus, according to Lemma 3 of Sect. 4, with in our case $\sigma = \lambda^{-1}\delta^2$, the solution $\Phi^t(p', x', I', \varphi')$ of the Hamiltonian system with Hamiltonian χ, corresponding to initial datum (p', x', I', φ'), is certainly defined for $t=1$, and one can set $\mathscr{C}_\lambda^{(k)} = \Phi^1$. Then one can consider the new Hamiltonian $H_\lambda^{(k+1)} = H_\lambda^{(k)} \circ \Phi^1$, and decompose it in the following way:

$$
\begin{aligned}
H_\lambda^{(k+1)} = {} & \lambda\Omega I' + \hat{h} + \lambda^{-1}g_k + \lambda^{-1-k}\bar{f}_k\\
& + \lambda^{-1-k}(f_k - \bar{f}_k) + \{\chi, \lambda\Omega I'\}\\
& + R_1[\hat{h}] + R_1[\lambda^{-1}g_k + \lambda^{-1-k}f_k] + R_2[\lambda\Omega I'],
\end{aligned}
\tag{3.7}
$$

where $R_1[F] = F \circ \Phi^1 - F$, and $R_2[F] = F \circ \Phi^1 - F - \{\chi, F\}$.

The second line of this expression contains φ'-dependent terms of order λ^{-1-k}, but vanishes just because of the choice (3.5) of χ. The new Hamiltonian $H^{(k+1)}$ has then the form (3.1), if one sets

$$
g_{k+1} = g_k + \lambda^{-k}\bar{f}_k \tag{3.8a}
$$

$$
\lambda^{-2-k}f_{k+1} = R_1[\hat{h}] + R_1[\lambda^{-1}g_k + \lambda^{-1-k}f_k] + R_2[\lambda\Omega I']. \tag{3.8b}
$$

One must now show that the inequalities (3.2) are satisfied with $k+1$ in place of k. Clearly (3.2a) is trivially true, while (3.2b) immediately follows from $\|\bar{f}_k\| \leqq \Lambda^k E$. To achieve (3.2c) we must use Lemma 2 and Lemma 3 to estimate the remainders. By (4.8) of Lemma 3 one has

$$
\|R_1[\hat{h}]\|_{\rho-(k+1)\delta} \leqq \|\{\chi, \hat{h}\}\|_{\rho-(k+1/2)\delta}, \tag{3.9}
$$

and by Lemma 2, recalling that $\hat{h}$ is independent of I and φ, so that one can take $\sigma=(\delta_p/2)/(\delta_x/2)\geqq\frac{1}{4}\delta^2$, one obtains

$$\|R_1[\hat{h}]\|_{\rho-(k+1)\delta}\leqq 4\delta^{-2}\|\chi\|_{\rho-k\delta}\hat{E}. \tag{3.10}$$

In a similar way, using $\|g_k\|_{\rho-k\delta}<2E$, which trivially follows from (3.2b) and (3.3), and taking $\sigma=\min[(\delta_p/2)(\delta_x/2),\ \lambda^{-1}(\delta_I/2)(\delta_\varphi/2)]\geqq\frac{1}{4}\lambda^{-1}\delta^2$ in Lemma 2, one immediately finds

$$\|R_1[\lambda^{-1}g_k+\lambda^{-1-k}f_k]\|_{\rho-(k+1)\delta}\leqq 12\delta^{-2}\|\chi\|_{\rho-k\delta}E. \tag{3.11}$$

Finally, using (4.9) of Lemma 3, and recalling that $\{\chi,\lambda\Omega I\}=-\lambda^{-1-k}(f_k-\bar{f}_k)$ by the choice (3.5) of χ, one finds

$$\|R_2[\lambda\Omega I']\|_{\rho-(k+1)\delta}\leqq 4\delta^{-2}\|\chi\|_{\rho-k\delta}E, \tag{3.12}$$

so that, recalling the definition (3.8b) of f_{k+1}, the estimate (3.6) of $\|\chi\|_{\rho-k\delta}$, and the definition of Λ, one finally has

$$\|f_{k+1}\|_{\rho-(k+1)\delta}<2^8\Lambda^k\delta^{-2}\Omega^{-1}E\hat{E}=\Lambda^{k+1}E, \tag{3.13}$$

as required.

