Available online at www.sciencedirect.com
SCIENCE DIRECT
J. Math. Pures Appl. 82 (2003) 613–664
JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES
www.elsevier.com/locate/matpur

Drift in phase space: a new variational mechanism with optimal diffusion time ☆

Massimiliano Berti [a,*], Luca Biasco [b], Philippe Bolle [c]

[a] *SISSA, Via Beirut 2-4, 34014, Trieste, Italy*
[b] *Dipartimento di Matematica, Università "Roma Tre", Largo S.L. Murialdo 1, 00146 Roma, Italy*
[c] *Département de mathématiques, Université d'Avignon, 33, rue Louis Pasteur, 84000 Avignon, France*

Received 9 November 2002

Abstract

We consider nonisochronous, nearly integrable, a priori unstable Hamiltonian systems with a (trigonometric polynomial) $O(\mu)$-perturbation which does not preserve the unperturbed tori. We prove the existence of Arnold diffusion with diffusion time $T_d = O((1/\mu)\ln(1/\mu))$ by a variational method which does not require the existence of "transition chains of tori" provided by KAM theory. We also prove that our estimate of the diffusion time T_d is optimal as a consequence of a general stability result derived from classical perturbation theory.
© 2003 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Résumé

Nous considérons des systèmes hamiltoniens presque intégrables, non isochrones et a priori instables par une perturbation en $O(\mu)$ qui ne préserve pas tels quels les tores invariants du système non perturbé (et qui est un polynôme trigonométrique). Nous montrons l'existence de la diffusion d'Arnold avec un temps de diffusion $T_d = O((1/\mu)\ln(1/\mu))$ par une méthode variationnelle qui n'impose pas de passer par des "chaînes de tores de transition" et par la théorie KAM. Nous montrons aussi que notre estimation du temps de diffusion T_d est optimale : c'est une conséquence d'un résultat général de stabilité qui provient de la théorie classique des perturbations.
© 2003 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

MSC: 37J40; 37J45

Keywords: Arnold diffusion; Variational methods; Shadowing theorem; Perturbation theory; Nonlinear functional analysis

☆ Supported by MURST Variational Methods and Nonlinear Differential Equations.
* Corresponding author.
E-mail addresses: berti@sissa.it (M. Berti), biasco@mat.uniroma3.it (L. Biasco), philippe.bolle@univ-avignon.fr (P. Bolle).

0021-7824/$ – see front matter © 2003 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.
doi:10.1016/S0021-7824(03)00032-1

1. Introduction and main results

Topological instability of action variables in multidimensional nearly integrable Hamiltonian systems is known as Arnold Diffusion. For autonomous Hamiltonian systems with two degrees of freedom KAM theory generically implies topological stability of the action variables, i.e., under the flow of the perturbed system the action variables stay close to their initial values for all times. On the contrary, for systems with more than two degrees of freedom, outside a large set of initial conditions provided by KAM theory, the action variables may undergo a drift of order one in a very long, but finite time called the "diffusion time". Arnold first showed up this instability phenomenon for a peculiar Hamiltonian in the famous paper [2].

As suggested by normal form theory near simple resonances, the Hamiltonian models which are usually studied have the form $H(I,\varphi,p,q) = (I_1^2/2) + \omega \cdot I_2 + (p^2/2) + \varepsilon(\cos q - 1) + \varepsilon\mu f(I,\varphi,p,q)$ where ε and μ are small parameters, $n := n_1 + n_2$, $(I_1, I_2, p) \in \mathbf{R}^n \times \mathbf{R}$ are the action variables and $(\varphi, q) = (\varphi_1, \varphi_2, q) \in \mathbf{T}^n \times \mathbf{T}$ are the angle variables. In Arnold's model $I_1, I_2 \in \mathbf{R}$, $\omega = 1$, $f(I,\varphi,p,q) = (\cos q - 1)(\sin\varphi_1 + \cos\varphi_2)$ and diffusion is proved for μ exponentially small w.r.t. $\sqrt{\varepsilon}$. Physically Hamiltonian H describes a system of n_1 "rotators" and n_2 harmonic oscillators weakly coupled with a pendulum through a perturbation term.

The mechanism proposed in [2] to prove the existence of Arnold diffusion and thereafter become classical, is the following one. For $\mu = 0$, the Hamiltonian system associated to H admits a continuous family of n-dimensional partially hyperbolic invariant tori $\mathcal{T}_I = \{\varphi \in \mathbf{T}^n,\ (I_1, I_2) = I,\ q = p = 0\}$ possessing stable and unstable manifolds $W_0^s(\mathcal{T}_I) = W_0^u(\mathcal{T}_I) = \{\varphi \in \mathbf{T}^n,\ (I_1, I_2) = I,\ (p^2/2) + \varepsilon(\cos q - 1) = 0\}$. The method used in [2] to produce unstable orbits relies on the construction, for $\mu \neq 0$, of "*transition chains*" of perturbed partially hyperbolic tori $\mathcal{T}_I^\mu$ close to $\mathcal{T}_I$ connected one to another by heteroclinic orbits. Therefore in general the first step is to prove the persistence of such hyperbolic tori $\mathcal{T}_I^\mu$ for $\mu \neq 0$ small enough, and to show that the perturbed stable and unstable manifolds $W_\mu^s(\mathcal{T}_I^\mu)$ and $W_\mu^u(\mathcal{T}_I^\mu)$ split and intersect transversally ("splitting problem"). The second step is to find a transition chain of perturbed tori: this is a difficult task since, for general nonisochronous systems, the surviving perturbed tori $\mathcal{T}_I^\mu$ are separated by the gaps appearing in KAM constructions. Two perturbed invariant tori $\mathcal{T}_I^\mu$ and $\mathcal{T}_{I'}^\mu$ could be too distant one from the other, forbidding the existence of a heteroclinic intersection between $W_\mu^u(\mathcal{T}_I^\mu)$ and $W_\mu^s(\mathcal{T}_{I'}^\mu)$: this is the so-called "*gap problem*". In [2] this difficulty is bypassed by the peculiar choice of the perturbation $f(I,\varphi,p,q) = (\cos q - 1) f(\varphi)$, whose gradient vanishes on the unperturbed tori $\mathcal{T}_I$, leaving them *all* invariant also for $\mu \neq 0$. The final step is to prove, by a "shadowing argument", the existence of a true diffusion orbit, close to a given transition chain of tori, for which the action variables I undergo a drift of O(1) in a certain time T_d called the *diffusion time*.

The first paper proving Arnold diffusion in presence of perturbations not preserving the unperturbed tori has been [12]. Extending Arnold's analysis, it is proved in [12] that, if the perturbation is a trigonometric polynomial in the angles φ, then, in some regions of the phase space, the "density" of perturbed invariant tori is high enough to allow the construction of a transition chain.

Regarding the shadowing problem, geometrical methods, see, e.g., [12–14,16], and variational ones, see, e.g., [9], have been applied, in the last years, in order to prove the existence of diffusion orbits shadowing a given transition chain of tori and to estimate the diffusion time. We also quote the important papers [7,8] which, even if dealing with Arnold's model perturbation only, have obtained, by variational methods, very good diffusion time estimates and have introduced new ideas for studying the shadowing problem. For isochronous systems new variational results concerning the shadowing and the splitting problem have been obtained in [4–6].

In this paper we provide an *alternative mechanism* to produce diffusion orbits. This method is not based on the existence of a transition chain of tori: we avoid the KAM construction of the perturbed hyperbolic tori, proving directly the existence of a drifting orbit as a local minimum of an action functional. At the same time our variational approach achieves the optimal diffusion time. We also prove that our diffusion time estimate is the optimal one as a consequence of a general stability result, proved via classical perturbation theory. As in [12] we deal with a perturbation which is a trigonometric polynomial in the angles and our diffusion orbits will not connect any two arbitrary frequencies of the action space, even if we manage to connect more frequencies than in [12], proving the drift also in some regions of the phase space where transition chains might not exist. Clearly if the perturbation is chosen as in Arnold's example we can drift in all the phase space with no restriction. The results proved here have been announced in [3].

In this paper we will assume, as in Arnold's paper, the parameter μ to be small enough in order to validate the so-called Poincaré–Melnikov approximation, when the first-order expansion term in μ for the splitting, the so-called Poincaré–Melnikov function, is the dominant one. For this reason, through this paper we will fix the "Lyapunov exponent" of the pendulum $\varepsilon := 1$, considering the so-called "a priori unstable" case. Actually our variational shadowing technique is not restricted to the a priori unstable case, but would allow, in the same spirit of [4–6], once a "splitting condition" is someway proved, to get diffusion orbits with the best diffusion time (in terms of some measure of the splitting).

We will consider nearly integrable *nonisochronous* Hamiltonian systems defined by:

$$\mathcal{H}_\mu = \frac{I^2}{2} + \frac{p^2}{2} + (\cos q - 1) + \mu f(I, \varphi, p, q, t), \tag{1.1}$$

where $(\varphi, q, t) \in \mathbf{T}^d \times \mathbf{T}^1 \times \mathbf{T}^1$ are the angle variables, $(I, p) \in \mathbf{R}^d \times \mathbf{R}^1$ are the action variables and $\mu \geqslant 0$ is a small real parameter. The Hamiltonian system associated with $\mathcal{H}_\mu$ writes

$$\dot\varphi = I + \mu \partial_I f, \quad \dot I = -\mu \partial_\varphi f, \quad \dot q = p + \mu \partial_p f, \quad \dot p = \sin q - \mu \partial_q f. \tag{$\mathcal{S}_\mu$}$$

The perturbation f is assumed to be a real trigonometric polynomial of order N in φ and t, namely:[1]

[1] $\bar f_{n,l}(I, p, q) = f_{-n,-l}(I, p, q)$ for all $(n, l) \in \mathbf{Z}^d \times \mathbf{Z}$ with $|(n, l)| \leqslant N$ where $\bar z$ denotes the complex conjugate of $z \in \mathbf{C}$.

$$f(I,\varphi,p,q,t)=\sum_{|(n,l)|\leqslant N} f_{n,l}(I,p,q)\mathrm{e}^{\mathrm{i}(n\cdot\varphi+lt)}. \tag{1.2}$$

The unperturbed Hamiltonian system $(\mathcal{S}_0)$ is completely integrable and in particular the energy $I_i^2/2$ of each rotator is a constant of the motion. The problem of *Arnold diffusion* in this context is whether, for $\mu\neq 0$, there exist motions whose net effect is to transfer O(1)-energy among the rotators. A natural complementary question regards the time of stability (or instability) for the perturbed system: what is the minimal time to produce an O(1)-exchange of energy, if any takes place, among the rotators?

For simplicity, even if it is not really necessary, we assume f to be a purely spatial perturbation, namely $f(\varphi,q,t)=\sum_{0\leqslant|(n,l)|\leqslant N} f_{n,l}(q)\exp(\mathrm{i}(n\cdot\varphi+lt))$. The functions $f_{n,l}$ are assumed to be smooth.

Let us define the “resonant web” $\mathcal{D}_N$, formed by the frequencies ω “resonant with the perturbation”:

$$\begin{aligned}\mathcal{D}_N&:=\left\{\omega\in\mathbf{R}^d \;\middle|\; \exists(n,l)\in\mathbf{Z}^{d+1}\text{ s.t. }0<|(n,l)|\leqslant N\text{ and }\omega\cdot n+l=0\right\}\\ &=\bigcup_{0<|(n,l)|\leqslant N} E_{n,l},\end{aligned} \tag{1.3}$$

where $E_{n,l}:=\{\omega\in\mathbf{R}^d \mid \omega\cdot n+l=0\}$. Let us also consider the Poincaré–Melnikov primitive:

$$\Gamma(\omega,\theta_0,\varphi_0):=-\int_{\mathbf{R}}\left[f(\omega t+\varphi_0,q_0(t),t+\theta_0)-f(\omega t+\varphi_0,0,t+\theta_0)\right]\mathrm{d}t,$$

where $q_0(t)=4\arctan(\exp t)$ is the separatrix of the unperturbed pendulum equation $\ddot q=\sin q$ satisfying $q_0(0)=\pi$.

The next theorem states that, for any connected component $\mathcal{C}\subset\mathcal{D}_N^c$, $\omega_I,\omega_F\in\mathcal{C}$, there exists a solution of $(\mathcal{S}_\mu)$ connecting a O(μ)-neighborhood of ω_I in the action space to a O(μ)-neighborhood of ω_F, in the time-interval $T_d=\mathrm{O}((1/\mu)|\ln\mu|)$.

Theorem 1.1. *Let $\mathcal{C}$ be a connected component of $\mathcal{D}_N^c$, $\omega_I,\omega_F\in\mathcal{C}$ and let $\gamma:[0,L]\to\mathcal{C}$ be a smooth embedding such that $\gamma(0)=\omega_I$ and $\gamma(L)=\omega_F$. Assume that, for all $\omega:=\gamma(s)$ $(s\in[0,L])$, $\Gamma(\omega,\cdot,\cdot)$ possesses a nondegenerate local minimum $(\theta_0^\omega,\varphi_0^\omega)$. Then $\forall\eta>0$ there exists $\mu_0=\mu_0(\gamma,\eta)>0$ and $C=C(\gamma)>0$ such that $\forall 0<\mu\leqslant\mu_0$ there exists a solution $(I_\mu(t),\varphi_\mu(t),p_\mu(t),q_\mu(t))$ of $(\mathcal{S}_\mu)$ and two instants $\tau_1<\tau_2$ such that $I_\mu(\tau_1)=\omega_I+\mathrm{O}(\mu)$, $I_\mu(\tau_2)=\omega_F+\mathrm{O}(\mu)$ and*

$$|\tau_2-\tau_1|\leqslant\frac{C}{\mu}|\ln\mu|. \tag{1.4}$$

Moreover $\mathrm{dist}(I_\mu(t),\gamma([0,L]))<\eta$ *for all* $\tau_1\leqslant t\leqslant\tau_2$.

In addition, the above result still holds for any perturbation $\mu(f+\mu\tilde f)$ with any smooth $\tilde f(\varphi,q,t)$.

We can also build diffusion orbits approaching the boundaries of $\mathcal{D}_N$ at distances as small as a certain power of μ: see for a precise statement Theorem 6.1.

Theorem 1.1 improves the corresponding result in [12] which enables to connect two frequencies ω_I and ω_F belonging to the same connected component $\mathcal{C} \subset \mathcal{D}^c_{N_1}$ for $N_1 = 14dN$ and with $\mathrm{dist}\{\{\omega_I, \omega_F\}, \mathcal{D}_{N_1}\} = \mathrm{O}(1)$. Such restrictions of [12] in connecting the action space through diffusion orbits arise because transition chains could not exist in all $\mathcal{C} \subset \mathcal{D}^c_N$ (see Remark 2.2). Unlikely our method enables to show up Arnold diffusion between any two frequencies $\omega_I, \omega_F \in \mathcal{C} \subset \mathcal{D}^c_N$ and along any path, since it does not require the existence of chains of true hyperbolic tori of $(\mathcal{S}_\mu)$.

Theorem 1.1 also improves the known estimates on the diffusion time. The first estimate obtained by geometrical method in [12], is $T_d = \mathrm{O}(\exp(1/\mu^2))$. In [13,14,16], still by geometrical methods, and in [9], by means of Mather's theory, the diffusion time has been proved to be just polynomially long in the splitting μ (the splitting angles between the perturbed stable and unstable manifolds $\mathcal{W}^{s,u}_\mu(\mathcal{T}^\mu_\omega)$ at a homoclinic point are, by classical Poincaré–Melnikov theory, $\mathrm{O}(\mu)$). We note that the variational method proposed by Bessi in [7] had already given, in the case of perturbations preserving all the unperturbed tori, the diffusion time estimate $T_d = \mathrm{O}(1/\mu^2)$. For isochronous systems the estimate on the diffusion time $T_d = \mathrm{O}((1/\mu)|\ln\mu|)$ has already been obtained in [4,5]. Very recently, in [14], the diffusion time (in the nonisochronous case) has been estimated as $T_d = \mathrm{O}((1/\mu)|\ln\mu|)$ by a method which uses "hyperbolic periodic orbits"; however the result of [14] is of local nature: the previous estimate holds only for diffusion orbits shadowing a transition chain close to some torus run with Diophantine flow.

We add that in [15] it was already conjectured that the optimal diffusion time in the a priori unstable case should be $T_d = \mathrm{O}((1/\mu)|\ln\mu|)$.

Our next statement (a stability result) concludes this quest for the minimal diffusion time T_d: it shows the optimality of our estimate $T_d = \mathrm{O}((1/\mu)|\ln\mu|)$.

Theorem 1.2. *Let $f(I,\varphi,p,q,t)$ be as in* (1.2)*, where the $f_{n,l}$ ($|(n,l)| \leqslant N$) are analytic functions. Then $\forall\kappa,\bar r,\tilde r > 0$ there exist $\mu_1,\kappa_0 > 0$ such that $\forall 0 < \mu \leqslant \mu_1$, for any solution $(I(t),\varphi(t),p(t),q(t))$ of $(\mathcal{S}_\mu)$ with $|I(0)| \leqslant \bar r$ and $|p(0)| \leqslant \tilde r$, there results*

$$\left|I(t) - I(0)\right| \leqslant \kappa \quad \forall t \text{ such that } |t| \leqslant \frac{\kappa_0}{\mu}\ln\frac{1}{\mu}. \tag{1.5}$$

Actually the proof of Theorem 1.2 contains much more information: in particular the stability time (1.5) is sharp only for orbits lying close to the separatrices. On the other hand, the orbits lying far away from the separatrices are much more stable, namely exponentially stable in time according to Nekhoroshev type time estimates, see (7.4) and (7.11). Indeed the diffusion orbit of Theorem 1.1 is found close to some pseudo-diffusion orbit whose (q,p) variables move along the separatrices of the pendulum.

As a byproduct of the techniques developed in this paper we have the following result (proved in Section 6) concerning "Arnold's example" [2] where

$$\mathcal{T}_\omega := \{I = \omega,\ \varphi \in \mathbf{T}^d,\ p = q = 0\}$$

are, for all $\omega \in \mathbf{R}^d$, even for $\mu \neq 0$, invariant tori of $(\mathcal{S}_\mu)$.

Theorem 1.3. *Let $f(\varphi, q, t) := (1 - \cos q)\tilde{f}(\varphi, t)$. Assume that for some smooth embedding $\gamma : [0, L] \to \mathbf{R}^d$, with $\gamma(0) = \omega_I$ and $\gamma(L) = \omega_F$, $\forall \omega := \gamma(s)$ $(s \in [0, L])$, $\Gamma(\omega, \cdot, \cdot)$ possesses a nondegenerate local minimum $(\theta_0^\omega, \varphi_0^\omega)$. Then $\forall \eta > 0$ there exists $\mu_0 = \mu_0(\gamma, \eta) > 0$, and $C = C(\gamma) > 0$ such that $\forall 0 < \mu \leqslant \mu_0$ there exists a heteroclinic orbit (η-close to γ) connecting the invariant tori $\mathcal{T}_{\omega_I}$ and $\mathcal{T}_{\omega_F}$. Moreover the diffusion time T_d needed to go from a μ-neighborhood of $\mathcal{T}_{\omega_I}$ to a μ-neighborhood of $\mathcal{T}_{\omega_F}$ is bounded by $(C/\mu)|\ln \mu|$ for some constant C.*

The method of proof of Theorem 1.1 (and Theorem 1.3) relies on a finite-dimensional reduction of Lyapunov–Schmidt type, variational in nature, introduced in [1] and later extended in [4–6] to the problem of Arnold diffusion. The diffusion orbit of Theorem 1.1 is found as a local minimum of the action functional close to some pseudo-diffusion orbit whose (p, q) variables move along the separatrices of the pendulum. The pseudo-diffusion orbits, constructed by the Implicit Function Theorem, are true solutions of $(\mathcal{S}_\mu)$ except possibly at some instants θ_i, for $i = 1, \ldots, k$, when they are glued continuously at the section $\{q = \pi, \bmod 2\pi\mathbf{Z}\}$ but the speeds $(\dot{\varphi}_\mu(\theta_i), \dot{q}_\mu(\theta_i)) = (I_\mu(\theta_i), p_\mu(\theta_i))$ may have a jump. The time interval $T_s = \theta_{i+1} - \theta_i$ is heuristically the time required to perform a single transition during which the rotators can exchange $\mathrm{O}(\mu)$-energy, i.e., the action variables vary of $\mathrm{O}(\mu)$. During each transition we can exchange only $\mathrm{O}(\mu)$-energy because the Melnikov contribution in the perturbed functional is $\mathrm{O}(\mu)$. Hence in order to exchange $\mathrm{O}(1)$ energy the number of transitions required will be $k = \mathrm{O}(1/\mu)$.