To complete the proof of the iterative lemma, one must verify the estimates (3.4) on the canonical transformation. These however are trivial consequences of (4.7) of Lemma 3: precisely, one has

$$\begin{aligned}|p_i-p_i'|\leqq\left\|\frac{\partial\chi}{\partial x_i}\right\|_{\rho-(k+1/2)\delta}&\leqq 2\delta_x^{-1}\|\chi\|_{\rho-k\delta}\\ &\leqq 2^4\delta_x^{-1}\lambda^{-2}\Omega^{-1}(\Lambda/\lambda)^kE\\ &\leqq 2^{-4}\lambda^{-1}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_p,\end{aligned} \tag{3.14}$$

and similarly for $|x_i-x_i'|$. Concerning $|\varphi-\varphi'|$, one has instead

$$\begin{aligned}|\varphi-\varphi'|\leqq\left\|\frac{\partial\chi}{\partial I}\right\|_{\rho-(k+1/2)\delta}&\leqq 2\delta_I^{-1}\lambda\|\chi\|_{\rho-k\delta}\\ &<2^{-4}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_\varphi,\end{aligned} \tag{3.15}$$

while to estimate $|I-I'|$ one can use the explicit expression (3.5) of χ, obtaining

$$\begin{aligned}|I-I'|\leqq\left\|\frac{\partial\chi}{\partial\varphi}\right\|_{\rho-(k+1/2)\delta}&\leqq 2\lambda^{-2-k}\Omega^{-1}\Lambda^kE\\ &=2\lambda^{-1}(\Lambda/\lambda)^{k+1}\Lambda^{-1}\Omega^{-1}E\\ &\leqq 2^{-7}(\Lambda/\lambda)^{k+1}(E/\hat{E})\delta_I\delta_\varphi.\end{aligned} \tag{3.16}$$

The iterative lemma is thus proven.

We now come back to the proof of the theorem. To this purpose, we make use of the iterative lemma for $k=0,\dots,r-1$, with $\delta=\rho/2r$, so that the constant Λ appearing in the lemma assumes the value $\Lambda=2^{10}\rho^{-2}\Omega^{-1}\hat{E}r^2\equiv\Lambda_0r^2$.

According to (3.3), one can choose any value of r not exceeding $\tilde{r}=(\lambda/2\Lambda_0)^{1/2}$, and set $H_\lambda'=H_\lambda^{(r)}$. The value of r is now precisely chosen in the interval $[1,\tilde{r}]$ in order to optimize the estimate on the φ'-dependent part of the $H_\lambda^{(r)}$, i.e., according to (3.2c), in order to minimize the quantity

$$R = \lambda^{-1}(\Lambda/\lambda)^r E = \lambda^{-1}(r^2 \Lambda_0/\lambda)^r E. \tag{3.17}$$

An elementary computation gives for r the optimal (noninteger) value

$$r^* = e^{-1}(\lambda/\Lambda_0)^{1/2}, \tag{3.18}$$

which is less than $\tilde{r}$, and greater than one if λ satisfies the further requirement $\lambda > \Lambda_0 e^2$. Clearly, one should treat only this case, because for $\lambda \leqq \Lambda_0 e^2 < 2^{13}\rho^{-2}\hat{E} = 4\lambda^*$, λ^* being given by (3.3), the theorem is trivially true (and completely useless). The best choice of r is then given by the integer satisfying

$$r^* - 1 < r \leqq r^*, \tag{3.19}$$

which leads to

$$\begin{aligned} R &< \lambda^{-1} e^{-2(r^*-1)} E = \lambda^{-1} e^{-\sqrt{4\lambda/\Lambda_0 e^2}+2} E \\ &< \lambda^{-1} e^{-\sqrt{\lambda/\lambda^*}+2} E, \end{aligned} \tag{3.20}$$

and thus to the second of (2.12).

The first of (2.12) is a trivial consequence of the corresponding inequality (3.2b), while (2.10a–d) follow from (3.4) by a sum on k: for example, one has

$$\begin{aligned} |p_i - p_i'| &\leqq 2^{-4}\lambda^{-1}(E/\hat{E})\delta_p \sum_{k=0}^{r-1} (\Lambda/\lambda)^{k+1} \\ &< 2^{-4}\lambda^{-1}(E/\hat{E})(\rho_p/2r) \\ &< 2^{-3}\lambda^{-1}(\lambda/\lambda^*)^{-1/2}(E/\hat{E})\rho_p, \end{aligned} \tag{3.21}$$

where $2r > r^*$ has been used.

The canonical transformation obtained by iteration of the lemma is clearly analytic in $D_{\rho/2}$, while the estimates (3.4) also imply that $\mathscr{C}_\lambda(D_{\rho/2}) \supset D_{\rho/4}$, as claimed. This completes the proof of the theorem. The idea to choose the order r as a function of the perturbative parameter as was done in (3.19), is the heart of Nekhoroshev's exponential estimate; [9] the optimization procedure followed here can be found in ref. [12].

4. Technical Lemmas

We introduce here a few elementary lemmas, used in the estimates of canonical transformations.

Lemma 1 (Cauchy inequality). *Let $f(z_1, \ldots, z_l)$ be defined in the polydisk*

$$\Delta = \{(z_1, \ldots, z_l) \in \mathbf{C}^l; \quad |z_i - z_i^0| \leqq \gamma_i\}, \tag{4.1}$$

$\gamma_1, \ldots, \gamma_l$ being positive constants, and analytic in the interior of Δ. Then one has, for $i = 1, \ldots, l$,

$$\left| \frac{\partial f}{\partial z_i}(z_1^0, \ldots, z_l^0) \right| \leqq \gamma_i^{-1} \| f \|, \tag{4.2}$$

where $\|\cdot\|$ denotes the supremum norm on Δ.

Proof. The proof is a well known easy application of Cauchy integral formula.

Lemma 2 (on Poisson brackets). *Let $(P, Q) = (P_1, \ldots, P_l, Q_1, \ldots, Q_l)$ be canonical variables and assume $f(P, Q)$, $g(P, Q)$ are defined in the polydisc*

$$\Delta = \{(P, Q) \in \mathbf{C}^{2l}; |P_i - P_i^0| \leqq \alpha_i, |Q_i - Q_i^0| \leqq \beta_i, i = 1, \ldots, l\}, \tag{4.3}$$

and are analytic in its interior; denote by $\|\cdot\|$ the supremum norm in Δ. Then for the Poisson bracket $\{f, g\}$ one has the estimate

$$|\{f, g\}\{P_0, Q_0)| \leqq \sigma^{-1} \|f\| \, \|g\|, \tag{4.4}$$

with $\sigma = \min_i \alpha_i \beta_i$.

Proof. By definition it is $\{f, g\} = \sum_{i=1}^{l} (\partial f/\partial p_i)(\partial g/\partial Q_i) - (\partial f/\partial Q_i)(\partial g/\partial P_i)$. Then, denoting

$$G(t) = g\left(P^0 - t\frac{\partial f}{\partial Q}(P^0, Q^0), Q^0 + t\frac{\partial f}{\partial P}(P^0, Q^0)\right), \tag{4.5}$$

one has $\{f, g\}(P^0, Q^0) = (dG/dt)(0)$. Now, one has clearly $|G(t)| \leqq \|g\|$ for t such that $|t(\partial f/\partial Q_i)(P^0, Q^0)| \leqq \alpha_i$ and $|t(\partial f/\partial P_i)(P^0, Q^0)| \leqq \beta_i$ for $i = 1, \ldots, l$, i.e., using the Cauchy inequality (4.2) to estimate the derivatives of f in (P^0, Q^0), certainly for $|t| \leqq T = \|f\|^{-1} \min_{1 \leqq i \leqq l} \alpha_i \beta_i$. Using again Cauchy inequality to estimate $(dG/dt)(0)$, (4.4) immediately follows.