We underline that the question of finding the optimal time and the mechanism for which we can avoid the construction of transition chains of tori are deeply connected. Indeed the main reason for which our drifting technique avoids the construction of KAM tori is the following one: if the time to perform a simple transition T_s is, say, just $T_s = \mathrm{O}(|\ln \mu|)$ then, on such "short" time intervals, it is valid to approximate the pseudo diffusion orbits with unperturbed solutions living on the stable and unstable manifolds of the unperturbed tori $W^s(\mathcal{T}_\omega) = W^u(\mathcal{T}_\omega) = \{I = \omega,\ \varphi \in \mathbf{T}^d,\ p^2/2 + (\cos q - 1) = 0\}$, when computing the value of the action functional. In this way we do not need to construct the true hyperbolic tori $\mathcal{T}_\omega^\mu$ (actually for our approximation we only need the time for a single transition to be $T_s \ll 1/\mu$).

The fact that it is possible to perform a single transition in a very short time interval like $T_s = \mathrm{O}(|\ln \mu|)$ is not obvious at all. In [7] the time to perform a single transition, in the example of Arnold, is $\mathrm{O}(1/\mu)$. This transition time arises in order to ensure that the variations of the kinetic part of the action functional associated with the rotators are small compared with the (positive definite) second derivative of the Poincaré–Melnikov primitive at its minimum point. Unfortunately this time is too long to use a simple approximation of the functional. The key observation that enables us to perform a single transition in a very short time interval concerns the behavior of the "gradient flow" of the unperturbed action functional of the rotators. This implies a sort of a priori estimate satisfied by the minimal diffusion orbits, see Remark 6.1. We think that estimate (6.18) is interesting in itself. In this way we can show that the variations of the action of the rotators are small enough, even on time intervals $T_s \ll 1/\mu$, and do not "destroy" the minimum of the Poincaré–Melnikov primitive.

When trying to build a pseudo-diffusion orbit which performs single transitions in very short time intervals we encounter another difficulty linked with the ergodization time. The time to perform a single transition T_s must be long enough to settle, at each instant θ_i, the projection (θ_i, φ_i) of the pseudo-orbit on the torus $\mathbf{T}^{d+1}$ sufficiently close to the minimum of the Poincaré–Melnikov function, i.e., the homoclinic point (in our method it is sufficient to arrive just O(1)-close, independently of μ, to the homoclinic point). This necessary request creates some difficulty since our pseudo-diffusion orbit may arrive $\mathrm{O}(\mu)$-close in the action space to resonant hyperplanes of frequencies whose linear flow does not provide a dense enough net of the torus. The way in which this problem is overcome is discussed in Section 5: we observe a phenomenon of “stabilization close to resonances” which forces the time for some single transitions to increase. Anyway the total time required to cross these (finite number of) resonances is still $T_d = \mathrm{O}((1/\mu)\ln(1/\mu))$, see (5.13) and the proof of Theorem 1.1. This discussion enables us to prove optimal fast-Arnold diffusion in large regions of the phase space and allows to improve the local diffusion results of [14].

We need therefore some results on the ergodization time of the torus for linear flows possibly resonant but only at a “sufficiently high order”. We present these results in Section 4. We point out that the main result of this section, Theorem 4.2, implies as corollaries Theorems B and D of [11], see Remark 4.1. It is of independent interest and could possibly improve the other results of [11].

This work is a further step of a research line, started in [4–6], for finding new mechanisms to prove Arnold diffusion. We expect that the variational method developed in this paper could be suitably refined in order to prove the existence of drifting orbits in the whole action space and then to prove such results for generic analytic perturbations too. Another possible application of these methods could regard infinite-dimensional Hamiltonian systems where the existence of “transition chains of infinite-dimensional hyperbolic tori” is quite far from being proved.

The paper is organized as follows: in Section 2 we perform the finite-dimensional reduction and we define the variational setting. In Section 3 we provide a suitable development of the reduced action functional. In Section 4 we prove the new results on the ergodization time. In Section 5 we define the unperturbed pseudo-orbit. In Section 6 we prove the existence of the diffusion orbit. In Section 7 we prove the stability result, that is to say the optimality of our diffusion time.

Notations. Through this paper the notation $a(z_1, \ldots, z_k) = \mathrm{O}(b(\mu))$ will mean that, for a suitable positive constant $C(\gamma, f) > 0$, $|a(z_1, \ldots, z_p)| \leqslant C(\gamma, f)|b(\mu)|$.

2. The variational setting and the finite-dimensional reduction

When the perturbation $f(\varphi, q, t) = \sum_{|(n,l)| \leqslant N} f_{n,l}(q) \exp(\mathrm{i}(n \cdot \varphi + lt))$ is purely spatial,[2] system $(\mathcal{S}_\mu)$ reduces to the second-order system

[2] We will develop all the computations for f. All the next arguments remain unchanged if the perturbation is $f + \mu \tilde{f}$, see the proof of Theorem 1.1.

$$\ddot\varphi = -\mu\, \partial_\varphi f(\varphi, q, t), \quad -\ddot q + \sin q = \mu\, \partial_q f(\varphi, q, t) \tag{2.1}$$

with associated Lagrangian

$$\mathcal{L}_\mu(\varphi, \dot\varphi, q, \dot q, t) = \frac{\dot\varphi^2}{2} + \frac{\dot q^2}{2} + (1 - \cos q) - \mu f(\varphi, q, t). \tag{2.2}$$

Using the Contraction Mapping Theorem we will prove in Lemma 2.1 that, near the unperturbed solutions $(\omega(t-\theta) + \varphi_0, q_0(t-\theta))$ living on the stable and unstable manifolds of the unperturbed tori $\mathcal{T}_\omega$, there exist, for μ small enough, solutions of the perturbed system (2.1) which connect the sections $\{\varphi = \varphi^+, q = -\pi, t = \theta^+\}$ and $\{\varphi = \varphi^-, q = \pi, t = \theta^-\}$ (under some assumptions). The diffusion orbit will be a chain of such connecting orbits.

We first introduce a few definitions and notations. For $\lambda := (\theta^+, \theta^-, \varphi^+, \varphi^-) \in \mathbf{R}^2 \times \mathbf{R}^{2d}$ with $\theta^+ < \theta^-$ we define $T_\lambda := \theta^- - \theta^+$ and the "mean frequency" $\omega_\lambda \in \mathbf{R}^d$ as $\omega_\lambda := (\varphi^- - \varphi^+)/(\theta^- - \theta^+)$. The "small denominator" of a frequency $\omega \in \mathbf{R}^d$ is defined by:

$$\beta(\omega) := \beta_N(\omega) := \min_{0<|(n,l)|\leqslant N} |n \cdot \omega + l|. \tag{2.3}$$

$\beta(\omega)$ measures how close the frequency ω lies to the resonant web $\mathcal{D}_N$ defined in (1.3). We use the abbreviation β_λ for $\beta(\omega_\lambda)$. We shall always assume through this paper that ω stays in a fixed bounded set containing the curve γ.

For T large enough, there exists a unique T-periodic solution Q_T of the pendulum equation, of small positive energy with $Q_T(0) = -\pi$, $Q_T(T) = \pi$. Moreover Q_T satisfies $\forall t \in [0, T/2) \cup (T/2, T]$,

$$\left|\partial_T Q_T(t)\right| \leqslant K_1 \mathrm{e}^{-K_2(T-t)}, \quad \left|\partial_T\big(Q_T(T - \cdot)\big)(t)\right| \leqslant K_1 \mathrm{e}^{-K_2(T-t)}$$

and

$$\begin{aligned} &\left|Q_T(t) - q_\infty(t)\right| + \left|\dot Q_T(t) - \dot q_\infty(t)\right| \leqslant K_1 \mathrm{e}^{-K_2 T}, \\ &\left|\dot Q_T(t)\right| \leqslant K_1 \max\left\{\mathrm{e}^{-K_2 t}, \mathrm{e}^{-K_2(T-t)}\right\}, \end{aligned} \tag{2.4}$$

for some positive constants K_1 and K_2, where q_∞ is defined by:

$$q_\infty(t) = q_0(t) - 2\pi \quad \text{if } t \in [0, T/2), \qquad q_\infty(t) = q_0(t - T) \quad \text{if } t \in (T/2, T].$$

Lemma 2.1. *There exists $\mu_2 > 0$ and constants $C_0, C_1, \bar c, c_1 > 0$ such that $\forall 0 < \mu \leqslant \mu_2$, $\forall \lambda = (\theta^+, \theta^-, \varphi^+, \varphi^-)$ such that $C_0 \beta_\lambda^2 > \mu$ and $C_1 |\ln \mu| \leqslant T_\lambda \leqslant C_0 \beta_\lambda / \mu$ there exists a unique solution $(\varphi_\mu(t), q_\mu(t)) := (\varphi_{\mu,\lambda}(t), q_{\mu,\lambda}(t))$ of (2.1), defined for $t \in (\theta^+ - 1, \theta^- + 1)$, satisfying $\varphi_\mu(\theta^\pm) = \varphi^\pm$, $q_\mu(\theta^\pm) = \mp\pi$ and*

$$\begin{aligned} &\text{(i)} \quad \left|\varphi_\mu(t) - \bar\varphi(t)\right| \leqslant \bar c \mu\big(1 + c_1 \mu T_\lambda^2\big)/\beta_\lambda^2, \qquad \left|\dot\varphi_\mu(t) - \omega\right| \leqslant \bar c \mu / \beta_\lambda, \\ &\text{(ii)} \quad \left|q_\mu(t) - Q_{T_\lambda}(t - \theta^+)\right| \leqslant \bar c \mu, \qquad \left|\dot q_\mu(t) - \dot Q_{T_\lambda}(t - \theta^+)\right| \leqslant \bar c \mu, \end{aligned} \tag{2.5}$$

where $\bar\varphi(t) := \omega_\lambda(t-\theta^+)+\varphi^+$. *Moreover* $\varphi_{\mu,\lambda}(t)$, $\dot\varphi_{\mu,\lambda}(t)$, $q_{\mu,\lambda}(t)$ *and* $\dot q_{\mu,\lambda}(t)$ *are* $\mathcal{C}^1$ *functions of* (t,λ).

The proof of Lemma 2.1 is given in Appendix A.

Remark 2.1. Roughly, the meaning of the above estimates is the following:

(1) We have imposed $C_1|\ln\mu| < T_\lambda := \theta^- - \theta^+$ so that by (2.4), on such intervals of time, the periodic solution Q_{T_λ} is $O(\mu)$ close to "separatrices" q_∞ of the unperturbed pendulum.
(2) Estimate (ii) implies that for $t \approx (\theta^+ + \theta^-)/2$ the perturbed solution q_μ may have $O(\mu)$ oscillations around the unstable equilibrium of the pendulum $q=0$, mod 2π, which is exactly what one expects perturbing with a general f. On the contrary for the class of perturbations considered in [2] as $f(\varphi,q,t) = (1-\cos q)f(\varphi,t)$ preserving *all* the invariant tori, estimate (ii) can be improved, getting $\max\{|q_\mu(t) - Q_{T_\lambda}(t-\theta^+)|, |\dot q_\mu(t) - \dot Q_{T_\lambda}(t-\theta^+)|\} = O(\mu\max\{\exp(-C|t-\theta^+|), \exp(-C|t-\theta^-|)\})$.
(3) For $\beta_\lambda \approx \sqrt{\mu}$ estimate (i) becomes meaningless: for a mean frequency ω_λ such that $n\cdot\omega_\lambda + l \approx \sqrt{\mu}$ for some $0 < |(n,l)| \leqslant N$ the perturbed transition orbits φ_μ are no more well-approximated by the straight lines $\bar\varphi(t) := \varphi^+ + \omega_\lambda(t-\theta^+)$.

Remark 2.2. Let us define $\mathcal{D}_N^\beta := \{\omega \in \mathbf{R}^d \mid |\omega\cdot n + l| > \beta,\ \forall 0 < |(n,l)| \leqslant N\}$. In [12] it is proved that hyperbolic invariant tori $\mathcal{T}_\omega^\mu$ of system $(\mathcal{S}_\mu)$ exist for Diophantine frequencies $\omega \in \mathcal{D}_{N_1}^{\beta_1}$, for some $\beta_1 = O(1)$ and some $N_1 = O(dN) > N$, namely avoiding more "resonances with the trigonometric polynomial f" than just N. The presence of such "resonant hyperplanes $E_{n,l}$" for $N < |(n,l)| < N_1$ may be reflected in estimate (i) by the term μT_λ^2. However such term, for our purposes, can be ignored.

From this point of view Lemma 2.1 could perhaps be interpreted as the first iterative step for looking at invariant hyperbolic tori in the perturbed system bifurcating from the unperturbed ones.

By Lemma 2.1, for $0 < \mu \leqslant \mu_2$, we can define on the set

$$\Lambda_\mu := \left\{\lambda = (\theta^+,\theta^-,\varphi^+,\varphi^-) \;\middle|\; C_0\beta_\lambda^2 > \mu,\ C_1|\ln\mu| \leqslant T_\lambda \leqslant \frac{C_0\beta_\lambda}{\mu}\right\},$$

the Lagrangian action functional $G_\mu : \Lambda_\mu \to \mathbf{R}$ as

$$G_\mu(\lambda) = G_\mu(\theta^+,\theta^-,\varphi^+,\varphi^-) := \int_{\theta^+}^{\theta^-} \mathcal{L}_\mu\big(\varphi_\mu(t),\dot\varphi_\mu(t),q_\mu(t),\dot q_\mu(t),t\big)\,dt. \tag{2.6}$$

We have:

Lemma 2.2. G_μ *is differentiable and* (*with the abbreviations* φ, q *for* φ_μ, q_μ)

$$\nabla_{\varphi^+} G_\mu(\lambda) = -\dot\varphi(\theta^+),$$
$$\partial_{\theta^+} G_\mu(\lambda) = \frac{1}{2}\bigl|\dot\varphi(\theta^+)\bigr|^2 + \frac{1}{2}\dot q^2(\theta^+) + \cos q(\theta^+) - 1 + \mu f(\varphi^+, \pi, \theta^+),$$
$$\nabla_{\varphi^-} G_\mu(\lambda) = \dot\varphi(\theta^-),$$
$$\partial_{\theta^-} G_\mu(\lambda) = -\left(\frac{1}{2}\bigl|\dot\varphi(\theta^-)\bigr|^2 + \frac{1}{2}\dot q^2(\theta^-) + \cos q(\theta^-) - 1 + \mu f(\varphi^-, \pi, \theta^-)\right).$$

Proof. By Lemma 2.1 the map $(\lambda, t) \mapsto (\varphi_{\mu,\lambda}(t), \dot\varphi_{\mu,\lambda}(t), q_{\mu,\lambda}(t), \dot q_{\mu,\lambda}(t))$ is C^1 on the set $\{(\lambda, t) \in \Lambda_\mu \times \mathbf{R} \mid \theta^+ \leqslant t \leqslant \theta^-\}$. Hence G_μ is differentiable and

$$\partial_{\theta^+} G_\mu(\lambda) = -\mathcal{L}_\mu\bigl(\varphi^+, \dot\varphi(\theta^+), -\pi, \dot q(\theta^+), \theta^+\bigr) + \int_{\theta^+}^{\theta^-} \dot\varphi(s)\cdot\partial_{\theta^+}\dot\varphi(s) + \dot q(s)\partial_{\theta^+}\dot q(s)\,\mathrm{d}s$$
$$+ \int_{\theta^+}^{\theta^-} \sin q(s)\partial_{\theta^+} q(s) - \mu\partial_\varphi f\bigl(\varphi(s), q(s), s\bigr)\cdot\partial_{\theta^+}\varphi(s)$$
$$- \mu\partial_q f\bigl(\varphi(s), q(s), s\bigr)\partial_{\theta^+} q(s)\,\mathrm{d}s.$$

Integrating by parts and using that $(q_{\mu,\lambda}, \varphi_{\mu,\lambda})$ satisfies (2.1) in (θ^+, θ^-), we obtain:

$$\partial_{\theta^+} G_\mu(\lambda) = -\mathcal{L}_\mu\bigl(\varphi^+, \dot\varphi(\theta^+), -\pi, \dot q(\theta^+), \theta^+\bigr) + \bigl[\dot q(s)\partial_{\theta^+} q(s) + \dot\varphi(s)\cdot\partial_{\theta^+}\varphi(s)\bigr]_{\theta^+}^{\theta^-}.$$

Now $q_{\mu,\lambda}(\theta^+) = -\pi$ for all λ hence $\dot q(\theta^+) + \partial_{\theta^+} q(\theta^+) = 0$. Similarly we get $\dot\varphi(\theta^+) + \partial_{\theta^+}\varphi(\theta^+) = 0$, $\partial_{\theta^+} q(\theta^-) = 0$, $\partial_{\theta^+}\varphi(\theta^-) = 0$. As a consequence

$$\partial_{\theta^+} G_\mu(\lambda) = \frac{1}{2}|\dot\varphi|^2(\theta^+) + \frac{1}{2}\dot q^2(\theta^+) + \bigl(\cos q(\theta^+) - 1\bigr) + \mu f(\varphi^+, \pi, \theta^+).$$

The other partial derivatives are computed in the same way. □

For $\beta > 0$ fixed, denoting $\lambda_i = (\theta_i, \theta_{i+1}, \varphi_i, \varphi_{i+1})$, we define on the set:

$$\Lambda_{\mu,k} := \Lambda_{\mu,k}^\beta$$
$$:= \bigl\{\lambda = (\theta_1, \dots, \theta_k, \varphi_1, \dots, \varphi_k) \in \mathbf{R}^k \times \mathbf{R}^{kd} \bigm| \forall 1 \leqslant i \leqslant k-1,\ \lambda_i \in \Lambda_\mu,\ \beta_{\lambda_i} \geqslant \beta\bigr\},$$

the reduced action functional $\mathcal{F}_\mu : \Lambda_{\mu,k} \to \mathbf{R}$ as

$$\mathcal{F}_\mu(\lambda) = \omega_I\varphi_1 - \frac{|\omega_I|^2}{2}\theta_1 + \mu\Gamma^u(\omega_I, \theta_1, \varphi_1) + \mu F(\omega_I, \theta_1, \varphi_1) + \sum_{i=1}^{k-1} G_\mu(\lambda_i)$$
$$- \omega_F\varphi_k + \frac{|\omega_F|^2}{2}\theta_k + \mu\Gamma^s(\omega_F, \theta_k, \varphi_k) - \mu F(\omega_F, \theta_k, \varphi_k),$$

where

$$\Gamma^u(\omega,\theta_0,\varphi_0) := -\int_{-\infty}^{0} \bigl[f\bigl(\omega t+\varphi_0, q_0(t), t+\theta_0\bigr) - f\bigl(\omega t+\varphi_0, 0, t+\theta_0\bigr)\bigr]\,\mathrm{d}t, \quad (2.7)$$

$$\Gamma^s(\omega,\theta_0,\varphi_0) := -\int_{0}^{+\infty} \bigl[f\bigl(\omega t+\varphi_0, q_0(t), t+\theta_0\bigr) - f\bigl(\omega t+\varphi_0, 0, t+\theta_0\bigr)\bigr]\,\mathrm{d}t, \quad (2.8)$$

are called respectively the unstable and the stable Poincaré–Melnikov primitive, and

$$F(\omega,\theta_0,\varphi_0) := -f_{0,0}\theta_0 - \sum_{0<|(n,l)|\leqslant N} f_{n,l}\frac{\mathrm{e}^{\mathrm{i}(n\cdot\varphi_0+l\theta_0)}}{\mathrm{i}(n\cdot\omega+l)}, \quad (2.9)$$

$f_{n,l} := f_{n,l}(0)$ being the Fourier coefficients of $f(\varphi,0,t)$.

Critical points of the "reduced action functional" $\mathcal{F}_\mu$ give rise to diffusion orbits whose action variables I go from a small neighborhood of ω_I to a small neighborhood of ω_F, as stated in Lemma 2.3 below. The "boundary terms"

$$\omega_I\varphi_1 - \frac{|\omega_I|^2}{2}\theta_1 + \mu\Gamma^u(\omega_I,\theta_1,\varphi_1) + \mu F(\omega_I,\theta_1,\varphi_1)$$

and

$$-\omega_F\varphi_k + \frac{|\omega_F|^2}{2}\theta_k + \mu\Gamma^s(\omega_F,\theta_k,\varphi_k) - \mu F(\omega_F,\theta_k,\varphi_k)$$

have been added also to enable us to find critical points of $\mathcal{F}_\mu$ w.r.t. all the variables (including $\theta_1,\varphi_1,\theta_k,\varphi_k$).

More precisely, for $\lambda = (\theta,\varphi) \in \Lambda_{\mu,k}$ we define the pseudo diffusion solutions $(\varphi_{\mu,\lambda}, q_{\mu,\lambda})$ on the interval $[\theta_1,\theta_k]$ by

$$\bigl(\varphi_{\mu,\lambda}(t), q_{\mu,\lambda}(t)\bigr) := \bigl(\varphi_{\mu,\lambda_i}(t), q_{\mu,\lambda_i}(t) + 2\pi(i-1)\bigr) \quad \text{for } t\in[\theta_i,\theta_{i+1}],$$

where $(\varphi_{\mu,\lambda_i}(t), q_{\mu,\lambda_i}(t))$ are given by Lemma 2.1. The pseudo diffusion solutions $(\varphi_{\mu,\lambda}, q_{\mu,\lambda})$ are then continuous functions which are true solutions of the equations of motion (2.1) on each interval (θ_i,θ_{i+1}), but the time derivatives $(\dot\varphi_{\mu,\lambda}, \dot q_{\mu,\lambda})$ may undergo a jump at time θ_i. We have

Lemma 2.3. *If $\tilde\lambda = (\tilde\theta,\widetilde\varphi) \in \Lambda_{\mu,k}$ is a critical point of $\mathcal{F}_\mu$, then $(\varphi_{\mu,\tilde\lambda}(t), q_{\mu,\tilde\lambda}(t))$ is a solution of* (2.1) *in the time interval $(\tilde\theta_1,\tilde\theta_k)$. Moreover $\dot\varphi_\mu(\tilde\theta_1) = \omega_I + \mathrm{O}(\mu)$, $\dot\varphi_\mu(\tilde\theta_k) = \omega_F + \mathrm{O}(\mu)$, i.e., $(\varphi_{\mu,\tilde\lambda}, q_{\mu,\tilde\lambda})$ is a diffusion orbit between ω_I and ω_F with diffusion time $T_d = |\tilde\theta_k - \tilde\theta_1|$.*

Proof. By Lemma 2.2 if $\nabla_{\varphi_i}\mathcal{F}_\mu(\tilde\lambda)=0$, then for $2\leqslant i\leqslant k-1$, $\dot\varphi_{\mu,\tilde\lambda}(\tilde\theta_i^-)=\dot\varphi_{\mu,\tilde\lambda}(\tilde\theta_i^+)$ and $\dot\varphi_{\mu,\tilde\lambda}(\tilde\theta_1)=\omega_I+\mathrm{O}(\mu)$, $\dot\varphi_{\mu,\tilde\lambda}(\tilde\theta_k)=\omega_F+\mathrm{O}(\mu)$. Moreover, if $\nabla_{\varphi_i}\mathcal{F}_\mu(\tilde\lambda)=0$ and $\partial_{\theta_i}\mathcal{F}_\mu(\tilde\lambda)=0$ then (for $2\leqslant i\leqslant k-2$), $\dot q^2_{\mu,\tilde\lambda}(\tilde\theta_i^+)=\dot q^2_{\mu,\tilde\lambda}(\tilde\theta_i^-)$. Now, by Lemma 2.1 and (2.4), $\dot q_{\mu,\tilde\lambda}(\tilde\theta_i^\pm)=\dot q_0(0)+\mathrm{O}(\mu)$. Hence $\dot q_{\mu,\tilde\lambda}(\tilde\theta_i^+)=\dot q_{\mu,\tilde\lambda}(\tilde\theta_i^-)$ and the proof is complete. □

3. The approximation of the reduced functional

In order to prove the existence of critical points of the reduced action functional $\mathcal{F}_\mu$ thanks to the properties of the Poincaré–Melnikov primitives $\Gamma(\omega,\cdot,\cdot)$ we need an appropriate expression of $\mathcal{F}_\mu$, see Lemma 3.5. We shall express $\mathcal{F}_\mu$ as the sum of a function whose definition contains the $\Gamma(\omega,\cdot,\cdot)$ (for which we can prove the existence of critical points) and of a remainder whose derivatives are so small that it cannot destroy the critical points of the first function.

The first lemma gives an approximation of G_μ (defined in (2.6)).

Lemma 3.1. *For $0<\mu\leqslant\mu_3$, for $\lambda\in\Lambda_\mu$ we have*:

$$G_\mu(\lambda)=\frac12\frac{|\varphi^--\varphi^+|^2}{(\theta^--\theta^+)}+\mu\Gamma^s(\omega_\lambda,\theta^+,\varphi^+)+\mu\Gamma^u(\omega_\lambda,\theta^-,\varphi^-)$$
$$-\mu\int_{\theta^+}^{\theta^-}f(\bar\varphi(t),0,t)\,\mathrm{d}t+R_0(\mu,\lambda),\tag{3.1}$$

where

$$\nabla_\lambda R_0(\mu,\lambda)=\mathrm{O}\left(\frac{\mu^2(1+\mu T_\lambda^2)}{\beta_\lambda^2}T_\lambda\right).\tag{3.2}$$

Proof. By Lemma 2.1, we can write

$$\varphi_{\mu,\lambda}(t)=\bar\varphi(t)+v_{\mu,\lambda}(t),\qquad q_{\mu,\lambda}(t)=Q_{T_\lambda}(t-\theta^+)+w_{\mu,\lambda}(t),$$

where

$$v_{\mu,\lambda}(\theta^+)=v_{\mu,\lambda}(\theta^-)=0,\quad \|\dot v_{\mu,\lambda}\|_{L^\infty(\theta^+,\theta^-)}=\mathrm{O}(\mu/\beta_\lambda),$$
$$\|v_{\mu,\lambda}\|_{L^\infty(\theta^+,\theta^-)}=\mathrm{O}\big(\big(\mu/\beta_\lambda^2\big)\big(1+\mu T_\lambda^2\big)\big)\quad\text{and}\quad w_{\mu,\lambda}(\theta^+)=w_{\mu,\lambda}(\theta^-)=0,$$
$$\|\dot w_{\mu,\lambda}\|_{L^\infty(\theta^+,\theta^-)}+\|w_{\mu,\lambda}\|_{L^\infty(\theta^+,\theta^-)}=\mathrm{O}(\mu).$$

In the following, in order to avoid cumbersome notation, we shall use the abbreviations v,w,Q for $v_{\mu,\lambda},w_{\mu,\lambda},Q_{T_\lambda}(\cdot-\theta^+)$, the dependency w.r.t. λ and μ being implicit. We have:

$$G_\mu(\lambda) = \int_{\theta^+}^{\theta^-} \frac{1}{2}\left|\dot{\overline{\varphi}}(t)\right|^2 + \dot{\overline{\varphi}}(t)\cdot\dot{v}(t) + \frac{1}{2}\left|\dot{v}(t)\right|^2 + \frac{1}{2}\dot{Q}^2(t) + \dot{Q}(t)\dot{w}(t) + \frac{1}{2}\dot{w}^2(t)$$
$$+ \int_{\theta^+}^{\theta^-} \left[1-\cos\bigl(Q(t)+w(t)\bigr)\right] - \mu f\bigl(\overline{\varphi}(t)+v(t), Q(t)+w(t), t\bigr)\,\mathrm{d}t.$$

Now since $v(\theta^+) = v(\theta^-) = 0$ and $w(\theta^+) = w(\theta^-) = 0$,

$$\int_{\theta^+}^{\theta^-} \dot{\overline{\varphi}}(t)\cdot\dot{v}(t)\,\mathrm{d}t = \int_{\theta^+}^{\theta^-} \omega_\lambda\cdot\dot{v}(t)\,\mathrm{d}t = 0$$

and

$$\int_{\theta^+}^{\theta^-} \dot{Q}(t)\dot{w}(t)\,\mathrm{d}t = \int_{\theta^+}^{\theta^-} -\ddot{Q}(t)w(t)\,\mathrm{d}t = \int_{\theta^+}^{\theta^-} -\bigl(\sin Q(t)\bigr)w(t)\,\mathrm{d}t.$$

As a result, $G_\mu(\lambda) = G_\mu^0(\lambda) + R_1(\lambda)$, where

$$G_\mu^0(\lambda) = \int_{\theta^+}^{\theta^-} \frac{1}{2}|\dot{\overline{\varphi}}|^2 + \frac{1}{2}\dot{Q}^2 + (1-\cos Q) - \mu f(\overline{\varphi}, Q, t),$$
$$R_1(\lambda) = \int_{\theta^+}^{\theta^-} \frac{1}{2}|\dot{v}|^2 + \frac{1}{2}\dot{w}^2 + \bigl(\cos Q - \cos(Q+w) - w\sin Q\bigr)$$
$$- \mu f(\overline{\varphi}+v, Q+w, t) + \mu f(\overline{\varphi}, Q, t).$$

We shall first prove that $|\nabla R_1| = \mathrm{O}(\mu^2(1+\mu T_\lambda^2)T_\lambda/\beta_\lambda^2)$. We have $\partial_{\theta^+}R_1 = r_1 + r_2 + r_3 + r_4 + r_5 + r_6$, where

$$r_1 := \int_{\theta^+}^{\theta^-} \dot{v}\cdot\frac{\mathrm{d}}{\mathrm{d}t}(\partial_{\theta^+}v) - \mu\partial_\varphi f(\overline{\varphi}+v, Q+w, t)\cdot(\partial_{\theta^+}v),$$
$$r_2 := \int_{\theta^+}^{\theta^-} \dot{w}\frac{\mathrm{d}}{\mathrm{d}t}(\partial_{\theta^+}w) + \bigl[\sin(Q+w) - \sin Q - \mu\partial_q f(\overline{\varphi}+v, Q+w, t)\bigr](\partial_{\theta^+}w),$$

$$r_3 := \int_{\theta^+}^{\theta^-} \big(-\sin Q + \sin(Q+w) - w\cos Q\big)\partial_{\theta^+} Q,$$

$$r_4 := \mu \int_{\theta^+}^{\theta^-} \big[\partial_\varphi f(\overline{\varphi}, Q, t) - \partial_\varphi f(\overline{\varphi}+v, Q+w, t)\big] \cdot \partial_{\theta^+}\overline{\varphi},$$

$$r_5 := \mu \int_{\theta^+}^{\theta^-} \big[\partial_q f(\overline{\varphi}, Q, t) - \partial_q f(\overline{\varphi}+v, Q+w, t)\big]\partial_{\theta^+} Q,$$

$$r_6 := -\frac{1}{2}\big|\dot v(\theta^+)\big|^2 - \frac{1}{2}\dot w(\theta^+)^2.$$

Now v and w satisfy

$$\begin{cases} -\ddot v(t) = \mu \partial_\varphi f\big(\overline{\varphi}(t)+v(t), Q(t)+w(t), t\big), \\ -\ddot w(t) + \sin\big(Q(t)+w(t)\big) = \mu\partial_q f\big(\overline{\varphi}(t)+v(t), Q(t)+w(t), t\big) + \sin Q(t). \end{cases}$$

Moreover, deriving w.r.t. θ^+ the equality $v(\theta^+) = 0$ we obtain that $(\partial_{\theta^+} v)(\theta^+) = -\dot v(\theta^+)$. Similarly $(\partial_{\theta^+} w)(\theta^+) = -\dot w(\theta^+)$, $(\partial_{\theta^+} v)(\theta^-) = 0$ and $(\partial_{\theta^+} w)(\theta^-) = 0$. Therefore an integration by parts gives $r_1 = |\dot v(\theta^+)|^2$, $r_2 = \dot w(\theta^+)^2$ hence $|r_1| + |r_2| = \mathrm{O}(\mu^2/\beta^2)$.

By the properties of Q_T, $\partial_{\theta^+} Q$ is bounded in the interval $[\theta^+, \theta^-]$ by a constant independent of λ. Moreover $-\sin Q(t) + \sin(Q(t)+w(t)) - w(t)\cos Q(t) = \mathrm{O}(w(t)^2)$. Therefore $r_3 = \mathrm{O}(\mu^2 T)$.

We have also, for some positive constant c,

$$|r_4| + |r_5| \leqslant c\mu T\Big[\sup_{t\in[\theta^+,\theta^-]} \big|\partial_{\theta^+} Q(t)\big| + \big|\partial_{\theta^+}\overline{\varphi}(t)\big|\Big]\Big[\sup_{t\in[\theta^+,\theta^-]} \big(\big|v(t)\big| + \big|w(t)\big|\big)\Big].$$

Since $\partial_{\theta^+}\overline{\varphi}$ is bounded independently of λ, we have by Lemma 2.1 $|r_4| + |r_5| = \mathrm{O}(\mu^2(1+\mu T_\lambda^2)T_\lambda/\beta_\lambda^2)$. Still by Lemma 2.1, $r_6 = \mathrm{O}(\mu^2/\beta^2)$. The estimate of the other derivatives of R_1 is obtained in the same way.

We now develop $G_\mu^0(\lambda)$ as

$$G_\mu^0(\lambda) = \frac{1}{2}\frac{|\varphi^- - \varphi^+|^2}{(\theta^- - \theta^+)} + \mu\Gamma^s(\omega_\lambda, \theta^+, \varphi^+) + \mu\Gamma^u(\omega_\lambda, \theta^-, \varphi^-) - \mu\int_{\theta^+}^{\theta^-} f(\overline{\varphi}(t), 0, t)\,\mathrm{d}t + R_2(\lambda) + R_3(\lambda),$$

where

$$R_2(\lambda) = \int_{\theta^+}^{\theta^-} \frac{1}{2}\dot{Q}^2(t) + \big(1 - \cos Q(t)\big)\,\mathrm{d}t = \int_0^{T_\lambda} \frac{1}{2}\dot{Q}^2_{T_\lambda}(t) + \big(1 - \cos Q_{T_\lambda}(t)\big)\,\mathrm{d}t, \tag{3.3}$$

$$R_3(\lambda) = \int_{\theta^+}^{\theta^-} -\mu\big[\big(f\big(\overline{\varphi}(t), Q(t), t\big) - f\big(\overline{\varphi}(t), 0, t\big)\big]\,\mathrm{d}t - \mu\Gamma^s(\omega_\lambda, \theta^+, \varphi^+)$$
$$- \mu\Gamma^u(\omega_\lambda, \theta^-, \varphi^-).$$

There remains to prove estimate (3.2) for ∇R_2 and ∇R_3. By (3.3) $\partial_{\varphi^\pm} R_2 = 0$ and $\partial_{\theta^+} R_2(\lambda) = -\partial_{\theta^-} R_2(\lambda)$ is the energy of the T_λ-periodic solution Q_{T_λ} of the pendulum equation. Now this energy is $\mathrm{O}(\mathrm{e}^{-c_2 T_\lambda})$. Hence (provided C_1 is large enough) $|\nabla R_2(\lambda)| = \mathrm{O}(\mu^2)$.

In order to estimate the derivatives of R_3, let us define $g(\varphi, q, t) := f(\varphi, q, t) - f(\varphi, 0, t)$. We have

$$R_3(\lambda) = \int_{\theta^+}^{\theta^-} -\mu g\big(\overline{\varphi}(t), Q(t), t\big)\,\mathrm{d}t - \mu\Gamma^s(\omega_\lambda, \theta^+, \varphi^+) - \mu\Gamma^u(\omega_\lambda, \theta^-, \varphi^-)$$
$$= \mu\big(a_3(\lambda) + b_3(\lambda)\big),$$

where

$$a_3(\lambda) := -\int_0^{T_\lambda/2} g\big(\omega_\lambda t + \varphi^+, Q_{T_\lambda}(t), t + \theta^+\big)\,\mathrm{d}t + \int_0^{\infty} g(\omega_\lambda t + \varphi^+, q_0(t), t + \theta^+)\,\mathrm{d}t,$$

$$b_3(\lambda) := -\int_{-T_\lambda/2}^{0} g\big(\omega_\lambda t + \varphi^-, Q_{T_\lambda}(t + T_\lambda), t + \theta^-\big)\,\mathrm{d}t$$
$$+ \int_{-\infty}^{0} g(\omega_\lambda t + \varphi^-, q_0(t), t + \theta^-)\,\mathrm{d}t.$$

We have:

$$a_3(\lambda) = -\int_0^{T_\lambda/2} \big[g\big(\omega_\lambda t + \varphi^+, Q_{T_\lambda}(t), t + \theta^+\big) - g\big(\omega_\lambda t + \varphi^+, q_0(t), t + \theta^+\big)\big]$$
$$+ \int_{T_\lambda/2}^{\infty} g\big(\omega_\lambda t + \varphi^+, q_0(t), t + \theta^+\big).$$

Recalling that $\sup_{t\in(0,T/2)}|\partial_T Q_T(t)| = \mathrm{O}(\mathrm{e}^{-c_2 T})$, $\sup_{t\in(0,T/2)}|Q_T(t)-q_0(t)| = \mathrm{O}(\mathrm{e}^{-c_2 T})$, it is easy to see that the derivatives of the first integral are $\mathrm{O}(T_\lambda \mathrm{e}^{-c_2 T_\lambda}) = \mathrm{O}(\mu)$ (still provided C_1 is large enough). Moreover, using that $(|g(\omega_\lambda t+\varphi^+, q_0(t), t)| + |\partial_\varphi g(\omega_\lambda t+\varphi^+, q_0(t), t)| + |\partial_t g(\omega_\lambda t+\varphi^+, q_0(t), t)|) = \mathrm{O}(q_0(t)-2\pi) = \mathrm{O}(\mathrm{e}^{-c_2 t})$ for $t\in(T_\lambda/2, +\infty)$, we find that the derivatives of the second integral are $\mathrm{O}(\mu)$ as well. Hence $|\nabla a_3(\lambda)| = \mathrm{O}(\mu)$. The same estimate holds for b_3. We then conclude that $\nabla R_3(\lambda) = \mathrm{O}(\mu^2)$, which completes the proof of Lemma 3.1. □

In Section 6 we will look for a critical point of $\mathcal{F}_\mu$ in the set:

$$E := \big\{\lambda = (\theta_1,\dots,\theta_k,\varphi_1,\dots,\varphi_k)\in \mathbf{R}^k\times\mathbf{R}^{kd} \,\big|\, \theta_i = \bar\theta_i + b_i,\ \varphi_i = \bar\varphi_i + a_i,\ |b_i|\leqslant 2\pi,\ |a_i|\leqslant 2\pi\big\}, \tag{3.4}$$

where $k, \bar\varphi_i, \bar\theta_i$ will be defined in Section 5. It will result that $E\subset \Lambda_{\mu,k}$ (for some $\beta>0$ depending on the curve γ). In particular, for all $\lambda\in E$

$$C_1|\ln\mu| \leqslant \theta_{i+1}-\theta_i < \frac{C_0\beta_i}{\mu}, \quad \forall i = 1,\dots,k-1, \tag{3.5}$$

where $\beta_i := \beta_{\lambda_i} := \beta(\omega_i)$ and $\omega_i := \omega_{\lambda_i} := (\varphi_{i+1}-\varphi_i)/(\theta_{i+1}-\theta_i)$. Moreover we will assume (see (5.8))

$$|\bar\omega_{i+1}-\bar\omega_i| \leqslant \rho\mu, \quad \text{where} \quad \bar\omega_i := \frac{\bar\varphi_{i+1}-\bar\varphi_i}{\bar\theta_{i+1}-\bar\theta_i}\ (1\leqslant i\leqslant k-1),\ \omega_0 := \omega_I,\ \omega_k := \omega_F \tag{3.6}$$

and $\rho>0$ is a small constant to be chosen later (see (6.3)). For the time being, assuming (3.5) and (3.6), we want to give a suitable expression of $\mathcal{F}_\mu$ in E. By Lemma 3.1, for $\lambda\in E$, we have

$$\begin{aligned}\mathcal{F}_\mu(\lambda) = {}& \sum_{i=1}^{k-1}\frac12\frac{|\varphi_{i+1}-\varphi_i|^2}{\theta_{i+1}-\theta_i} + \omega_I\varphi_1 - \omega_F\varphi_k - \frac{|\omega_I|^2}{2}\theta_1 + \frac{|\omega_F|^2}{2}\theta_k \\ & + \sum_{i=1}^{k}\mu\big(\Gamma^u(\omega_{i-1},\theta_i,\varphi_i) + \Gamma^s(\omega_i,\theta_i,\varphi_i)\big) + \mu F(\omega_I,\theta_1,\varphi_1) \\ & - \sum_{i=1}^{k-1}\mu\int_{\theta_i}^{\theta_{i+1}} f\big(\omega_i(t-\theta_i)+\varphi_i, 0, t\big)\,\mathrm{d}t - \mu F(\omega_F,\theta_k,\varphi_k) \\ & + \sum_{i=1}^{k-1} R_0(\mu,\lambda_i),\end{aligned} \tag{3.7}$$

where $|\nabla_\lambda R_0(\mu,\lambda)|$ satisfies (3.2). We shall write $\mathcal{F}_\mu$ in an appropriate form thanks to the following lemmas. The first one says how close the "mean frequencies" ω_i are to the unperturbed $\overline{\omega}_i$.