Lemma 3 (on canonical transformations). *Let $(P_1, \ldots, P_l, Q_1, \ldots, Q_l)$ be canonical coordinates in a domain $D \subset \mathbf{C}^{2l}$, and for a given $\delta = (\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_l)$, denote by $\Delta_\delta(P, Q)$ the polydisc $\{(P', Q') \in \mathbf{C}^{2l};\ |P'_i - P_i| \leqq \alpha_i,\ |Q'_i - Q_i| \leqq \beta_i,\ i = 1, \ldots, l\}$, and by $D - \delta$ the subset of D defined by $D - \delta = \{(P, Q) \in D; \Delta_\delta(P, Q) \subset D\}$. For any $D' \subset D$, denote by $\|\cdot\|_{D'}$ the supremum norm in D'.*

Consider now any Hamiltonian $\chi(P, Q)$ defined in D, and analytic in its interior; denote by $\Phi^t(P, Q) = (P^t(P, Q), Q^t(P, Q))$ the solution of the corresponding equations of motion at time t, with initial datum (P, Q), and assume

$$4|t| \, \|\chi\|_D \leqq \sigma \equiv \min_{1 \leqq i \leqq l} \alpha_i \beta_i. \tag{4.6}$$

Then Φ^t is an analytic canonical transformation: $D - \delta \to D$, which for $1 \leqq i \leqq l$ satisfies the estimates

$$\begin{aligned} |P_i^t - P_i| &\leqq |t| \left\| \frac{\partial \chi}{\partial Q_i} \right\|_{D - \delta/2}, \\ |Q_i^t - Q_i| &\leqq |t| \left\| \frac{\partial \chi}{\partial P_i} \right\|_{D - \delta/2}. \end{aligned} \tag{4.7}$$

Moreover, for any function F analytic in the interior of D, the transformed function $F^t = F \circ \Phi^t$ satisfies the estimates

$$R_1^t[F] \equiv \|F^t - F\|_{D - \delta} \leqq |t| \, \|\{\chi, F\}\|_{D - \delta/2} \tag{4.8}$$

and

$$R_2^t[F] \equiv \| F^t - F - t\{\chi, F\} \|_{D-\delta} \leqq \frac{|t|^2}{2} \| \{\chi, \{\chi, F\}\} \|_{D-\delta/2} . \tag{4.9}$$

Proof. By Cauchy inequality (4.2) one has

$$\begin{aligned} \left\| \frac{\partial \chi}{\partial P_i} \right\|_{D-\delta/2} &\leqq 2\alpha_i^{-1} \| \chi \|_D , \\ \left\| \frac{\partial \chi}{\partial Q_i} \right\|_{D-\delta/2} &\leqq 2\beta_i^{-1} \| \chi \|_D . \end{aligned} \tag{4.10}$$

Then the existence and uniqueness theorem for the solutions of ordinary differential equations allows one to conclude that for initial data $(P,Q) \in D - \delta$ the solution $\Phi^t(P,Q)$ is well defined, and does not escape $D - \delta/2$, for t satisfying the inequality (4.6). Moreover, $\Phi^t(P,Q)$ is analytic both as a function of t and of (P,Q), and is canonical for each t.

Then, to achieve (4.7), one simply writes $|P_i^t - P_i| \leqq \left| \int_0^t (\partial \chi/\partial Q_i)(P^s, Q^s) ds \right| \leqq |t| \, \| (\partial \chi / \partial Q_i) \|_{D-\delta/2}$, and similarly for $|Q_i^t - Q_i|$, while to achieve (4.8) one makes use of Lemma 2, and writes the identity

$$F^t(P,Q) - F(P,Q) = \int_0^t \{\chi, F\} (\Phi^{t'}(P,Q)) dt' , \tag{4.11}$$

which gives

$$\| F^t - F \|_{D-\delta} \leqq |t| \, \| \{\chi, F\} \|_{D-\delta/2} . \tag{4.12}$$