Lemma 3.2. *Let* $\lambda = (\theta_1,\dots,\theta_k,\varphi_1,\dots,\varphi_k)$ *belong to* E. *Then*

$$|\omega_i - \overline{\omega}_i| = \mathrm{O}\left(\frac{1}{\theta_{i+1}-\theta_i}\right) = \mathrm{O}\left(\frac{1}{|\ln\mu|}\right). \tag{3.8}$$

Moreover,

$$\begin{aligned}&\Gamma^u(\omega_{i-1},\theta_i,\varphi_i) + \Gamma^s(\omega_i,\theta_i,\varphi_i) = \Gamma(\overline{\omega}_i,\theta_i,\varphi_i) + R_4(\lambda_i),\\ &\quad \textit{where } \nabla R_4 = \mathrm{O}\big(1/|\ln\mu|\big).\end{aligned} \tag{3.9}$$

Proof. Set $\Delta\theta_i := \theta_{i+1}-\theta_i$, $\Delta a_i := a_{i+1}-a_i$ and $\Delta b_i := b_{i+1}-b_i$. By an elementary computation we get $\omega_i - \overline{\omega}_i = -\overline{\omega}_i\Delta b_i/\Delta\theta_i + \Delta a_i/\Delta\theta_i$. By the definition of E and (3.5), estimate (3.8) follows.

From the definition of Γ^u, Γ^s and the exponential decay of q_0 it results that $\partial_\omega\Gamma^{u,s}$ is bounded by a uniform constant, as well as its partial derivatives. Hence (3.9) is a straightforward consequence of (3.8) and of (3.6). □

Lemma 3.3. *For* $0<\mu\leqslant\mu_4$

$$\begin{aligned}&\mu F(\omega_I,\theta_1,\varphi_1) - \sum_{i=1}^k \mu\int_{\theta_i}^{\theta_{i+1}} f(\omega_i(t-\theta_i)+\varphi_i,0,t)\,\mathrm{d}t - \mu F(\omega_F,\theta_k,\varphi_k)\\ &\quad = \sum_{i=1}^k R_5^i(\mu,\lambda_{i-1},\lambda_i),\end{aligned} \tag{3.10}$$

where, for all i[3]

$$\begin{aligned}&\nabla R_5^i(\mu,\theta_{i-1},\varphi_{i-1},\theta_i,\varphi_i,\theta_{i+1},\varphi_{i+1})\\ &\quad = \mathrm{O}\left(\frac{\mu}{\beta_{i-1}^2(\theta_i-\theta_{i-1})} + \frac{\mu}{\beta_i^2(\theta_{i+1}-\theta_i)} + \frac{\mu|\beta_i-\beta_{i-1}|}{\beta_{i-1}\beta_i}\right).\end{aligned} \tag{3.11}$$

Proof. We have

$$-\int_{\theta_i}^{\theta_{i+1}} f\big(\varphi_i+\omega_i(t-\theta_i),0,t\big)\,\mathrm{d}t = F(\omega_i,\theta_{i+1},\varphi_{i+1}) - F(\omega_i,\theta_i,\varphi_i)$$

[3] In the cases $i=1$, $i=k$ we only have $R_5^1 = R_5^1(\mu,\theta_1,\varphi_1,\theta_2,\varphi_2)$ and $R_5^k = R_5^k(\mu,\theta_{k-1},\varphi_{k-1},\theta_k,\varphi_k)$.

$$= \big(F(\omega_i, \theta_{i+1}, \varphi_{i+1})\big) - \big(F(\omega_{i-1}, \theta_i, \varphi_i)\big) + \big(F(\omega_{i-1}, \theta_i, \varphi_i) - F(\omega_i, \theta_i, \varphi_i)\big),$$

where $F(\omega, \cdot, \cdot)$ is defined in (2.9). We obtain:

$$\mu F(\omega_I, \theta_1, \varphi_1) - \sum_{i=1}^{k-1} \mu \int_{\theta_i}^{\theta_{i+1}} f\big(\varphi_i + \omega_i(t - \theta_i), 0, t\big)\,\mathrm{d}t - \mu F(\omega_F, \theta_k, \varphi_k) = \sum_{i=1}^{k} R_5^i,$$

where

$$\begin{aligned} R_5^i &:= R_5^i(\mu, \theta_{i-1}, \varphi_{i-1}, \theta_i, \varphi_i, \theta_{i+1}, \varphi_{i+1}) := \mu\big(F(\omega_{i-1}, \theta_i, \varphi_i) - F(\omega_i, \theta_i, \varphi_i)\big) \\ &= -\mu \sum_{0<|(n,l)|\leqslant N} f_{n,l} \frac{\mathrm{e}^{\mathrm{i}(n\cdot\varphi_i + l\theta_i)}}{\mathrm{i}} \left(\frac{1}{(n\cdot\omega_{i-1} + l)} - \frac{1}{(n\cdot\omega_i + l)} \right). \end{aligned}$$

Now we prove (3.11). Let us consider for example $\partial_{\theta_i} R_5^i$. We have:

$$\begin{aligned} \partial_{\theta_i} R_5^i &= \mu \partial_{\theta_i}\big(F(\omega_{i-1}, \theta_i, \varphi_i) - F(\omega_i, \theta_i, \varphi_i)\big) \\ &= \mu\left(\partial_\omega F(\omega_{i-1}, \theta_i, \varphi_i).\frac{-\omega_{i-1}}{(\theta_i - \theta_{i-1})} - \partial_\omega F(\omega_i, \theta_i, \varphi_i).\frac{\omega_i}{(\theta_{i+1} - \theta_i)} \right) \\ &\quad - \mu\left(\sum_{0<|(n,l)|\leqslant N} f_{n,l}\, l\, \mathrm{e}^{\mathrm{i}(n\cdot\varphi_i + l\theta_i)} \left(\frac{1}{(n\cdot\omega_{i-1} + l)} - \frac{1}{(n\cdot\omega_i + l)} \right) \right), \end{aligned} \tag{3.12}$$

where

$$\partial_\omega F(\omega, \theta_0, \varphi_0) = \sum_{0<|(n,l)|\leqslant N} f_{n,l} \frac{n\,\mathrm{e}^{\mathrm{i}(n\cdot\varphi_0 + l\theta_0)}}{\mathrm{i}(n\cdot\omega + l)^2}. \tag{3.13}$$

Estimate (3.11) follows immediately from (3.12) and (3.13). The other partial derivatives of R_5^i can be estimated similarly. □

Finally, to get a suitable expression of $\mathcal{F}_\mu$, we find convenient to introduce coordinates $(b, c) \in \mathbf{R}^{(1+d)k}$ defined by (3.4) and

$$c_i = a_i - \overline{\omega}_i b_i, \quad \forall i = 1, \ldots, k \tag{3.14}$$

(we are just performing a linear change of coordinates adapted to the direction of the unperturbed flow at each i-transition $(b_i, a_i) = b_i(1, \overline{\omega}_i) + (0, c_i)$).

Lemma 3.4. *We have*:

$$\sum_{i=1}^{k-1}\frac{1}{2}\frac{|\varphi_{i+1}-\varphi_i|^2}{(\theta_{i+1}-\theta_i)}+\omega_I\varphi_1-\omega_F\varphi_k-\frac{|\omega_I|^2}{2}\theta_1+\frac{|\omega_F|^2}{2}\theta_k$$
$$=\frac{1}{2}\sum_{i=1}^{k-1}\frac{|c_{i+1}-c_i|^2}{\Delta\bar\theta_i+(b_{i+1}-b_i)}+\sum_{i=1}^{k}R_6^i(\mu,\theta_i,\varphi_i,\theta_{i+1},\varphi_{i+1}), \tag{3.15}$$

where $\Delta\bar\theta_i:=\bar\theta_{i+1}-\bar\theta_i$ *and*[4]

$$\nabla R_6^i(\mu,\theta_{i-1},\varphi_{i-1},\theta_i,\varphi_i,\theta_{i+1},\varphi_{i+1})=\mathrm{O}(\Delta\overline{\omega}_i)=\mathrm{O}(\rho\mu). \tag{3.16}$$

Proof. Let $\{\gamma_i\}_{i=1,\dots,k-1}$ be defined by $\varphi_{i+1}-\varphi_i=\overline{\omega}_i(\theta_{i+1}-\theta_i)+\gamma_i$. We can write $\omega_I\varphi_1-\omega_F\varphi_k$ as

$$\omega_I\varphi_1-\omega_F\varphi_k=\sum_{i=1}^{k-1}\big((\overline{\omega}_{i-1}-\overline{\omega}_i)\varphi_i-\overline{\omega}_i(\varphi_{i+1}-\varphi_i)\big)+\varphi_k(\overline{\omega}_{k-1}-\omega_F)$$
$$=\sum_{i=1}^{k-1}\big((\overline{\omega}_{i-1}-\overline{\omega}_i)\varphi_i-|\overline{\omega}_i|^2(\theta_{i+1}-\theta_i)-\overline{\omega}_i\gamma_i\big)$$
$$+\varphi_k(\overline{\omega}_{k-1}-\omega_F). \tag{3.17}$$

We can also write:

$$-\frac{|\omega_I|^2}{2}\theta_1+\frac{|\omega_F|^2}{2}\theta_k=\sum_{i=1}^{k-1}\left(\left(\frac{|\overline{\omega}_i|^2}{2}-\frac{|\overline{\omega}_{i-1}|^2}{2}\right)\theta_i+\frac{|\overline{\omega}_i|^2}{2}(\theta_{i+1}-\theta_i)\right)$$
$$+\left(\frac{|\omega_F|^2}{2}-\frac{|\overline{\omega}_{k-1}|^2}{2}\right)\theta_k, \tag{3.18}$$

$$\sum_{i=1}^{k-1}\frac{1}{2}\frac{|\varphi_{i+1}-\varphi_i|^2}{(\theta_{i+1}-\theta_i)}=\sum_{i=1}^{k-1}\frac{|\overline{\omega}_i|^2}{2}(\theta_{i+1}-\theta_i)+\frac{1}{2}\frac{|\gamma_i|^2}{(\theta_{i+1}-\theta_i)}+\overline{\omega}_i\gamma_i. \tag{3.19}$$

Summing (3.17), (3.18) and (3.19) we get

$$\sum_{i=1}^{k-1}\frac{1}{2}\frac{|\varphi_{i+1}-\varphi_i|^2}{(\theta_{i+1}-\theta_i)}+\omega_I\varphi_1-\omega_F\varphi_k-\frac{|\omega_I|^2}{2}\theta_1+\frac{|\omega_F|^2}{2}\theta_k$$
$$=\sum_{i=1}^{k-1}\frac{1}{2}\frac{|\gamma_i|^2}{(\theta_{i+1}-\theta_i)}+\sum_{i=1}^{k-1}\left(\frac{|\overline{\omega}_i|^2}{2}-\frac{|\overline{\omega}_{i-1}|^2}{2}\right)\theta_i+(\overline{\omega}_{i-1}-\overline{\omega}_i)\varphi_i+\varphi_k(\overline{\omega}_{k-1}-\omega_F)$$

4 For $i=k$ we have $R_6^k=R_6^k(\mu,\theta_k,\varphi_k)$.

$$+\left(\frac{|\omega_F|^2}{2}-\frac{|\overline{\omega}_{k-1}|^2}{2}\right)\theta_k. \tag{3.20}$$

Substituting $\overline{\varphi}_i + a_i$ for φ_i and $\bar{\theta}_i + b_i$ for θ_i, we get $\gamma_i = (a_{i+1} - a_i) - \overline{\omega}_i(b_{i+1} - b_i)$. Moreover the nonconstant terms in the right-hand side of (3.20) (i.e., those depending on a_i, b_i) are the first one and

$$\sum_{i=1}^{k}(\overline{\omega}_{i-1}-\overline{\omega}_i)a_i+\left(\frac{|\overline{\omega}_i|^2}{2}-\frac{|\overline{\omega}_{i-1}|^2}{2}\right)b_i =: \sum_{i=1}^{k} R^i(\mu,\theta_i,\varphi_i)$$

with $\nabla R^i(\mu,\theta_i,\varphi_i) = \mathrm{O}(\Delta\overline{\omega}_i)$. Finally, expressing γ_i in terms of (b_i, c_i) we get

$$\gamma_i = (a_{i+1}-a_i)-\overline{\omega}_i(b_{i+1}-b_i) = (c_{i+1}-c_i)+b_{i+1}\Delta\overline{\omega}_i$$

and then from (3.20), developing the square, we get (3.16). □

From (3.7) Lemmas 3.2, 3.3 and 3.4 we obtain the expression of $\mathcal{F}_\mu$ in the new coordinates (b, c) required to apply the variational argument of Section 6.

Lemma 3.5. *There exists $\mu_5, C_2 > 0$ such that $\forall 0 < \mu \leqslant \mu_5$, if*

$$\beta_i \geqslant C_2 \max\left\{\mu^{1/2}(\theta_{i+1}-\theta_i)^{1/2},\ \mu(\theta_{i+1}-\theta_i)^{3/2},\ (\theta_{i+1}-\theta_i)^{-1/2}\right\}, \tag{3.21}$$

then

$$\mathcal{F}_\mu(b,c) = \frac{1}{2}\sum_{i=1}^{k-1}\frac{|c_{i+1}-c_i|^2}{\Delta\bar{\theta}_i+(b_{i+1}-b_i)}+\mu\sum_{i=1}^{k}\Gamma(\overline{\omega}_i,\bar{\theta}_i+b_i,\overline{\varphi}_i+\overline{\omega}_i b_i+c_i)$$
$$+ R_7(b,c), \tag{3.22}$$

$$R_7(b,c) := \sum_{i=1}^{k} R_7^i(\mu, b_{i-1}, c_{i-1}, b_i, c_i, b_{i+1}, c_{i+1}), \tag{3.23}$$

where[5]

$$|\nabla R_7^i| \leqslant C_2\rho\mu. \tag{3.24}$$

Proof. It is easy to see that (3.6), (3.8) and (3.21) imply (provided μ is small enough) that

$$\frac{\beta_{i-1}}{2} \leqslant \beta_i \leqslant 2\beta_{i-1}, \quad |\beta_i-\beta_{i-1}| = \mathrm{O}\left(\frac{1}{\theta_i-\theta_{i-1}}+\frac{1}{\theta_{i+1}-\theta_i}+\mu\right). \tag{3.25}$$

[5] In the cases $i = 1, i = k$ we have $R_7^1 = R_7^1(\mu,\theta_1,\varphi_1,\theta_2,\varphi_2)$ and $R_7^k = R_7^k(\mu,\theta_{k-1},\varphi_{k-1},\theta_k,\varphi_k)$.

Noting that $\partial_{c_i} = \partial_{\varphi_i}$ and $\partial_{b_i} = \overline{\omega}_i \partial_{\varphi_i} + \partial_{\theta_i}$, estimate (3.24) follows from (3.2), (3.9), (3.11), (3.25) and (3.16). □

4. Ergodization times

In order to define $\overline{\varphi}_i, \bar{\theta}_i$ $(1 \leqslant i \leqslant k)$ we need some results, stated in this section, on the ergodization time of the torus $\mathbf{T}^l := \mathbf{R}^l/\mathbf{Z}^l$ for linear flows possibly resonant but only at a "sufficiently high level".

Let $\Omega \in \mathbf{R}^l$; it is well known that, if $\Omega \cdot p \neq 0, \forall p \in \mathbf{Z}^l \setminus \{0\}$, then the trajectories of the linear flow $\{\Omega t + A\}_{t \in \mathbf{R}}$ are dense on $\mathbf{T}^l$ for any initial point $A \in \mathbf{T}^l$. It is also intuitively clear that the trajectories of the linear flow $\{\Omega t + A\}_{t \in \mathbf{R}}$ will make an arbitrary fine δ-net $(\delta > 0)$ if Ω is resonant only at a sufficiently high level, namely if $\Omega \cdot p \neq 0,\ \forall p \in \mathbf{Z}^l$ with $0 < |p| \leqslant M(\delta)$ for some large enough $M(\delta)$. Let us make more precise and quantitative these considerations.

For any $\Omega \in \mathbf{R}^l$ define the ergodization time $T(\Omega, \delta)$ required to fill $\mathbf{T}^l$ within $\delta > 0$ as

$$T(\Omega, \delta) = \inf\bigl\{t \in \mathbf{R}_+ \bigm| \forall x \in \mathbf{R}^l,\ d\bigl(x, A + [0, t]\Omega + \mathbf{Z}^l\bigr) \leqslant \delta\bigr\},$$

where d is the Euclidean distance and A some point of $\mathbf{R}^l$. $T(\Omega, \delta)$ is clearly independent of the choice of A. Above and in what follows, $\inf E$ is equal to $+\infty$ if E is empty. For $R > 0$ let

$$\alpha(\Omega, R) = \inf\bigl\{|p \cdot \Omega| \bigm| p \in \mathbf{Z}^l,\ p \neq 0,\ |p| \leqslant R\bigr\}.$$

Theorem 4.1. $\forall l \in \mathbf{N}$ *there exists a positive constant* a_l *such that,* $\forall \Omega \in \mathbf{R}^l$, $\forall \delta > 0$, $T(\Omega, \delta) \leqslant (\alpha(\Omega, a_l/\delta))^{-1}$. *Moreover* $T(\Omega, \delta) \geqslant (1/4)\alpha(\Omega, 1/4\delta)^{-1}$.

In the above theorem α^{-1} is equal to 0 if $\alpha = +\infty$ and to $+\infty$ if $\alpha = 0$.

Remark 4.1. Assume that Ω is a C-τ Diophantine vector, i.e., there exist $C > 0$ and $\tau \geqslant l - 1$ such that $\forall k \in \mathbf{Z}^l$ $|k \cdot \Omega| \geqslant C/|k|^\tau$. Then $\alpha(\Omega, R) \geqslant C/R^\tau$ and so $T(\Omega, \delta) \leqslant a_l^\tau / C\delta^\tau$. This estimate was proved in Theorem D of [11]. Also Theorem B of [11] is an easy consequence of Theorem 4.1.

Theorem 4.1 is a direct consequence of more general statements, see Theorem 4.2 and Remark 4.2. Let us introduce first some notations. Let Λ be a lattice of $\mathbf{R}^l$, i.e., a discrete subgroup of $\mathbf{R}^l$ such that $\mathbf{R}^l/\Lambda$ has finite volume. For all $\Omega \in \mathbf{R}^l$ we define:

$$T(\Lambda, \Omega, \delta) = \inf\bigl\{t \in \mathbf{R}_+ \bigm| \forall x \in \mathbf{R}^l,\ d(x, [0, t]\Omega + \Lambda) \leqslant \delta\bigr\},$$

($T(\Lambda, \Omega, \delta)$ is the time required to have a δ-net of the torus $\mathbf{R}^l/\Lambda$ endowed with the metric inherited from $\mathbf{R}^l$). For $R > 0$, let

$$\Lambda^* = \bigl\{p \in \mathbf{R}^l \bigm| \forall \lambda \in \Lambda,\ p \cdot \lambda \in \mathbf{Z}\bigr\} \quad \text{and} \quad \Lambda^*_R = \bigl\{p \in \Lambda^* \bigm| 0 < |p| \leqslant R\bigr\}$$

(Λ^* is a lattice of $\mathbf{R}^l$ which is conjugated to Λ). We define:

$$\alpha(\Lambda, \Omega, R) = \inf\bigl\{|p \cdot \Omega| \bigm| p \in \Lambda^*_R\bigr\}.$$

The following result holds:

Theorem 4.2. $\forall l \in \mathbf{N}$ *there exists a positive constant* a_l *such that, for all lattice* Λ *of* $\mathbf{R}^l$, $\forall \Omega \in \mathbf{R}^l$, $\forall \delta > 0$, $T(\Lambda, \Omega, \delta) \leqslant (\alpha(\Lambda, \Omega, a_l/\delta))^{-1}$.