Finally, to obtain (4.9), one writes

$$\begin{aligned} F^t(P,Q) - F(P,Q) - t\{\chi, F\}(P,Q) &= \int_0^t dt' \int_0^{t'} dt'' \{\chi, \{\chi, F\}\} (\Phi^{t''}(P,Q)) , \\ \| F^t - F - t\{\chi, F\} \|_{D-\delta} &\leqq \int_0^{|t|} dt' \int_0^{t'} dt'' \, \| \{\chi, \{\chi, F\}\} \|_{D-\delta/2} \\ &\leqq \tfrac{1}{2} |t|^2 \, \| \{\chi, \{\chi, F\}\} \|_{D-\delta/2} . \end{aligned} \tag{4.13}$$

5. Concluding Remarks

As already commented in the introduction, the theorem stated in Sect. 2 naturally applies to the problem of the introduction of a single constraint in a dynamical system, in order to reduce the effective number of degrees of freedom from $n+1$ to n. No essential difference is expected when v constraints are introduced, if the frequency vector is $\omega = \lambda \Omega$, with λ large and $\Omega = (\Omega_1, \ldots, \Omega_v)$ satisfying a Diophantine condition, apart from a bad v-dependence in all constants entering the theorem. A more interesting case is that of resonant frequencies, in particular completely resonant ones, because of its relevance for statistical mechanics. Unfortunately, it is not possible to extend the present technique to the resonant case, because for resonant systems the actions are not separately constant, and in particular cannot be kept out of zero, where the Hamiltonian is no more analytic. So the resonant case will be treated in the forthcoming second part of this work with a different technique.

The existence of time scales growing exponentially with λ was exhibited by numerical computations, reported in ref. [14] for the case $\nu = 1$, and in ref. [14, 16] for $\nu > 1$. These numerical experiments are quite delicate, and must be considered very carefully, but in any case they give some evidence of a λ-dependence of the time-scale of the form $\exp(\lambda/\lambda^*)$, instead of the weaker form $\exp(\lambda/\lambda^*)^a$, $a \leq \frac{1}{2}$, which we find in our perturbative approach. This fact could be an indication that the perturbative approach is still far from providing a complete understanding of the mechanism which leads to the exponentially long time scales.

As a final remark, let us comment on the possible use of the above theorem to support the ideas expressed by Jeans in his 1903 paper [7]. The heart of this paper is the study of the collision of a point-particle with an internally structured, fast vibrating molecule; Jeans's purpose is to show that at the end of the collision the vibrational energy of the molecule is left unchanged, up to an extremely small amount, exponentially small with λ. This is different from the conclusion obtained in the corollary of Sect. 2, because Jeans looks for an exponentially small energy exchange during a single collision, i.e. in a short time, while we did consider exponentially long time scales, allowing much larger energy exchanges.

In fact, it seems to be very easy to apply our theorem to Jeans's case, if one considers (as Jeans does) the case of short-range interaction potentials, say potentials decaying with the distance r like e^{-r/r_0}. Indeed, take as the time origin a moment when the two colliding molecules are still far apart, at a distance $r_1 \gg r_0$, precisely $r_1 = (\lambda/\lambda^*)^{1/2} r_0$, and consider the energy exchange after a time t_1, when the molecules are again at a distance r_1; due to the choice of r_1, it turns out that t_1 is of order $(\lambda/\lambda^*)^{1/2}$, so that, from the expression (2.11) of the new Hamiltonian, in place of (2.13a) one obtains $|\lambda\Omega I'(t_1) - \lambda\Omega I'(0)| \sim (\lambda/\lambda^*)^{1/2} e^{-(\lambda/\lambda^*)^{1/2}}$. To obtain the corresponding energy exchange in the old variables, i.e. the quantity $|\lambda\Omega I(t_1) - \lambda\Omega I(0)|$ (which corresponds to Jeans's purpose), one must add to the above exponential estimate the contribution of the canonical transformation, precisely $|\lambda\Omega I(t_1) - \lambda\Omega I'(t_1)| + |\lambda\Omega I(0) - \lambda\Omega I'(0)|$. To estimate such quantity one can use (2.10c): however, the estimate being necessary only in a region with $r \simeq r_1$, one can take $E \sim e^{-r_1/r_0} \sim e^{-(\lambda/\lambda^*)^{1/2}}$, so that the exponential estimate for the energy transfer is preserved, and Jeans's conjecture (apart from the precise form of the exponential dependence) is thus proven.