Remark 4.2. It is fairly obvious that $T(\Lambda, \Omega, \delta) \geqslant (1/4)\alpha(\Lambda, \Omega, 1/4\delta)^{-1}$. Indeed, assume that $\Lambda^*_{1/4\delta} \neq \emptyset$ and let $p \in \Lambda^*_{1/4\delta}$ be such that $p \cdot \Omega = \alpha := \alpha(\Lambda, \Omega, 1/4\delta)$. Let $x \in \mathbf{R}^l$ satisfy $p \cdot x = 1/2$. Then $\forall t \in [0, 1/4\alpha)$, $\forall \lambda \in \Lambda$,

$$\bigl|x - (t\Omega + \lambda)\bigr| \geqslant \frac{|p \cdot (x - t\Omega - \lambda)|}{|p|} \geqslant 4\delta\bigl|p \cdot x - tp \cdot \Omega - p \cdot \lambda\bigr|,$$

and $p \cdot x - p \cdot \lambda \in (1/2) + \mathbf{Z}$, whereas $|tp \cdot \Omega| = t\alpha < 1/4$. Hence $|x - (t\Omega + \lambda)| > \delta$.

In the next section we will apply Theorem 4.1 when $\Omega = (\omega, 1) \in \mathbf{R}^{d+1}$. The proof of Theorem 4.2 is given in the Appendix B. We could give an explicit expression of a_l. However it is not useful for our purpose and the constants a_l which can be derived from our proof are certainly far from being optimal.

5. The unperturbed pseudo-diffusion orbit

Consider the set Q_M of "nonergodizing frequencies"

$$Q_M := \bigl\{\omega \in \mathbf{R}^d \bigm| \exists (n, l) \in \mathbf{Z}^{d+1} \text{ with } 0 < \bigl|(n, l)\bigr| \leqslant M, \text{ and } \omega \cdot n + l = 0\bigr\} = \bigcup_{h \in S_M} E_h,$$

where $S_M := \{h = (n, l) \in (\mathbf{Z}^d \setminus \{0\}) \times \mathbf{N} \mid 0 < |h| \leqslant M,\ h \neq jh',\ \forall j \in \mathbf{Z},\ h' \in (\mathbf{Z}^d \setminus \{0\}) \times \mathbf{N}\}$ and $E_h = E_{n,l} := \{\omega \in \mathbf{R}^d \mid (\omega, 1) \cdot h = \omega \cdot n + l = 0\}$. By Theorem 4.1 (or Theorem 4.2, with $\Lambda = 2\pi \mathbf{Z}^{d+1}$), for $\delta > 0$, if ω belongs to

$$Q^c_M = \bigl\{\omega \in \mathbf{R}^d \bigm| \omega \cdot n + l \neq 0,\ \forall 0 < \bigl|(n, l)\bigr| \leqslant M\bigr\}, \tag{5.1}$$

with $M = 8\pi a_{d+1}/\delta$, then the flow of $(\omega, 1)$ provides a $\delta/4$-net of the torus $\mathbf{T}^{d+1}$.

Moreover if $\omega \notin Q_M$ then for all $(n, l) \in \mathbf{Z}^d \setminus \{0\} \times \mathbf{Z}$,

$$|n \cdot \omega + l| = |n| \operatorname{dist}(\omega, E_{n,l}) \geqslant \operatorname{dist}(\omega, E_{n,l}) \geqslant \operatorname{dist}(\omega, Q_M) > 0. \tag{5.2}$$

By Theorem 4.1 (or Theorem 4.2), we deduce from (5.2) the estimate,

$$T\bigl((\omega, 1), \delta/4\bigr) \leqslant \frac{2\pi}{\operatorname{dist}(\omega, Q_M)}, \tag{5.3}$$

which measures the divergence of the ergodization time $T((\omega, 1), \delta)$ as ω approaches the set Q_M.

Definition 5.1. Given $M > 0$, a connected component $\mathcal{C}$ of $\mathcal{D}_N^c$ and $\omega_I, \omega_F \in \mathcal{C}$, we say that an embedding $\gamma \in C^2([0, L], \mathcal{C})$ is a *Q_M-admissible connecting curve between ω_I and ω_F* if the following properties are satisfied:

(a) $\gamma(0) = \omega_I$, $\gamma(L) = \omega_F$, $|\dot\gamma(s)| = 1$ $\forall s \in (0, L)$,
(b) $\forall h = (n, l) \in S_M$, $\forall s \in [0, L]$ such that $\gamma(s) \in E_h$, $n \cdot \dot\gamma(s) \neq 0$.

Condition (b) means that for all $h \in S_M$, $\gamma([0, L])$ may intersect E_h transversally only. It is easy to see that condition (b) implies that $\mathcal{I}(\gamma) = \{s \in [0, L] \mid \gamma(s) \in Q_M\}$ is finite and that there exists $\nu > 0$ such that for all $s \in \mathcal{I}(\gamma)$, for all $h = (n, l) \in S_M$ such that $\gamma(s) \in E_h$, $|\dot\gamma(s) \cdot n|/|n| \geqslant \nu$.

If a curve α is not admissible we can always find "close to it" an admissible one γ. Indeed the following lemma holds.

Lemma 5.1. *Let $M > 0$, $\mathcal{C}$ be a connected component of $\mathcal{D}_N^c$, $\omega_I, \omega_F \in \mathcal{C}$ and let $\alpha \in C^2([0, L_0], \mathcal{C})$ be an embedding with $\alpha(0) = \omega_I$ and $\alpha(L_0) = \omega_F$. Then, $\forall \eta > 0$, there exists a curve γ, Q_M-admissible between ω_I and ω_F, satisfying* $\operatorname{dist}(\gamma(s), \alpha([0, L_0])) < \eta$, $\forall s \in [0, L]$.

Proof. First it is easy to see that there exists an embedding $\alpha_1 : [0, L_1] \to \mathcal{C}$ such that $\alpha_1(0) = \omega_I$, $\alpha_1(L_1) = \omega_F$, $\operatorname{dist}(\alpha_1(s), \alpha([0, L_0])) \leqslant \eta/4$ and $\forall h = (n, l) \in S_M$, $\omega_I \notin E_h$ (respectively $\omega_F \notin E_h$) or $\dot\alpha_1(0) \cdot n \neq 0$ (respectively $\dot\alpha_1(L_1) \cdot n \neq 0$).

Let $r > 0, \nu_1 > 0$ be such that $\forall s \in [0, r] \cup [L_1 - r, L_1]$, $\forall h = (n, l) \in S_M$, $\operatorname{dist}(\alpha_1(s), E_h) \geqslant \nu_1$ or $|\dot\alpha_1(s) \cdot n| \geqslant \nu_1$. Let $\phi : [0, L_1] \to [0, 1]$ be a smooth function such that $\phi(0) = \phi(L_1) = 0$ and $\forall s \in [r, L_1 - r]$ $\phi(s) = 1$.

We shall prove that for all $\varepsilon > 0$ there exists $\omega_\varepsilon \in \mathbf{R}^d$, $|\omega_\varepsilon| < \varepsilon$, such that $\forall h = (n, l) \in S_M$, for all $s \in [r, L_1 - r]$ such that $\alpha_1(s) \in E_h + \omega_\varepsilon$, $\dot\alpha_1(s) \cdot n \neq 0$. For $h = (n, l) \in S_M$, let $\mathcal{J}_h = \{s \in [r, L_1 - r] \mid n \cdot \dot\alpha_1(s) = 0\}$ and $\mathcal{V}_h = \{\alpha_1(s) - u \mid s \in \mathcal{J}_h, u \in E_h\}$. Let $\psi_h : [r, L_1 - r] \times E_h \to \mathbf{R}^d$ be defined by $\psi_h(s, u) = \alpha_1(s) - u$. $D\psi_h(s, u)$ is singular iff $s \in \mathcal{J}_h$. Therefore $\mathcal{V}_h$ is the set of the critical values of ψ_h and by Sard's lemma, $\operatorname{meas}(\mathcal{V}_h) = 0$. Hence for all $\varepsilon > 0$ there exists $\omega_\varepsilon \in \mathbf{R}^d$ such that $|\omega_\varepsilon| < \varepsilon$, $\omega_\varepsilon \notin \mathcal{V}_h$ for all $h \in S_M$. Our claim follows.

Now we can define $\alpha_2 : [0, L_1] \to \mathcal{C}$ by $\alpha_2(s) = \alpha_1(s) - \phi(s)\omega_\varepsilon$. It is easy to check that, provided ε is small enough, α_2 is an embedding which satisfies condition (b). γ is obtained from α_2 by a simple time reparametrization. □

If $\Gamma(\alpha(s), \cdot, \cdot)$ possesses, for each s, a nondegenerate local minimum $(\theta_0^{\alpha(s)}, \varphi_0^{\alpha(s)})$, then, by the Implicit Function Theorem, along any curve γ sufficiently close to α, $\Gamma(\gamma(s), \cdot, \cdot)$ possesses local minima $(\theta_0^{\gamma(s)}, \varphi_0^{\gamma(s)})$ such that

$$D^2_{(\theta,\varphi)} \Gamma\big(\gamma(s), \theta_0^{\gamma(s)}, \varphi_0^{\gamma(s)}\big) > \lambda \,\mathrm{Id}, \quad \forall s \in [0, L], \tag{5.4}$$

for some constant $\lambda > 0$ depending on α. Therefore, by the above lemma, it is enough to prove the existence of drifting orbits along admissible curves γ. Property (5.4) will be used in Lemma 6.1.

Given a Q_M-admissible curve γ, let us call $s_1^*, \dots, s_r^*$ the elements of $\mathcal{I}(\gamma)$, and $\omega_1^* = \gamma(s_1^*), \dots, \omega_r^* = \gamma(s_r^*)$ the corresponding frequencies. Since, $\forall m = 1, \dots, r$, $(\theta_0^{\omega_m^*}, \varphi_0^{\omega_m^*})$ is a nondegenerate local minimum of $\Gamma(\omega_m^*, \cdot, \cdot)$, there is a neighborhood W_m of ω_m^* such that, $\forall \omega \in W_m$, $\Gamma(\omega, \cdot)$ admits a nondegenerate local minimum $(\theta_0^\omega, \varphi_0^\omega)$, the map $\omega \mapsto (\theta_0^\omega, \varphi_0^\omega)$ being Lipschitz-continuous on W_m. Therefore we shall assume without loss of generality that for all $m = 1, \dots, r$,

$$\forall (\omega, \omega') \in \big(W_m \cap \gamma\big([0, L]\big)\big)^2 \big|\big(\theta_0^\omega, \varphi_0^\omega\big) - \big(\theta_0^{\omega'}, \varphi_0^{\omega'}\big)\big| \leqslant K|\omega - \omega'|. \tag{5.5}$$

It is easy to prove that, if γ is an admissible curve, there exists $d_0 > 0$ such that

($*$) $\{s \in [0, L] \mid \mathrm{dist}(\gamma(s), Q_M) \leqslant d_0\}$ is the union of a finite number of disjoint intervals $[S_1, S_1'], \dots, [S_r, S_r']$; for all $m = 1, \dots, r$ each interval $[S_m, S_m']$ intersects $\mathcal{I}(\gamma)$ at a unique point s_m^* and $\gamma([S_m, S_m']) \subset W_m$. Moreover $(s \mapsto \mathrm{dist}(\gamma(s), Q_M))$ is decreasing on $[S_m, s_m^*)$, increasing on $(s_m^*, S_m']$, and $\mathrm{dist}(\gamma(s), Q_M) \geqslant (\nu/2)|s - s_m^*|$ for all $s \in [S_m, S_m']$.

Now we are able to define the "unperturbed transition chain": for some small constant $\rho > 0$ which will be specified later we choose $k \in \mathbf{N}$ and $k + 1$ "intermediate frequencies":

$$\omega_I =: \overline{\omega}_0, \overline{\omega}_1, \dots, \overline{\omega}_{k-1}, \overline{\omega}_k := \omega_F$$

with $\overline{\omega}_i := \gamma(s_i)$ for certain $0 =: s_0 < s_1 < \cdots < s_{k-1} < s_k := L$ verifying

$$\frac{\rho\mu}{2} \leqslant s_{i+1} - s_i \leqslant \rho\mu, \quad \forall i = 0, \dots, k-1. \tag{5.6}$$

By (5.6) there results that

$$\frac{L}{\rho\mu} \leqslant k \leqslant \frac{2L}{\rho\mu}, \tag{5.7}$$

moreover it follows from (a) that

$$|\overline{\omega}_{i+1} - \overline{\omega}_i| \leqslant \rho\mu, \quad \forall i = 0, \dots, k-1. \tag{5.8}$$

This condition has been used before in Lemma 3.4. Given k time instants $\bar{\theta}_1 := \theta_0^{\overline{\omega}_1} < \bar{\theta}_2 < \cdots < \bar{\theta}_i < \cdots < \bar{\theta}_k$, we define the $\{\overline{\varphi}_i\}_{i=1,\dots,k}$ by the iteration formula:

$$\overline{\varphi}_1 = \varphi_0^{\overline{\omega}_1}, \quad \overline{\varphi}_{i+1} = \overline{\varphi}_i + \overline{\omega}_i\big(\bar{\theta}_{i+1} - \bar{\theta}_i\big). \tag{5.9}$$

The choice of the instants $\{\bar{\theta}_i\}_{i=1,\dots,k}$ is specified in the next lemma: the main request is that $(\bar{\theta}_i, \overline{\varphi}_i)$ must arrive δ-close $\bmod 2\pi\mathbf{Z}^{d+1}$, to the local minimum point $(\theta_0^{\overline{\omega}_i}, \varphi_0^{\overline{\omega}_i})$

of the Poincaré–Melnikov primitive $\Gamma(\overline{\omega}_i, \cdot, \cdot)$, see (5.11)–(5.12). From (5.3) we derive that if $\overline{\omega}_i$ is $1/|\ln\mu|$ far from the set Q_M of "nonergodizing frequencies" we can reach this goal for "short" time intervals $\bar{\theta}_{i+1} - \bar{\theta}_i \approx |\ln\mu|$. In order to cross the set Q_M of "nonergodizing frequencies" we need to use longer time intervals $\bar{\theta}_{i+1} - \bar{\theta}_i \approx 1/\operatorname{dist}(Q_M, \overline{\omega}_i)$ if $\sqrt{\mu}/|\ln\mu| < \operatorname{dist}(Q_M, \overline{\omega}_i) < 1/|\ln\mu|$. When the $\overline{\omega}_i$ are "close" (less than $\sqrt{\mu}/|\ln\mu|$-distant) to the set of nonergodizing hyperplanes Q_M we choose again $\bar{\theta}_{i+1} - \bar{\theta}_i \approx |\ln\mu|$. We also estimate in (5.13) the total time $\bar{\theta}_k - \bar{\theta}_1 = \sum_{i=1}^{k} \bar{\theta}_{i+1} - \bar{\theta}_i$.

Lemma 5.2. $\forall\delta > 0$ *there exists* $\mu_6 > 0$ *such that* $\forall 0 < \mu \leqslant \mu_6$ *there exist* $\{\bar{\theta}_i\}_{i=1,\ldots,k}$ *with* $\bar{\theta}_1 = \theta_0^{\overline{\omega}_1}$ *satisfying,*

(i) *if* $\operatorname{dist}(\overline{\omega}_i, Q_M) > \sqrt{\mu}/|\ln\mu|$, *then*

$$\max\left\{C_1|\ln\mu|, \frac{2\pi}{\operatorname{dist}(\overline{\omega}_i, Q_M)}\right\} < \bar{\theta}_{i+1} - \bar{\theta}_i < 2\max\left\{C_1|\ln\mu|, \frac{2\pi}{\operatorname{dist}(\overline{\omega}_i, Q_M)}\right\}, \tag{5.10}$$

where $M = 8\pi a_{d+1}/\delta$;

(ii) *if* $\operatorname{dist}(\overline{\omega}_i, Q_M) \leqslant \sqrt{\mu}/|\ln\mu|$ *then* $C_1|\ln\mu| < \bar{\theta}_{i+1} - \bar{\theta}_i < 2C_1|\ln\mu|$, *and such that*

$$\operatorname{dist}\left(\left(\bar{\theta}_i, \overline{\varphi}_i\right), \left(\theta_0^{\overline{\omega}_i}, \varphi_0^{\overline{\omega}_i}\right) + 2\pi\mathbf{Z}^{d+1}\right) < \delta, \quad \forall i = 1, \ldots, k, \tag{5.11}$$

where $\overline{\varphi}_1, \ldots, \overline{\varphi}_k$ *are defined by* (5.9)*. Equivalently,* $\forall i = 1, \ldots, k$*, there exist* $h_i \in \mathbf{Z}^{d+1}$ *and* $\chi_i \in \mathbf{R}^{d+1}$ *such that*

$$\left(\bar{\theta}_i, \overline{\varphi}_i\right) = \left(\theta_0^{\overline{\omega}_i}, \varphi_0^{\overline{\omega}_i}\right) + 2\pi h_i + \chi_i \quad \text{with } |\chi_i| < \delta. \tag{5.12}$$

Moreover there exists a constant $K(\gamma)$ *such that*

$$\bar{\theta}_k - \bar{\theta}_1 \leqslant K(\gamma)\frac{|\ln\mu|}{\rho\mu}. \tag{5.13}$$

Proof. Let $\mu_6 > 0$ be so small that $\sqrt{\mu_6}/|\ln\mu_6| < d_0$ and $\sqrt{|\ln\mu_6|} \geqslant 32\sqrt{C_1}/(\nu\sqrt{\delta\rho})$.

Let us define $(\bar{\theta}_1, \overline{\varphi}_1) := (\theta_0^{\overline{\omega}_1}, \varphi_0^{\overline{\omega}_1})$. Assume that $(\bar{\theta}_1, \ldots, \bar{\theta}_i)$ has been defined. If $\operatorname{dist}(\overline{\omega}_i, Q_M) > \sqrt{\mu}/|\ln\mu|$ then by (5.3) there certainly exists $(\bar{\theta}_{i+1}, \overline{\varphi}_{i+1})$ satisfying (5.9), (5.10), such that

$$\operatorname{dist}\left(\left(\bar{\theta}_{i+1}, \overline{\varphi}_{i+1}\right), \left(\theta_0^{\overline{\omega}_{i+1}}, \varphi_0^{\overline{\omega}_{i+1}}\right) + 2\pi\mathbf{Z}^{d+1}\right) < \delta/4.$$

We now consider the case in which $\overline{\omega}_i$ is close to some "nonergodizing" hyperplanes of Q_M. If $\operatorname{dist}(\overline{\omega}_{i-1}, Q_M) > \sqrt{\mu}/|\ln\mu|$ and $\operatorname{dist}(\overline{\omega}_i, Q_M) \leqslant \sqrt{\mu}/|\ln\mu|$ we proceed as follows. We have $\overline{\omega}_i = \gamma(s_i)$, with $s_i \in [S_q, S_q']$ for some q, $1 \leqslant q \leqslant r$. Moreover, by property $(*)$ there exists $p^* \in \mathbf{N}$ such that $\{j \in \{1, \ldots, k\} \mid s_j \in [S_q, S_q']$ and

$\text{dist}(\overline{\omega}_j, Q_M) \leqslant \sqrt{\mu}/|\ln\mu|\} = \{i,\dots,i+p^*-1\}$, and $s_i \leqslant s_q^* \leqslant s_{i+p^*-1}$. We shall use the abbreviations s^* for s_q^*, and ω^* for ω_q^*. We claim that

$$1 \leqslant p^* \leqslant p := \left[\frac{\sqrt{\delta}}{4\sqrt{C_1\rho\mu|\ln\mu|}}\right]. \tag{5.14}$$

In fact, by (5.6) and $(*)$

$$\begin{aligned}\frac{\nu\rho}{4}\mu(p^*-1) &\leqslant \frac{\nu}{2}\big[(s_{i+p^*-1}-s^*)+(s_*-s_i)\big]\\ &\leqslant \text{dist}(\overline{\omega}_{i+p^*-1}, Q_M)+\text{dist}(\overline{\omega}_i, Q_M) \leqslant 2\frac{\sqrt{\mu}}{|\ln\mu|}.\end{aligned}$$

Hence $p^* \leqslant 8(\nu\rho\sqrt{\mu}|\ln\mu|)^{-1}$, which implies (5.14), by the choice of μ_6.

Now we can define the $\bar{\theta}_{i+1},\dots,\bar{\theta}_{i+p^*}$. The flow of $(\omega^*,1)$, as any linear flow on a torus, has the following property: there exists $T^*(\omega^*,\delta) > 0$ (abbreviated as T^*) such that any time interval of length T^* contains t satisfying $\text{dist}((t\omega^*,t), 2\pi\mathbf{Z}^{d+1}) \leqslant \delta/4$.