This comment has been inserted here mainly as a tribute to the ideas and the intuition of Jeans, which appear today to have been essentially right.

Appendix: The Realization of Holonomic Constraints

The aim of this Appendix is to recall the problem of the realization of constraints of classical mechanics [1–5]; the basic purpose is to justify the interest in Hamiltonian systems of the form (1), to which the present paper is devoted. For obvious simplicity reasons, our exposition will be not as formal as it could be; for a more formal introduction to the problem, the reader is deferred to the above quoted references, in particular to ref. [3] which we will follow more closely.

Let us consider a Lagrangian system with $n+\nu$ degrees of freedom, and assume it reduces to n degrees of freedom by the introduction of ν holonomic constraints. Assume one is able to introduce adapted coordinates (x,ξ), $x=(x_1,\ldots,x_n)$, $\xi=(\xi_1,\ldots,\xi_\nu)$, such that $\xi_1=0,\ldots,\xi_\nu=0$ are the equations of the constraints. If the free system is described by the "free Lagrangian"

$$L(x,\xi,\dot{x},\dot{\xi})=T(x,\xi,\dot{x},\dot{\xi})-V(x,\xi), \tag{A.1}$$

then the constrained system is well known to be described by the "constrained Lagrangian"

$$\begin{aligned}\hat{L}(x,\dot{x})&=\hat{T}(x,\dot{x})-\hat{V}(x),\\ \hat{T}(x,\dot{x})&\equiv T(x,0,\dot{x},0),\\ \hat{V}(x)&\equiv V(x,0).\end{aligned} \tag{A.2}$$

For example, for a point-mass m confined to a spherical surface of radius R around the origin, using the ordinary spherical coordinates (r,θ,φ) and denoting $x=(\theta,\varphi)$, $\xi=r-R$, one has

$$\begin{aligned}L&=\tfrac{1}{2}m[\dot{\xi}^2+(\xi+R)^2\dot{\theta}^2+((\xi+R)\sin\theta)^2\dot{\varphi}^2]-V(\theta,\varphi,\xi),\\ \hat{L}&=\tfrac{1}{2}m(R^2\dot{\theta}^2+R^2\sin\theta^2\dot{\varphi}^2)-V(\theta,\varphi,0).\end{aligned} \tag{A.3}$$

This purely geometrical introduction of constraints is easy and powerful, but completely ignores the physics, i.e., how constraining forces are practically produced. Physically one expects that the constraining forces are the elastic reactions to very small deformations of some material objects, characterized by very large elastic constants. In the above example, one can think the mass m is bonded to the origin by a spring of unstretched length R and large elastic constant; for more general systems, one would like to represent each (holonomic and bilateral) constraining device by a suitable "confining potential" $\lambda^2W(x,\xi)$, with $W(x,0)=0$, and $W(x,\xi)>0$ for $\xi\neq 0$. This means one aims to study, in place of the already constrained system (A.2), the complete Lagrangian system

$$L_\lambda(x,\xi,\dot{x},\dot{\xi})=T(x,\xi,\dot{x},\dot{\xi})-V(x,\xi)-\lambda^2W(x,\xi). \tag{A.4}$$

The physical intuition is that one should recover, in some sense, the constrained system, in the limit $\lambda\to\infty$.

Let

$$T=\tfrac{1}{2}\sum_{i,j=1}^{n}A_{ij}(x,\xi)\dot{x}_i\dot{x}_j+\tfrac{1}{2}\sum_{i,j=1}^{\nu}B_{ij}(x,\xi)\dot{\xi}_i\dot{\xi}_j+\sum_{i=1}^{n}\sum_{j=1}^{\nu}C_{ij}(x,\xi)\dot{x}_i\dot{\xi}_j \tag{A.5}$$

be the kinetic energy of the system. Now, one can always introduce adapted coordinates, such that one has $C_{ij}(x,0)=0$ (orthogonality of the x and ξ coordinates on the constraint), and $B_{ij}(x,0)=\delta_{ij}$ (as is obtained by the substitution $\xi=B^{-1/2}(x,0)\xi'$ in (A.4), which is possible because the matrix B is symmetric and positive definite). For example, the Lagrangian (A.3) is already written in the adapted coordinates, apart from the trivial substitution $\xi=m^{-1/2}\xi'$. Using adapted coordinates, the Lagrangian (A.4) can be given the form

$$L_\lambda(x,\xi,\dot{x},\dot{\xi}) = \hat{L}(x,\dot{x}) + \tfrac{1}{2}\sum_{i=1}^{\nu} \dot{\xi}_i^2 - \lambda^2 W(x,\xi) + F(x,\xi,\dot{x},\dot{\xi}), \tag{A.6}$$

where $\hat{L}$ is given by (A.2), and F vanishes for vanishing ξ.

Denote by $(X_\lambda^t(x,\xi,\dot{x},\dot{\xi}),\ \Xi_\lambda^t(x,\xi,\dot{x},\dot{\xi}))$ the solution to the Lagrangian problem (A.6), with initial datum $(x,\xi,\dot{x},\dot{\xi})$. The basic question posed in refs. [1–5] is whether, at fixed t, one can achieve

$$\lim_{\lambda\to\infty} X_\lambda^t(x,0,\dot{x},\dot{\xi}) = \hat{X}^t(x,\dot{x}), \tag{A.7}$$

$\hat{X}^t$ being the solution of the constrained Lagrangian problem (A.2). As is proven in the above quoted references, a sufficient condition for the validity of (A.7) is that $W(x,\xi)$ does not depend on x, i.e. one has $W(x,\xi) = w(\xi)$. Such a condition is clearly satisfied by our model example (A.3), if the spring connecting the mass m to the origin is isotropic, but otherwise is not. As widely discussed for example in ref. [3], such an apparently strong condition cannot in general be released, unless in (A.7) one imposes $\dot{\xi} = 0$ in the initial datum, beside $\xi = 0$.

Passing now to the Hamiltonian formalism, and assuming the x-independence of the confining potential, one clearly obtains, from (A.6), a Hamiltonian of the form

$$H(p,x,\pi,\xi) = \tfrac{1}{2}\sum_{i=1}^{\nu} \pi_j^2 + \lambda^2 w(\xi) + \hat{h}(p,x) + f(p,x,\pi,\xi), \tag{A.8}$$

where $p = (p_1,\dots,p_n)$ and $\pi = (\pi_1,\dots,\pi_\nu)$ are the momenta conjugated to x and ξ, while $\hat{h}$ is the Hamiltonian corresponding to $\hat{L}$, and f collects the terms which vanish for $\xi = 0$. Such an Hamiltonian already has the form (1.1); the expression (1.2) of h_ω corresponds to the particular choice $w(\xi) = \frac{1}{2}\sum_{i=1}^{\nu} \Omega_i^2\xi_i^2$. The only relevant point in this choice is that none of the Ω_i should vanish, while the possible presence of higher order terms in ξ would be completely irrelevant; these terms have not been inserted just for simplicity.