Therefore (provided $C_1|\ln\mu_6| > T^*$) we can define $\bar{\theta}_{i+1},\dots,\bar{\theta}_{i+p^*}$ such that

$$\begin{aligned}&C_1|\ln\mu| \leqslant \bar{\theta}_{i+j+1}-\bar{\theta}_{i+j} \leqslant 2C_1|\ln\mu|,\\ &\text{dist}\big(\big(\bar{\theta}_{i+j},\widetilde{\varphi}_{i+j}\big),\big(\bar{\theta}_i,\overline{\varphi}_i\big)+2\pi\mathbf{Z}^{d+1}\big) \leqslant \delta/4,\end{aligned} \tag{5.15}$$

where $\widetilde{\varphi}_{i+j} = \overline{\varphi}_i + \omega^*(\bar{\theta}_{i+j}-\bar{\theta}_i)$. For $1 \leqslant j \leqslant p^*$, let

$$\overline{\varphi}_{i+j} = \overline{\varphi}_i + \sum_{q=1}^{j}\overline{\omega}_{i+q-1}\big(\bar{\theta}_{i+q}-\bar{\theta}_{i+q-1}\big). \tag{5.16}$$

We now check that for all $j = 1,\dots,p^*$, $(\bar{\theta}_{i+j},\overline{\varphi}_{i+j})$, as defined in (5.15) and (5.16), satisfy estimate (5.11), namely

$$\begin{aligned}\text{dist}_T\big(\big(\bar{\theta}_{i+j},\overline{\varphi}_{i+j}\big),\big(\theta_0^{\overline{\omega}_{i+j}},\varphi_0^{\overline{\omega}_{i+j}}\big)\big) &:= \text{dist}\big(\big(\bar{\theta}_{i+j},\overline{\varphi}_{i+j}\big),\big(\theta_0^{\overline{\omega}_{i+j}},\varphi_0^{\overline{\omega}_{i+j}}\big)+2\pi\mathbf{Z}^{d+1}\big)\\ &\leqslant \delta.\end{aligned} \tag{5.17}$$

We have by (5.16) that

$$\begin{aligned}&\text{dist}_T\big(\big(\bar{\theta}_{i+j},\overline{\varphi}_{i+j}\big),\big(\bar{\theta}_i,\overline{\varphi}_i\big)\big)\\ &\quad\leqslant \text{dist}_T\big(\big(\bar{\theta}_{i+j},\widetilde{\varphi}_{i+j}\big),\big(\bar{\theta}_i,\overline{\varphi}_i\big)\big)+\left|\sum_{q=1}^{j}(\overline{\omega}_{i+q-1}-\omega^*)\big(\bar{\theta}_{i+q}-\bar{\theta}_{i+q-1}\big)\right|\\ &\quad\leqslant \delta/4+2C_1|\ln\mu|\sum_{q=1}^{p^*}\big|s_{i+q-1}-s^*\big| \quad \big(\text{by (5.15) and (a)}\big)\end{aligned}$$

$$\leqslant \delta/4 + 2C_1|\ln\mu|p^*(s_{i+p^*-1} - s_i)$$
$$\leqslant \delta/4 + 2C_1|\ln\mu|p^2\rho\mu \leqslant 3\delta/8,$$

by (5.6) and (5.14). Therefore, by (5.5),

$$\operatorname{dist}_T\big((\bar\theta_{i+j},\bar\varphi_{i+j}),\big(\theta_0^{\overline{\omega}_{i+j}},\varphi_0^{\overline{\omega}_{i+j}}\big)\big) \leqslant \frac{3\delta}{8} + \operatorname{dist}_T\big((\bar\theta_i,\bar\varphi_i),\big(\theta_0^{\overline{\omega}_i},\varphi_0^{\overline{\omega}_i}\big)\big) + K|\overline{\omega}_{i+j} - \overline{\omega}_i|$$
$$\leqslant \frac{3\delta}{8} + \frac{\delta}{4} + K\rho\mu p < \delta$$

by (5.14), provided μ_6 has been chosen small enough.

There remains to prove (5.13). By $(*)$ we can write

$$A_m := \left\{ s \in [S_m, S'_m] \;\middle|\; \frac{\sqrt{\mu}}{|\ln\mu|} \leqslant \operatorname{dist}(\gamma(s), Q_M) \leqslant \frac{1}{2C_1|\ln\mu|} \right\} = [U_m, V_m] \cup [V'_m, U'_m],$$

with $S_m < U_m < V_m < s_m^* < V'_m < U'_m < S'_m$ (in the case when $\omega^* = \omega_{I,F}$, A_m is just an interval). Moreover, by (a), $s_m^* - V_m, V'_m - s_m^* \geqslant \sqrt{\mu}/|\ln\mu|$. Define $A := \bigcup_{m=1}^r A_m$. We have $\bar\theta_k - \bar\theta_1 = \sigma_0 + \sum_{m=1}^r \sigma_m$, where

$$\sigma_0 := \sum_{1\leqslant i\leqslant k-1, s_i\notin A} (\bar\theta_{i+1} - \bar\theta_i), \qquad \sigma_m := \sum_{1\leqslant i\leqslant k-1, s_i\in A_m} (\bar\theta_{i+1} - \bar\theta_i).$$

For $s_i \notin A$, $\bar\theta_{i+1} - \bar\theta_i \leqslant 2C_1|\ln\mu|$, hence $\sigma_0 \leqslant 2C_1k|\ln\mu| \leqslant 4C_1L\ln\mu/(\rho\mu)$. For $i \in A_m$, $\bar\theta_{i+1} - \bar\theta_i \leqslant 4\pi(\operatorname{dist}(\overline{\omega}_i, Q_M))^{-1} \leqslant 8\pi/(\nu|s_i - s_m^*|)$ by $(*)$, and hence, using that by (5.6) $s_{i+1} \geqslant s_i + \rho\mu/2$,

$$\sigma_m \leqslant \frac{8\pi}{\nu} \sum_{1\leqslant i\leqslant k-1, s_i\in A_m} \frac{1}{|s_i - s_m^*|} \leqslant \frac{16\pi}{\nu\rho\mu} \sum_{1\leqslant i\leqslant k-1, s_i\in A_m} \frac{s_{i+1} - s_i}{|s_i - s_m^*|}.$$

Estimating the above sum with an integral we easily get:

$$\sigma_m \leqslant \frac{8\pi}{\nu(s_m^* - V_m)} + \frac{16\pi}{\nu\rho\mu}\int_{U_m}^{V_m} \frac{\mathrm{d}s}{s_m^* - s} + \frac{8\pi}{\nu(V'_m - s_m^*)} + \frac{16\pi}{\nu\rho\mu}\int_{V'_m}^{U'_m} \frac{\mathrm{d}s}{s - s_m^*};$$

(5.13) can be easily deduced by the bound on $s_m^* - V_m, V'_m - s_m^*$. $\square$

In the next section we will prove the existence of a diffusion orbit (φ_μ, q_μ) close to the "unperturbed pseudo-diffusion orbit" $(\overline{\varphi}(t), \bar q(t)) : (\bar\theta_1, \bar\theta_k) \to \mathbf{R}^{d+1}$ defined, for $t \in [\bar\theta_i, \bar\theta_{i+1}]$, as $\overline{\varphi}(t) := \overline{\varphi}_i + \bar\omega_i(t - \bar\theta_i)$ and $\bar q_{|[\theta_i,\theta_{i+1}]} := Q_{\bar\theta_{i+1}-\bar\theta_i}(\cdot - \bar\theta_i) \pmod{2\pi}$.

6. The diffusion orbit

We need the following property of the Melnikov function $\widetilde{\Gamma}(\omega,\cdot,\cdot)$ defined w.r.t. to the variables (b,c) by

$$\widetilde{\Gamma}(\omega,b,c) := \Gamma\big(\omega, \theta_0^\omega + b, \varphi_0^\omega + b\omega + c\big).$$

Lemma 6.1. *Assume that $\Gamma(\omega,\cdot,\cdot)$ possesses a nondegenerate local minimum in $(\theta_0^\omega,\varphi_0^\omega)$. Then there exist $r>0$, $\bar b>0$, $\nu_j>0$ $(j=1,2)$ depending only on γ such that $\forall \omega=\gamma(s)$, $s\in[0,L]$*

(i) $\partial_c\widetilde{\Gamma}(\omega,b,c)\cdot c \geqslant \nu_2>0$ ***or*** $|\partial_b\widetilde{\Gamma}(\omega,b,c)|\geqslant \nu_1>0$ *for* $|c|=r, |b|\leqslant \bar b$,
(ii) $\partial_b\widetilde{\Gamma}(\omega,b,c)\times \operatorname{sign}(b)\geqslant \nu_1>0$ *for* $|c|\leqslant r$ *and* $b=\pm\bar b$.

Proof. We can assume that (5.4) is satisfied. Since $\Gamma(\omega,\cdot,\cdot)$ possesses a nondegenerate minimum in $(\theta_0^\omega,\varphi_0^\omega)$, $\widetilde{\Gamma}(\omega,b,c)$ possesses in $(0,0)$ a nondegenerate minimum. Hence we write $\widetilde{\Gamma}(\omega,b,c)$, up to a constant, as $\widetilde{\Gamma}(\omega,b,c)=Q_2(b,c)+Q_3(b,c)$ where $Q_2(b,c)=: \beta_\omega b^2/2+(\alpha_\omega\cdot c)b+(\gamma_\omega c\cdot c)/2$ is a positive definite quadratic form ($\beta_\omega\in\mathbf{R}, \alpha_\omega\in\mathbf{R}^d$, $\gamma_\omega\in \mathrm{Mat}(d\times d)$) and $Q_3=\mathrm{O}(|b|^3+|c|^3)$. More precisely, by (5.4), there exists $\varepsilon>0$ such that $\beta_\omega>\varepsilon$, and $d_\omega(c):=\beta_\omega(\gamma_\omega c\cdot c)-(\alpha_\omega\cdot c)^2>\varepsilon|c|^2$ for all $\omega\in\gamma([0,L])$. In addition, by the smoothness of Γ and the fact that $\omega=\gamma(s)$ lives in a compact subset of $\mathbf{R}^d$, there exists a constant M such that, $\forall\omega\in\gamma([0,L])$, $|\alpha_\omega|+|\beta_\omega|+|\gamma_\omega|\leqslant M$, $|\nabla Q_3(b,c)|\leqslant M(b^2+|c|^2)$.

We have $\partial_b Q_2(b,c)=\beta_\omega b+\alpha_\omega\cdot c$ and $\partial_c Q_2(b,c)\cdot c=b\alpha_\omega\cdot c+(\gamma_\omega c\cdot c)$.

Let us define $\bar\nu_1:=\inf_{\omega\in\gamma([0,L])}\varepsilon/(4|\alpha_\omega|)>0$ and $\bar\nu_2:=\inf_{\omega\in\gamma([0,L])}\varepsilon/(4\beta_\omega)>0$. Then consider $\nu_1:=\bar\nu_1 r$, $\nu_2=\bar\nu_2 r^2$ and $\bar b:=r\sup_{\omega\in\gamma([0,L])}(3\bar\nu_1+|\alpha_\omega|)/\beta_\omega$, $r\in(0,1]$. We now prove that, provided $r>0$ has been chosen sufficiently small, conditions (i) and (ii) are satisfied with the above choice of the constants. Indeed if $(|\alpha_\omega\cdot c|+2\bar\nu_1 r)/\beta_\omega\leqslant |b|\leqslant\bar b$ and $|c|\leqslant r$ then $\partial_b\widetilde{\Gamma}(\omega,b,c)\cdot\operatorname{sign}(b)\geqslant \beta_\omega|b|-|\alpha_\omega\cdot c|-|\partial_b Q_3(b,c)|\geqslant 2\bar\nu_1 r-\mathrm{O}(r^2)\geqslant\nu_1$ for r sufficiently small. In particular this proves (ii). On the other hand, if $|b|<(|\alpha_\omega\cdot c|+2\bar\nu_1 r)/\beta_\omega$ and $|c|=r$, then

$$\begin{aligned}
\partial_c\widetilde{\Gamma}(\omega,b,c)\cdot c &= b(\alpha_\omega\cdot c)+(\gamma_\omega c\cdot c)+\partial_c Q_3(b,c)\cdot c \geqslant (\gamma_\omega c\cdot c)-\big|b(\alpha_\omega\cdot c)\big|+\mathrm{O}\big(r^3\big)\\
&\geqslant \frac{\varepsilon r^2+(\alpha_\omega\cdot c)^2-|\alpha_\omega\cdot c|(|\alpha_\omega\cdot c|+2\bar\nu_1 r)}{\beta_\omega}+\mathrm{O}\big(r^3\big)\\
&\geqslant \frac{\varepsilon-2\bar\nu_1|\alpha_\omega|}{\beta_\omega}r^2+\mathrm{O}\big(r^3\big)\geqslant\frac{\varepsilon}{2\beta_\omega}r^2-\mathrm{O}\big(r^3\big)\geqslant 2\bar\nu_2 r^2+\mathrm{O}\big(r^3\big).
\end{aligned}$$

Hence (i) is satisfied for r small enough. □

The partial derivatives of $\widetilde{\Gamma}$ are Lipschitz-continuous w.r.t. (b,c) uniformly in $\omega\in\gamma([0,L])$. Therefore, by Lemma 6.1, there exists $\delta>0$ such that, $\forall\eta\in\mathbf{R}$ with $|\eta|\leqslant\delta$, $\forall\xi\in\mathbf{R}^d$ with $|\xi|\leqslant\delta$, $\forall\omega\in\gamma([0,L])$,

$$\partial_c \widetilde{\Gamma}(\omega, b+\eta, c+\xi)\cdot c \geqslant 3\nu_2/4 > 0 \quad \text{or}$$

$$\left|\partial_b \widetilde{\Gamma}(\omega, b+\eta, c+\xi)\right| \geqslant 3\nu_1/4 > 0 \quad \text{for } |c| = r,\ |b| \leqslant \bar{b}, \tag{6.1}$$

$$\partial_b \widetilde{\Gamma}(\omega, b+\eta, c+\xi) \times \operatorname{sign}(b) \geqslant 3\nu_1/4 > 0 \quad \text{for } |c| \leqslant r \text{ and } b = \pm\bar{b}. \tag{6.2}$$

Moreover let us fix $\rho > 0$ such that

$$\rho \leqslant \min\{\nu_1/2, \nu_2/r\}/(6C_2), \tag{6.3}$$

where C_2 appears in (3.24). These are the positive constants (δ, ρ) that we use in order to define, for $0 < \mu < \mu_6$, $\overline{\omega}_i$, $\bar{\theta}_i$, $\overline{\varphi}_i$ by Lemma 5.2.

Since $\gamma([0, L])$ is a compact subset of $\mathcal{D}_N^c$, $\inf_{s\in[0,L]} \beta(\gamma(s)) > 0$ and, by the choice of $\bar{\theta}_i$, for μ small enough (3.21) is satisfied. Therefore, by Lemma 3.5 and (5.12), there exists $\mu_7 > 0$ such that, $\forall 0 < \mu \leqslant \mu_7$,

$$\mathcal{F}_\mu(b, c) = \frac{1}{2}\sum_{i=1}^{k-1} \frac{|c_{i+1} - c_i|^2}{\Delta\bar{\theta}_i + (b_{i+1} - b_i)} + \mu \sum_{i=1}^{k} \widetilde{\Gamma}(\overline{\omega}_i, \eta_i + b_i, \xi_i + c_i) + R_7, \tag{6.4}$$

where $|\eta_i| \leqslant \delta$, $|\xi_i| \leqslant \delta$, R_7 is given by (3.23) and satisfies (3.24).

We minimize the functional $\mathcal{F}_\mu$ on the closure of

$$W := \left\{(b, c) := (b_1, c_1, \ldots, b_k, c_k) \in \mathbf{R}^{(d+1)k} \,\middle|\, |b_i| < \bar{b},\ |c_i| < r,\ \forall i = 1, \ldots, k\right\}.$$

Since $\overline{W}$ is compact, $\mathcal{F}_\mu$ attains its minimum in $\overline{W}$, say at $(\tilde{b}, \tilde{c})$. By Lemma 2.3 the existence of the diffusion orbit will be proved once we show that $(\tilde{b}, \tilde{c}) \in W$, see Lemma 6.3. Let us define for $i = 1, \ldots, k-1$

$$w_i := w_i(b, c) := \frac{c_{i+1} - c_i}{\theta_{i+1} - \theta_i} = \frac{c_{i+1} - c_i}{\Delta\bar{\theta}_i + (b_{i+1} - b_i)},$$

and $w_0 = w_k = 0$. From (5.9) and (3.14), w_i can be written as

$$w_i = \frac{\varphi_{i+1} - \varphi_i}{(\theta_{i+1} - \theta_i)} - \overline{\omega}_i - \frac{\Delta\overline{\omega}_i b_{i+1}}{(\theta_{i+1} - \theta_i)} = (\omega_i - \overline{\omega}_i) + \mathrm{O}\left(\frac{\mu}{|\ln\mu|}\right). \tag{6.5}$$

By the expression of $\mathcal{F}_\mu$ in (6.4) we have, for all $i = 1, \ldots, k$,

$$\partial_{c_i}\mathcal{F}_\mu(b, c) = w_{i-1} - w_i + \mu\partial_c\widetilde{\Gamma}(\overline{\omega}_i, \eta_i + b_i, \xi_i + c_i) + R_i, \tag{6.6}$$

$$\partial_{b_i}\mathcal{F}_\mu(b, c) = \frac{1}{2}\left(|w_i|^2 - |w_{i-1}|^2\right) + \mu\partial_b\widetilde{\Gamma}(\overline{\omega}_i, \eta_i + b_i, \xi_i + c_i) + S_i, \tag{6.7}$$

where $R_i := \partial_{c_i} R_7$, $S_i := \partial_{b_i} R_7$ satisfy, by (3.24) and (6.3)

$$|R_i|, |S_i| \leqslant \frac{\mu}{2}\min\left\{\frac{\nu_1}{2}, \frac{\nu_2}{r}\right\}. \tag{6.8}$$

By (6.6)–(6.7), a way to see critical points of $\mathcal{F}_\mu$ is to show that the terms $w_{i-1} - w_i$ and $|w_i|^2 - |w_{i-1}|^2$ are small w.r.t the $O(\mu)$-contribution provided by the Melnikov function. By (3.8) $|\omega_i - \overline{\omega}_i| = O(1/(\theta_{i+1} - \theta_i))$ and hence, using (6.5), an estimate for each w_i separately is given by $w_i = O(1/|\bar{\theta}_{i+1} - \bar{\theta}_i|) + O(\mu/|\ln\mu|)$. Hence each $|w_i|$ is $O(\mu)$-small if the time to make a transition $|\bar{\theta}_{i+1} - \bar{\theta}_i| = O(1/\mu)$, as in [7]. These time intervals are too large to obtain the approximation for the reduced action functional $\mathcal{F}_\mu$ given in Lemma 3.5 and (6.4). Therefore we need more refined estimates: the proof of Theorem 1.1 (and Theorem 1.3) relies on the following crucial property for $\widetilde{w}_i := w_i(\tilde{b}, \tilde{c})$, satisfied by the minimum point $(\tilde{b}, \tilde{c})$.

Lemma 6.2. *We have* (*for* $i = 1, \dots, k$)

$$\text{(i)} \quad |\widetilde{w}_i - \widetilde{w}_{i-1}| = O(\mu), \qquad \text{(ii)} \quad |\widetilde{w}_i| = O\left(\frac{\sqrt{\mu}}{\sqrt{|\ln\mu|}}\right). \tag{6.9}$$

Proof. Estimate (6.9)(i) is a straightforward consequence of (6.6) and (6.8) if $|\tilde{c}_i| < r$, since in this case $\partial_{c_i}\mathcal{F}_\mu(\tilde{b}, \tilde{c}) = 0$. We now prove that (6.9)(i) holds also if $|\tilde{c}_i| = r$ for some i. Indeed if $|\tilde{c}_i| = r$ then

$$\partial_{c_i}\mathcal{F}_\mu(\tilde{b}, \tilde{c}) = \alpha_\mu \tilde{c}_i, \quad \text{for some } \alpha_\mu \leqslant 0, \tag{6.10}$$

(since $(\tilde{b}, \tilde{c})$ is a minimum point) and then by (6.6), (6.10) and (6.8) we deduce:

$$\widetilde{w}_{i-1} - \widetilde{w}_i = \alpha_\mu \tilde{c}_i + O(\mu). \tag{6.11}$$

Let us decompose $\widetilde{w}_{i-1}$ and $\widetilde{w}_i$ in the "radial" and "tangent" directions to the ball $S_i = \{|b_i| \leqslant \bar{b},\ |c_i| \leqslant r\}$:

$$\widetilde{w}_{i-1} = a_i \tilde{c}_i + u_i, \quad \text{with } u_i \cdot \tilde{c}_i = 0, \tag{6.12}$$

$$-\widetilde{w}_i = a_i' \tilde{c}_i + u_i', \quad \text{with } u_i' \cdot \tilde{c}_i = 0. \tag{6.13}$$

Since $|\tilde{c}_{i-1}| \leqslant |\tilde{c}_i| = r$, $|\tilde{c}_{i+1}| \leqslant |\tilde{c}_i| = r$, there results that

$$a_i r^2 = \widetilde{w}_{i-1} \cdot \tilde{c}_i \geqslant 0 \quad \text{and} \quad a_i' r^2 = -\widetilde{w}_i \cdot \tilde{c}_i \geqslant 0, \tag{6.14}$$

so that $a_i, a_i' \geqslant 0$. Summing (6.12) and (6.13) and using (6.11) we obtain:

$$\left(a_i + a_i'\right)\tilde{c}_i + \left(u_i + u_i'\right) = O(\mu) + \alpha_\mu \tilde{c}_i,$$

with $a_i, a_i', -\alpha_\mu \geqslant 0$. This implies that $\alpha_\mu = O(\mu/r)$ and from Eq. (6.11) we get (6.9)(i).

We can now prove (6.9)(ii). Let $i_0 \in \{1, \dots, k-1\}$ be such that $\forall 1 \leqslant i \leqslant k-1$, $|\widetilde{w}_{i_0}| \geqslant |\widetilde{w}_i|$. For $j \in \{1, \dots, k-1\}$, $j \neq i_0$ we can write $\widetilde{w}_j = \widetilde{w}_{i_0} + s_j$ with $s_j = \sum_{i=i_0}^{j-1}(\widetilde{w}_{i+1} - \widetilde{w}_i)$ and hence, by (6.9)(i)

$$|s_j| \leqslant \sum_{i=i_0}^{j-1} |\widetilde{w}_{i+1} - \widetilde{w}_i| \leqslant C\mu|j - i_0| \tag{6.15}$$

for some constant $C > 0$. Hence

$$\tilde{c}_j - \tilde{c}_{i_0} = \sum_{i=i_0}^{j-1} \widetilde{w}_i\big(\tilde{\theta}_{i+1} - \tilde{\theta}_i\big) = \widetilde{w}_{i_0}\big(\tilde{\theta}_j - \tilde{\theta}_{i_0}\big) + \sum_{i=i_0}^{j-1} s_i\big(\tilde{\theta}_{i+1} - \tilde{\theta}_i\big) \tag{6.16}$$

and then by (6.15)

$$\begin{aligned}|\tilde{c}_j - \tilde{c}_{i_0}| &\geqslant \big|\widetilde{w}_{i_0}\big|\big|\tilde{\theta}_j - \tilde{\theta}_{i_0}\big| - C\mu|j - i_0|\big|\tilde{\theta}_j - \tilde{\theta}_{i_0}\big| \\ &= \big(|\widetilde{w}_{i_0}| - C\mu|j - i_0|\big)\big|\tilde{\theta}_j - \tilde{\theta}_{i_0}\big|.\end{aligned} \tag{6.17}$$

Since $|\tilde{\theta}_{i+1} - \tilde{\theta}_i| > C_1|\ln\mu| + \mathrm{O}(1)$ (by (3.4)), $\forall i = 1, \ldots, k-1$, $|\tilde{\theta}_j - \tilde{\theta}_{i_0}| > C_1|j - i_0| \cdot |\ln\mu|$. Take $\bar{j} \in \{1, \ldots, k-1\}$ such that $|\bar{j} - i_0| = [(\sqrt{\mu}\sqrt{|\ln\mu|})^{-1}] + 1$ (such a $\bar{j}$ certainly exists since, by (5.7), $k \approx 1/\mu$ for μ small). Then we obtain, using that $|\tilde{c}_i| \leqslant r$ for all $i = 1, \ldots, k$,

$$2r \geqslant |\tilde{c}_j - \tilde{c}_{i_0}| \geqslant \left(|\widetilde{w}_{i_0}| - C\frac{\sqrt{\mu}}{\sqrt{|\ln\mu|}} - C\mu\right)C_1\frac{\sqrt{|\ln\mu|}}{\sqrt{\mu}},$$

i.e., $|\widetilde{w}_{i_0}| \leqslant (2r + CC_1)\sqrt{\mu}/(C_1\sqrt{|\ln\mu|}) + C\mu$. We have thus proved the important property (6.9)(ii). □

Remark 6.1. By (6.5), $(\widetilde{\omega}_i - \overline{\omega}_i) = \widetilde{w}_i + \mathrm{O}(\mu/|\ln\mu|)$, so that, by (5.8), (6.9) implies

$$|\widetilde{\omega}_i - \overline{\omega}_i| = \mathrm{O}\left(\frac{\sqrt{\mu}}{\sqrt{|\ln\mu|}}\right), \qquad |\widetilde{\omega}_{i+1} - \widetilde{\omega}_i| = \mathrm{O}(\mu). \tag{6.18}$$

Note that, from (3.8), we would just obtain $|\widetilde{\omega}_i - \overline{\omega}_i| = \mathrm{O}(1/|\ln\mu|)$. (6.18) can be seen as an a priori estimate satisfied by the minimum point $(\tilde{\theta}, \widetilde{\varphi})$.

The following lemma proves the existence of a local minimum of the reduced action functional in the interior of W and hence of a true diffusion orbit.

Lemma 6.3. *Let $(\tilde{b}, \tilde{c})$ be a minimum point of $\mathcal{F}_\mu$ over $\overline{W}$. Then $(\tilde{b}, \tilde{c}) \in W$, namely*

$$|\tilde{c}_i| < r \quad \textit{for all } i \in \{1, \ldots, k\} \tag{6.19}$$

and

$$|\tilde{b}_i| < \bar{b} \quad \textit{for all } i \in \{1, \ldots, k\}. \tag{6.20}$$

Proof. By (6.9) we have $||\widetilde{w}_{i+1}|^2 - |\widetilde{w}_i|^2| \leqslant |\widetilde{w}_{i+1} - \widetilde{w}_i| \cdot (|\widetilde{w}_{i+1}| + |\widetilde{w}_i|) = \mathrm{O}(\mu^{3/2})$, and hence, from (6.7) we derive:

$$\partial_{b_i} \mathcal{F}_\mu(\tilde{b}, \tilde{c}) = \mu \partial_b \widetilde{\Gamma}\big(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i\big) + \mathrm{O}\big(\mu^{3/2}\big) + S_i. \tag{6.21}$$

Let us first assume by contradiction that $\exists i$ such that $|\tilde{c}_i| = r$ and $|\tilde{b}_i| < \bar{b}$. In this case we claim that

$$\partial_c \widetilde{\Gamma}\big(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i\big) \cdot \tilde{c}_i \leqslant \nu_2/2 \quad \text{and} \quad \big|\partial_b \widetilde{\Gamma}\big(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i\big)\big| \leqslant \nu_1/2 \tag{6.22}$$

contradicting (6.1), since $|\eta_i|, |\xi_i| \leqslant \delta$. Let us prove (6.22). Since $(\tilde{b}, \tilde{c})$ is a minimum point

$$\begin{aligned}\partial_{c_i} \mathcal{F}_\mu\big(\tilde{b}, \tilde{c}\big) \cdot \tilde{c}_i &= (\widetilde{w}_{i-1} - \widetilde{w}_i) \cdot \tilde{c}_i + \mu \partial_c \widetilde{\Gamma}\big(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i\big) \cdot \tilde{c}_i + R_i \cdot \tilde{c}_i \\ &= \alpha_\mu \tilde{c}_i \cdot \tilde{c}_i = \alpha_\mu r^2 \leqslant 0.\end{aligned}$$

By (6.14) and (6.8) it follows that $\partial_c \widetilde{\Gamma}(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i) \cdot \tilde{c}_i \leqslant \nu_2/2$. Moreover since $|\tilde{b}_i| < \bar{b}$ we have $\partial_{b_i} \mathcal{F}_\mu(\tilde{b}, \tilde{c}) = 0$, and by (6.21), (6.8) it follows that $|\partial_b \widetilde{\Gamma}(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i)| \leqslant \nu_1/2$ (provided μ is small enough). Estimate (6.22) is then proved. As a result, if (6.20) holds, so does (6.19).

Let us finally prove (6.20). If by contradiction $\exists i$ with $|\tilde{b}_i| = \bar{b}$, by (6.21), (6.8) and since $(\tilde{b}, \tilde{c})$ is a minimum point, arguing as before, we deduce that $\partial_b \widetilde{\Gamma}(\overline{\omega}_i, \eta_i + \tilde{b}_i, \xi_i + \tilde{c}_i) \times \operatorname{sign}(\tilde{b}_i) \leqslant \nu_1/2$. This contradicts (6.2) since $|\eta_i|, |\xi_i| \leqslant \delta$. □

Proof of Theorem 1.1. Lemmas 6.3 and 2.3 imply the existence of a diffusion orbit

$$z_\mu(t) := \big(\varphi_\mu(t), q_\mu(t), I_\mu(t), p_\mu(t)\big)$$

with $\dot{\varphi}_\mu(\tilde{\theta}_1) = \omega_I + \mathrm{O}(\mu)$ and $\dot{\varphi}_\mu(\tilde{\theta}_k) = \omega_I + \mathrm{O}(\mu)$ ($z_\mu(\cdot)$ connects a $\mathrm{O}(\mu)$-neighborhood of $\mathcal{T}_{\omega_I}$ to a $\mathrm{O}(\mu)$-neighborhood of $\mathcal{T}_{\omega_F}$ in the time-interval (τ_1, τ_2) where $\tau_1 := (\tilde{\theta}_1 + \tilde{\theta}_2)/2$, $\tau_2 := (\tilde{\theta}_{k-1} + \tilde{\theta}_k)/2$). The estimate on the diffusion time is a straightforward consequence of (5.13) and the fact that $\tilde{\theta}_{1,k} = \bar{\theta}_{1,k} + \mathrm{O}(1)$. That $\operatorname{dist}(I_\mu(t), \gamma([0, L])) < \eta$ for all t, provided μ is small enough, results from (6.18) and the estimates of Lemma 2.1.

Finally we observe that, if the perturbation is $\mu(f + \mu \tilde{f})$, then Lemma 2.1 still applies with the same estimates. Moreover in the development of the reduced functional the term containing $\mu^2 \tilde{f}$ gives, in time intervals $\bar{\theta}_{i+1} - \bar{\theta}_i \leqslant const.|\ln \mu|/\sqrt{\mu}$, negligible contributions $\mathrm{o}(\mu)$. Therefore the same variational proof applies. □

Proof of Theorem 1.3. If the perturbation is of the form $f(\varphi, q, t) = (1 - \cos q) f(\varphi, t)$, by Remark 2.1(2), we can prove that the development (3.22) holds along any path γ of the action space (without any condition as (3.21)). Therefore the previous variational argument applies. □

For $\beta > 0$ small let $\mathcal{D}_N^\beta$ be the set of frequencies "β-nonresonant with the perturbation" $\mathcal{D}_N^\beta := \{\omega \in \mathbf{R}^d \mid |\omega \cdot n + l| > \beta,\ \forall 0 < |(n, l)| \leqslant N\}$. If β becomes small with μ our estimate on the diffusion time required to approach to the boundaries of $\mathcal{C} \cap \mathcal{D}_N^{\beta}$ slightly deteriorates. In the same hypotheses as in Theorem 1.1 we have the following result.

Theorem 6.1. $\forall R > 0$, $\forall 0 \leqslant a < 1/4$, *there exists* $\mu_8 > 0$ *such that* $\forall 0 < \mu \leqslant \mu_8$, $\forall \omega_I, \omega_F \in \mathcal{C} \cap \mathcal{D}_N^{\mu^a} \cap B_R(0)$ *there exist a diffusion orbit* $(\varphi_\mu(t), q_\mu(t), I_\mu(t), p_\mu(t))$ *of* $(\mathcal{S}_\mu)$ *and two instants* $\tau_1 < \tau_2$ *with* $I_\mu(\tau_1) = \omega_I + \mathrm{O}(\mu)$, $I_\mu(\tau_2) = \omega_F + \mathrm{O}(\mu)$ *and*

$$|\tau_2 - \tau_1| = \mathrm{O}\big(1/\mu^{1+a}\big). \tag{6.23}$$

Proof. For simplicity we consider the case in which $\beta(\omega_I) = \mathrm{O}(\mu^a)$ and $\beta(\omega_F) = \mathrm{O}(1)$. With respect to Theorem 1.1 we only need to prove the existence of a diffusion orbit connecting ω_I to some fixed ω^* lying in the same connected component of $\mathcal{D}_N^c \cap B_R(0)$ containing ω_I. In order to construct an orbit connecting ω_I to ω^* we can define $\overline{\omega}_i := \omega_I + i(\omega^* - \omega_I)/k$, for $0 \leqslant i \leqslant k$ and $k := [|\omega^* - \omega_I|/\rho\mu] + 1$. We obtain that $\beta_j = \beta(\overline{\omega}_j) \geqslant C(\mu^a + j\rho\mu)$ for some $C > 0$ and we choose $\bar{\theta}_{j+1} - \bar{\theta}_j \geqslant \mathit{const}.\beta_j^{-2}$ verifying in this way the hypotheses of Lemma 3.5. If ω_I belongs to some Q_M the transition times $|\ln \mu|/\sqrt{\mu}$ needed to cross Q_M (see Lemma 5.2) still satisfy (3.21). We finally obtain a diffusion time $\bar{\theta}_k - \bar{\theta}_1 = \sum_{j=1}^{k-1} (\bar{\theta}_{j+1} - \bar{\theta}_j) = \mathrm{O}(1/\mu^{1+a})$. □

7. The stability result and the optimal time

In this section we will prove, via classical perturbation theory, stability results for the action variables, implying, in particular, Theorem 1.2. We shall use the following notations: for $l \in \mathbf{N}$, $A \subset \mathbf{C}^l$ and $r > 0$, we define $A_r := \{z \in \mathbf{C}^l \mid \mathrm{dist}(z, A) \leqslant r\}$ and $\mathbf{T}_s^l := \{z \in \mathbf{C}^l \mid |\mathrm{Im}\, z_j| < s,\ \forall 1 \leqslant j \leqslant l\}$ (thought of as a complex neighborhood of $\mathbf{T}^l$). Given two bounded open sets $B \subset \mathbf{C}^2$, $D \subset \mathbf{C}^l$ and $f(I, \varphi, p, q)$, real analytic function with holomorphic extension on $D_\sigma \times \mathbf{T}_{s+\sigma}^l \times B_\sigma$ for some $\sigma > 0$, we define the following norm

$$\|f\|_{B,D,s} = \sum_{k \in Z^l} \sup_{\substack{(p,q) \in B \\ I \in D}} \big|\hat{f}_k(I, p, q)\big| \mathrm{e}^{|k|s},$$

where $\hat{f}_k(I, p, q)$ denotes the k-Fourier coefficient of the periodic function $\varphi \to f(I, \varphi, p, q)$.

Let us consider Hamiltonian $\mathcal{H}_\mu$ defined in (1.1) and assume that $f(I, \varphi, p, q, t)$, defined in (1.2), is a real analytic function, possessing, for some $r, \bar{r}, \tilde{r}, s > 0$, complex analytic extension on $\{I \in \mathbf{R}^d \mid |I| \leqslant \bar{r}\}_r \times \mathbf{T}_s^d \times \{p \in \mathbf{R} \mid |p| \leqslant \tilde{r}\}_r \times \mathbf{T}_s \times \mathbf{T}_s$.

It is convenient to write Hamiltonian $\mathcal{H}_\mu$ in autonomous form. For this purpose let us introduce the new action-angle variables (I_0, φ_0) with $t = \varphi_0$, that will still be denoted by $I := (I_0, I_1, \ldots, I_n)$ and $\varphi := (\varphi_0, \varphi_1, \ldots, \varphi_n)$. Defining $h(I) := I_0 + |I|^2/2$ and $E := E(p, q) := p^2/2 + (\cos q - 1)$, $\mathcal{H}_\mu$ is then equivalent to the autonomous Hamiltonian,

$$H := H(I, \varphi, p, q) := h(I) + E(p, q) + \mu f(I, \varphi, p, q). \tag{7.1}$$

Clearly, Hamiltonian H is a real analytic function, with complex analytic extension on

$$\left\{ I \in \mathbf{R}^{d+1} \,\middle|\, |I| \leqslant \bar{r} \right\}_r \times \mathbf{T}_s^{d+1} \times \left\{ p \in \mathbf{R} \,\middle|\, |p| \leqslant \tilde{r} \right\}_r \times \mathbf{T}_s.$$

In the sequel we will denote by $z(t) := (I(t), \varphi(t), p(t), q(t))$ the solution of the Hamilton equations associated to Hamiltonian (7.1) with initial condition $z(0) = (I(0), \varphi(0), p(0), q(0))$.

The proof of the stability of the action variables is divided in two steps:

(i) (*Stability far from the separatrices of the pendulum*:) prove stability in the region:

$$\mathcal{E}_1 := \mathcal{E}_1^+ \cup \mathcal{E}_1^- := \left\{ (I, \varphi, p, q) \,\middle|\, E(p,q) \geqslant \mu^{c_d} \right\} \\ \cup \left\{ (I, \varphi, p, q) \,\middle|\, -2 + \mu^{c_d} \leqslant E(p,q) \leqslant -\mu^{c_d} \right\}$$

in which we can apply the Nekhoroshev Theorem obtaining actually stability for exponentially long times,

(ii) (*Stability close to the separatrices of the pendulum and to the elliptic equilibrium point*:) prove stability in the region:

$$\mathcal{E}_2 := \mathcal{E}_2^+ \cup \mathcal{E}_2^- := \left\{ (I, \varphi, p, q) \,\middle|\, -2\mu^{c_d} \leqslant E(p,q) \leqslant 2\mu^{c_d} \right\} \\ \cup \left\{ (I, \varphi, p, q) \,\middle|\, -2 \leqslant E(p,q) \leqslant -2 + 2\mu^{c_d} \right\}$$

in which we use some *ad hoc* arguments,

where $0 < c_d < 1$ is a positive constant that will be chosen later on, see (7.12).

We first prove (i). In the regions[6] $\widetilde{\mathcal{E}}_1^{\pm} := \Pi_{q,p} \mathcal{E}_1^{\pm}$ we first write the pendulum Hamiltonian $E(p,q)$ in action-angle variables. In the region[7] $\widetilde{\mathcal{E}}_1^+ \cup \{p > 0\}$ the new action variable P is defined by the formula

$$P := P^+(E) := \frac{\sqrt{2}}{\pi} \int_0^{\pi} \sqrt{E + (1 + \cos\psi)}\, \mathrm{d}\psi,$$

while in the region $\widetilde{\mathcal{E}}_1^-$ the new action variable is

$$P := P^-(E) = \frac{2\sqrt{2}}{\pi} \int_0^{\psi_0(E)} \sqrt{E + (1 + \cos\psi)}\, \mathrm{d}\psi,$$

[6] $\Pi_{p,q}$ denotes the projection onto the (p,q) variables.

[7] The case with $p < 0$ is completely analogous.

where $\psi_0(E)$ is the first positive number such that $E + (1 + \cos\psi_0(E)) = 0$. We will use the following lemma, proved in [10], regarding the analyticity radii of these action-angle variables close to the separatrices of the pendulum.

Lemma 7.1. *There exist intervals $D^\pm \subset \mathbf{R}$, symplectic transformations $\phi^\pm = \phi^\pm(P, Q)$ real analytic on $D^\pm \times \mathbf{T}$ with holomorphic extension on $D^\pm_{r_0} \times \mathbf{T}_{s_0}$ and functions $E^\pm$ real analytic on $D^\pm$ with holomorphic extension on $D^\pm_{r_0}$ such that $\phi^\pm(D^\pm \times \mathbf{T}) = \widetilde{\mathcal{E}}^\pm_1$ and*

$$E\big(\phi^\pm(P, Q)\big) = E^\pm(P),$$

with $r_0 = const\,\mu^{c_d}$ and $s_0 = const/|\ln\mu|$. Moreover, for E bounded, the following estimates on the derivatives hold [8]

$$\frac{\mathrm{d}E^\pm}{\mathrm{d}P}(P^\pm(E)) \approx \ln^{-1}\left(1 + \frac{1}{\sqrt{|E|}}\right), \tag{7.2}$$

$$\pm\frac{\mathrm{d}^2E^\pm}{\mathrm{d}P^2}(P^\pm(E)) \approx \frac{1}{|E|}\ln^{-3}\left(1 + \frac{1}{\sqrt{|E|}}\right). \tag{7.3}$$

After this change of variables Hamiltonian H becomes

$$\begin{aligned} H^\pm &:= H^\pm(I, \varphi, P, Q) := h^\pm(I, P) + \mu f^\pm(I, \varphi, P, Q) \\ &:= h(I) + E^\pm(P) + \mu f^\pm(I, \varphi, P, Q), \end{aligned}$$

where $f^\pm(I, \varphi, P, Q) := f(I, \varphi, \phi^\pm(P, Q))$.