As explained in the introduction, we are interested in the possible existence of long time-scales, growing exponentially with λ according to (1.5), on which the energy exchanges between the x and ξ degrees of freedom are almost forbidden, so that in particular $\hat{h}$ is practically constant, as in the case of an ideally constrained system. It is perhaps worth noticing that on such long time scales it is not conceivable to obtain point-wise convergence of orbits, comparable with (A.7), which holds only for $\lambda \to \infty$ at fixed t. Indeed, it should be clear that, for example, whenever the constrained system to be realized has unstable orbits (positive Lyapunov exponents), the limit (A.7) must be highly non-uniform in time; in fact, from the proofs reported in the above quoted references (which essentially rely on the uniqueness theorem for the solutions of ordinary differential equations) one only gets quite poor estimates of the form

$$\operatorname{dist}(X_\lambda^t, \hat{X}^t) \leq K\lambda^{-1}e^{\mu t}, \quad K,\ \mu > 0, \tag{A.9}$$

which, although being in general optimal, lose any usefulness after a time scale of order $\lambda^{-1} \log \mu$; only in the particular case of integrable systems, it should be possible to replace the exponential of time in (A.9) by a linear function, but in any case one would remain far from the long time scale (1.5).

References

1. Rubin, H. Hungar, P.: Commun. Pure Appl. Math. **10**, 65 (1957)
2. Arnold, V. I.: Méthodes mathématiques de la mécanique classique Moscow: MIR 1976, p. 97
3. Gallavotti, G.: Meccanica elementare Boringhieri, Torino: 1980 [English edition: The elements of mechanics. Berlin, Heidelberg, New York: Springer 1983], Chap. 3.
4. Takens, F.: Motion under the influence of a strong constraining force. In: Global theory of dynamical systems. Nitecki Z. Robinson C. (eds.), Lecture Notes in Mathematics, Vol. 819. Berlin, Heidelberg, New York: Springer 1979
5. Van Kampen, N. G.: Elimination of fast variables. Phys. Rep. **124**, 69 (1985)
6. Boltzmann, L.: Nature **51**, 413 (1895)
7. Jeans, J. H.: On the vibrations set up in molecules by collisions. Phil. Mag. **6**, 279 (1903)
8. Jeans, J. H.: On the partition of energy between matter and aether. Phil. Mag. **10**, 91 (1905)
9. Nekhoroshev, N. N.: Usp. Mat. Nauk. **32**, 5 (1977) [Russ. Math. Surv. **32**, 1 (1977); Tr. Sem. Petrows. No. 5, 5 (1979)]
10. Benettin, G. Galgani, L. Giorgilli, A.: Celest. Mech. **37**, 1 (1985)
11. Gallavotti, G.: Quasi integrable mechanical systems. In: Critical phenomena, random systems, gauge theories. Osterwalder, K., Stora, R. (eds.). Amsterdam, New York, Oxford, Tokyo: North-Holland 1986
12. Giorgilli, A. Galgani, L.: Celest. Mech. **37**, 95 (1985)
13. Benettin, G. Gallavotti, G.: Stability of motions near resonances in quasi-integrable Hamiltonian systems. Stat. Phys. **44**, 293 (1986)
14. Benettin, G : Examples of long time scales in Hamiltonian dynamical systems. In: Proceedings of the L'Aquila Conference "Recent Advances in Hamiltonian Dynamics," June 1986 (to appear)
15. Benettin, G. Galgani, L. Giorgilli, A.: Boltzmann's ultraviolet cutoff and Nekhoroshev's theorem on Arnold diffusion. Nature **311**, 444 (1985)
16. Benettin, G. Galgani, L. Giorgilli, A.: Exponential law for the equipartition times among translational and vibrational degrees of freedom. Phys. Lett. **A 120**, 23 (1987)

Communicated by J.-P. Eckmann

Received January 25, 1987; in revised from April 16, 1987

Note added in proofs. A problem related to ours was studied by A. I. Neishtadt [PPM USSR Vol. 48, no. 2, 133–139, 1984], in giving estimates for adiabatic invariants. The main difference is that our problem requires to consider for the action I a domain with a radius inversely proportional to λ (see Sect. 2), and this significantly alters the perturbative scheme. We are indebted to Prof. J. Moser for indicating to us Neishtadt paper. Earlier rigorous estimates for normal forms were obtained in J. Moser, Nachr, Akad. Wiss. Goett., 1955, 87.