7.1. *Stability in the region $\mathcal{E}^+_1$*

In the region $\mathcal{E}^+_1$, the proof of the stability of the actions variables follows by a straightforward application of the Nekhoroshev Theorem as proved in Theorem 1 of [19]. In order to apply such theorem we need some definitions. For $l, m > 0$, a function $h := h(J)$ is said to be l, m*-quasi-convex* on $A \subset \mathbf{R}^{d+1}$, if at every point $J \in A$ at least one of the inequalities

$$\big|\big\langle h'(J), \xi\big\rangle\big| > l|\xi|, \qquad \big\langle h''(J)\xi, \xi\big\rangle \geqslant m|\xi|^2$$

holds for each $\xi \in \mathbf{R}^{d+1}$. Using the previous lemma it is possible to prove that, for every $\bar r > 0$, the Hamiltonian h^+ is l, m*-quasi-convex* in the set $S := D^+_{r_0} \times \{I \in \mathbf{R}^{d+1} \mid |I| \leqslant \bar r\}_{r_0}$ with $l, m = \mathrm{O}(1)$. In the previous set also holds

$$\big\|(h^+)''\big\| =: M = \mathrm{O}\big(\mu^{-c_d}\ln^{-3}(1/\mu)\big), \qquad \big\|(h^+)'\big\| =: \Omega_0 = \mathrm{O}(1).$$

[8] If $f(x), g(x)$ are positive function, with the symbol $f \approx g$ we mean that $\exists c_1, c_2 > 0$ such that $c_1 g(x) \leqslant f(x) \leqslant c_2 g(x)$, $\forall x$.

Putting

$$\varepsilon := \mu \|f^+\|_{S,s_0} = \mathrm{O}(\mu), \qquad \alpha := \frac{(1-2c_d(d+3))}{2(d+2)},$$
$$\varepsilon_0 := 2^{-10} r_0^2 m \left(\frac{m}{11M}\right)^{2(d+2)} = \mathrm{O}\big(\mu^{2c_d(d+3)} \ln^{6(d+2)}(1/\mu)\big),$$

we obtain that, if the initial data $(I(0), \varphi(0), p(0), q(0)) \in \mathcal{E}_1^+$, that is $P(0) \in D^+$, then

$$\big|I(t) - I(0)\big| \leqslant const.\mu^{\alpha} \ln^{-3}(1/\mu), \quad \text{for}$$
$$|t| \leqslant const. \exp\big(const.\mu^{-\alpha} \ln^2(1/\mu)\big). \tag{7.4}$$

If $c_d < 1/2(d+3)$ then $\alpha > 0$ and we obtain stability for exponentially long times.

7.2. *Stability in the region $\mathcal{E}_1^-$*

In the region $\mathcal{E}_1^-$ we cannot use the Nekhoroshev Theorem as proved in [19], because E^- is concave and so h^- is not quasi-convex. However we can still apply the Nekhoroshev Theorem in its original and more general form as proved in [17] (see also [18]); in fact the function h^- proves to be *steep* (see Definition 1.7.C, p. 6 of [17]).

For simplicity we prove the *steepness* of the function h^- in the case $d = 1$ only. In this case $h^- = h^-(I_0, I_1, P) = I_0 + I_1^2/2 + E^-(P)$. We need more informations on the function E^-. In the following, in order to simplify the notation, we will forget the apex $^-$ writing, for example, $E = E^-$ and $P = P^-$.

By (1.11) of [17], since $\nabla h^- \neq 0$, a sufficient condition for h^- to be steep is that the system

$$\eta_1 + I\eta_2 + E'(P)\eta_3 := 0, \qquad \eta_2^2 + E''(P)\eta_3^2 := 0, \qquad E'''(P)\eta_3^3 := 0, \tag{7.5}$$

has no real solution apart from the trivial one $\eta_1 = \eta_2 = \eta_3 = 0$.

Making the change of variable $\psi = \arccos(1 - \widetilde{E} + \xi\widetilde{E})$, where $\widetilde{E} = E + 2$, we get[9]

$$\dot{P}(E) = \int_0^1 F_1(\xi; E)\,\mathrm{d}\xi, \qquad \ddot{P}(E) = 3^{-1/2} \int_0^1 F_2(\xi; E)\,\mathrm{d}\xi,$$
$$\dddot{P}(E) = \int_0^1 F_3(\xi; E)\,\mathrm{d}\xi, \tag{7.6}$$

where

9 We will denote with "$\dot{}$" the derivative with respect to E, and with "$'$" the derivative with respect to P.

$$F_1(\xi;E) := \frac{\sqrt{2}}{\pi\sqrt{\xi}\sqrt{1-\xi}\sqrt{\widetilde{E}\xi - E}}, \qquad F_2(\xi;E) := \frac{\sqrt{6}\sqrt{1-\xi}}{2\pi\sqrt{\xi}(\widetilde{E}\xi - E)^{3/2}},$$
$$F_3(\xi;E) := \frac{3\sqrt{2}(1-\xi)^{3/2}}{4\pi\sqrt{\xi}(\widetilde{E}\xi - E)^{5/2}}. \tag{7.7}$$

From the equation $E(P(E)) = E$, deriving with respect to E, we obtain that

$$E'''\big(P(E)\big) = -\big(\dot{P}(E)\big)^{-5}\big[\dot{P}(E)\dddot{P}(E) - 3\big(\ddot{P}(E)\big)^2\big].$$

We want to prove that

$$E'''\big(P(E)\big) < 0, \tag{7.8}$$

for every E with $-2 < E < 0$. This is equivalent to prove that $\dot{P}(E)\dddot{P}(E) > 3(\ddot{P}(E))^2$. Using (7.7) we see that $F_1F_3 = F_2^2$ and hence, noting that $F_3(\xi;E)$ is not proportional to $F_1(\xi;E)$ for every E fixed, we conclude that $\int F_1 \int F_3 > (\int F_2)^2$ by a straightforward application of Cauchy–Schwartz inequality and (7.8) follows from (7.6).

By (7.8) the unique solution of the system (7.5) is the trivial one $\eta_1 = \eta_2 = \eta_3 = 0$, hence the function h^- is steep. It is simple to prove that the so-called *steepness coefficients* and *steepness indices* (see again Definition 1.7.C, p. 6 of [17]) can be taken uniformly for $-2 + \mu^{c_d} \leqslant E \leqslant -\mu^{c_d}$: that is they do not depend on μ.

Now we are ready to apply the Nekhoroshev Theorem in the formulation given in Theorem 4.4 of [17]. In order to use the notations of [17] we need the following substitutions:[10]

$$(I,P) \to I, \quad (\varphi, Q) \to \varphi, \quad H^- \to H, \quad h^- \to H_0, \quad \mu f^- \to H_1, \quad r_0 \to \rho,$$
$$\big\{I \in \mathbf{R}^{d+1} \bigm| |I| \leqslant \bar{r}\big\} \times D^- \to G, \quad \big\{I \in \mathbf{R}^{d+1} \bigm| |I| \leqslant \bar{r}\big\}_{r_0} \times \mathbf{T}^{d+1}_{s_0} \times D^+_{r_0} \times \mathbf{T}_{s_0} \to F.$$

Defining $m := \sup_F \|\partial^2 H_0/\partial I^2\|$ and remembering (7.3) and the definition of r_0, we have:

$$m \leqslant const.\mu^{-c_d}\ln^{-3}(1/\mu), \quad \rho = const.\mu^{c_d}. \tag{7.9}$$

In order to apply the theorem we have only to verify the following condition,

$$M := \sup_F |H_1| < M_0, \tag{7.10}$$

where M_0 depends only on the steepness coefficients and steepness indices (which are independent of μ) and on m and ρ (which depend on μ). Moreover we use the fact that the dependence of M_0 on m and ρ is, "polynomial" (although it is quite cumbersome): that

[10] We observe that we do not need to introduce the (p,q) variables so in our case $C = +\infty$.

is there exist constant $\tilde{c}_d, \bar{c}_d > 0$ such that $M_0(m,\rho) \geqslant const.m^{-\tilde{c}_d}\rho^{\bar{c}_d}$ (see Section 6.8 of [18]). So condition (7.10) becomes, using (7.9),

$$\mu \leqslant const.\mu^{c_d(\tilde{c}_d+\bar{c}_d)}\ln^{3\tilde{c}_d}(1/\mu),$$

which is verified choosing $c_d < (\tilde{c}_d + \bar{c}_d)^{-1}$.

Now we can apply the Nekhoroshev Theorem as formulated in Theorem 4.4 of [17], obtaining that if $(I(0), \varphi(0), p(0), q(0)) \in \mathcal{E}_1^-$ then

$$\begin{aligned} &\bigl|I(t) - I(0)\bigr| \leqslant \mathrm{d}/2 := M^b/2 = \mathrm{O}\bigl(\mu^b\bigr) \\ &\forall |t| \leqslant T := \frac{1}{M}\exp\left(\frac{1}{M}\right)^a = \mathrm{O}\left(\frac{1}{\mu}\exp\left(\frac{1}{\mu}\right)^a\right), \end{aligned} \tag{7.11}$$

where $a, b > 0$ are some constants depending only on the steepness properties of H_0. Finally, choosing

$$c_d < \min\bigl\{(2d+6)^{-1},\ (\tilde{c}_d + \bar{c}_d)^{-1}\bigr\}, \tag{7.12}$$

we have proved the exponential stability in the region $\mathcal{E}_1$.

7.3. Stability in the region $\mathcal{E}_2^+$

In the following we will denote $I^* := (I_1, \ldots, I_d)$ the projection on the last d coordinates. We shall prove the following lemma:

Lemma 7.2. $\forall \kappa > 0,\ \exists \kappa_0, \mu_8 > 0$ *such that* $\forall 0 < \mu \leqslant \mu_8$, *if* $(I(t), \varphi(t), p(t), q(t)) \in \mathcal{E}_2^+$ *for* $0 < t \leqslant \overline{T}$, *then*

$$\bigl|I^*(t) - I^*(0)\bigr| \leqslant \frac{\kappa}{2} \quad \forall t \leqslant \min\left\{\frac{\kappa_0}{\mu}\ln\frac{1}{\mu}, \overline{T}\right\}.$$

It is quite obvious that for initial conditions $(I(0), \varphi(0), p(0), q(0)) \in \mathcal{E}_2^+$, Theorem 1.2 follows from Lemma 7.2 and the exponential stability in the region $\mathcal{E}_1$.

In order to prove Lemma 7.2 let us define, for some fixed $0 < \delta < \pi/4$, the following two regions in the phase space: $U := \{(I, \varphi, p, q) \mid |q| \leqslant \delta \bmod 2\pi,\ |E(p,q)| \leqslant 2\mu^{c_d}\}$ and $V := \{(I, \varphi, p, q) \mid |q| > \delta \bmod 2\pi,\ |E(p,q)| \leqslant 2\mu^{c_d}\}$. We first note that[11]

$$\begin{aligned} &z(t) \in V\ \forall t_1 < t < t_2,\ \bigl|q(t_1)\bigr|, \bigl|q(t_2)\bigr| = \delta \bmod 2\pi \\ &\quad \Rightarrow \quad t_2 - t_1 < c_1,\ \bigl|I(t_2) - I(t_1)\bigr| \leqslant c_2(t_2 - t_1)\mu. \end{aligned} \tag{7.13}$$

[11] In the following we will use c_i to denote some positive constant independent on μ.

Indeed in this case $\forall t_1 < t < t_2$, $c_3 \leqslant |\dot q(t)| \leqslant c_4$. This implies that $t_2 - t_1 \leqslant c_1$ and then, integrating the equation of motion $\dot I = -\mu \partial_\varphi f$ in (t_1, t_2), we immediately get (7.13). We also claim that

$$\forall t_1 < t < t_2,\ z(t) \in U \text{ and } \big|q(t_1)\big|, \big|q(t_2)\big| = \delta \bmod 2\pi \Rightarrow t_2 - t_1 \geqslant c_5 |\ln \mu|. \tag{7.14}$$

We denote with t_U^i (respectively t_V^i) the ith time for which the orbit enters in (respectively goes out from) U, so that $t_U^i < t_V^i < t_U^{i+1} < t_V^{i+1}$ for $0 \leqslant i \leqslant i_0$. From (7.14) it follows that $i_0 \leqslant c_6 \kappa_0/\mu$ and, from (7.13), that the time T_V spent by the orbit in the region V is bounded by $c_7 \kappa_0/\mu$.

In order to prove (7.14) we use the following normal form result for the pendulum Hamiltonian $E(p, q)$ in a neighborhood of its hyperbolic equilibrium point (see, e.g., [12]).

Lemma 7.3. *There exist $R, \tilde\delta > 0$, an analytic function g, with $g'(0) = -1$ and an analytic canonical transformation*

$$\Phi : B \to \big\{|p| \leqslant \tilde\delta\big\} \times \big\{|q| \leqslant \delta \bmod 2\pi\big\} \quad \textit{where } B := \big\{|P|, |Q| \leqslant R\big\},$$

such that $E(\Phi(P, Q)) = g(PQ)$.

In the coordinates (Q, P) the local stable and unstable manifolds are respectively $W^s_{\mathrm{loc}} = \{P = 0\}$ and $W^u_{\mathrm{loc}} = \{Q = 0\}$ and Hamiltonian (7.1) writes as

$$\widetilde H := \widetilde H(I, \varphi, P, Q) := h(I) + g(PQ) + \mu \tilde f(I, \varphi, P, Q)$$

where $\tilde f(I, \varphi, P, Q) := f(I, \varphi, \Phi(P, Q))$.

We are now able to prove (7.14). Certainly there exists an instant $t_1^* \in [t_1, t_2)$ for which $(p(t_1^*), q(t_1^*)) \in \Phi(B)$ but, $\forall t_1 < t < t_1^*$, $(p(t), q(t)) \notin \Phi(B)$. It follows that, if we take the representant $q(t_1) \in [-\delta, \delta]$, then $p(t_1^*)q(t_1^*) < 0$. We will denote with $Z(t) := (I(t), \varphi(t), P(t), Q(t)) = (I(t), \varphi(t), \Phi^{-1}(p(t), q(t)))$ the corresponding solution of the Hamiltonian system associated to $\widetilde H$. From the fact that $|q(t_1^*)| = \delta$ or $(p(t_1^*), q(t_1^*)) \in \partial\Phi(B)$ and that $|g(PQ)| \leqslant \mu^{c_d}$, $p(t_1^*)q(t_1^*) < 0$, it follows that $|P(t_1^*)| \leqslant c_8 \mu^{c_d}$ and $|Q(t_1^*)| \geqslant c_9$.

In the same way there exists an instant t_2^* with $t_1 < t_1^* < t_2^* < t_2$ for which $(P(t_2^*), Q(t_2^*)) \in B$ but, $\forall t > t_2^*$ $(P(t), Q(t)) \notin B$; in particular it results $|P(t_2^*)| \geqslant c_{10}$. We claim that $t_2^* - t_1^* \geqslant c_{11} \ln(1/\mu)$. Indeed $P(t)$ satisfies the Hamilton's equation $\dot P(t) = -g'(P(t)Q(t))P(t) - \mu \partial_Q \tilde f(I(t), \varphi(t), P(t), Q(t))$ with initial condition $|P(t_1^*)| \leqslant c_8 \mu^{c_d}$. Since $|P(t_2^*)| \geqslant c_{10}$, we can derive from Gronwall's lemma that $t_2^* - t_1^* \geqslant c_{11} \ln(1/\mu)$, which implies (7.14).

By the following normal-form lemma there exists a close to the identity symplectic change of coordinates removing the nonresonant angles φ in the perturbation up to $\mathrm{O}(\mu^2)$. It can be proved by standard perturbation theory (see for similar lemmas Section 5 of [12]).

Lemma 7.4. *Let $\beta > 0$. There exist $R, \rho > 0$ so small that, defining $\lambda := \min_{|\xi| \leqslant R^2} |g'(\xi)|$, $S := \max_{|\xi| \leqslant R^2} |g''(\xi)|$, then $\lambda \geqslant 2SR^2$ and $\rho \leqslant \min\{\lambda/4N,\ R^2/8s,\ \beta/2N, r\}$. Let Λ be* *a sublattice of* $\mathbf{Z}^{d+1}$. *Let* $\mathcal{D} \subset \mathbf{R}^{d+1}$ *be bounded and* β*-nonresonant mod* Λ, *i.e.,* $\forall I \in \mathcal{D}$, $h \in \mathbf{Z}^{d+1} \setminus \Lambda$, $|h| \leqslant N$ *it results* $|(1, I^*) \cdot h| \geqslant \beta$. *Suppose that*

$$\varepsilon := \mu \|\tilde{f}\|_{B,D,s} \leqslant 2^{-11}\beta_* \rho s, \tag{7.15}$$

where[12] $D := \mathcal{D}_\rho$, $\beta_* := \min\{\beta, \lambda/2\}$. *Then there exists an analytic canonical transformation*:

$$\begin{aligned} \Psi : \overline{D} \times \mathbf{T}^{d+1}_{s/4} \times \overline{B} \to D \times \mathbf{T}^{d+1}_{s} \times B, \\ (\bar{I}, \bar{\varphi}, \overline{P}, \overline{Q}) \mapsto (I, \varphi, P, Q), \end{aligned} \tag{7.16}$$

with $\overline{B} := \{|\overline{P}|, |\overline{Q}| \leqslant R/8\}$, $\overline{D} := \mathcal{D}_{\rho/4}$, *such that*

$$\overline{H} := \overline{H}(\bar{I}, \bar{\varphi}, \overline{P}, \overline{Q}) := \widetilde{H} \circ \Psi = h(\bar{I}) + \bar{g}(\bar{I}, \bar{\varphi}, \overline{P}\,\overline{Q}) + \bar{f}(\bar{I}, \bar{\varphi}, \overline{P}, \overline{Q})$$

with $\bar{g}(\bar{I}, \bar{\varphi}, \xi) := g(\xi) + f^*(\bar{I}, \bar{\varphi}, \xi)$, $f^*(\bar{I}, \bar{\varphi}, \xi) = \sum_{h \in \Lambda,\ |h| \leqslant N} f^*_h(\bar{I}, \xi)\, \mathrm{e}^{\mathrm{i}h\cdot\bar{\varphi}}$ *and* $\|f^*\|_{\overline{B},\overline{D},s/4} \leqslant \varepsilon$. *Moreover the following estimates hold*

$$|\bar{I} - I| \leqslant \frac{2^4\varepsilon}{\beta_* s}, \qquad |\overline{P} - P|, |\overline{Q} - Q| \leqslant \frac{2^5\varepsilon}{R\beta_*}, \qquad \|\bar{f}\|_{\overline{B},\overline{D},s/4} \leqslant \frac{2^9\varepsilon^2}{\beta_*\rho s}. \tag{7.17}$$

Let $\mathcal{L}$ be the (finite) set of the maximal sublattices $\Lambda = \langle h_1, \ldots, h_s \rangle \subset \mathbf{Z}^{d+1}$ for some independent $h_i \in \mathbf{R}^{d+1}$ with $|h_i| \leqslant N$ for $i = 1, \ldots, s \leqslant d$. For $\Lambda \in \mathcal{L}$ we define the Λ-resonant frequencies $R^\Lambda := \{I^* \in \mathbf{R}^d \mid (1, I^*) \cdot h = 0,\ \forall h \in \Lambda\}$ and the set of the s-order resonant frequencies $Z^s := \bigcup_{\dim \Lambda = s} R^\Lambda$.

Setting $h_i = (l_i, n_i)$ with $l_i \in \mathbf{R}$, $n_i \in \mathbf{R}^d$, we remark that if $R^\Lambda \neq \emptyset$ then $n_1, \ldots, n_s$ are independent. We also define the $(d - s)$-dimensional linear subspace (associated with the affine subspace R^Λ) $L^\Lambda := \bigcap_{i=1}^s n_i^\perp \subset \mathbf{R}^d$ and we denote by Π^Λ the orthogonal projection from $\mathbf{R}^d$ onto L^Λ.

Since $\mathcal{L}$ is a finite set, $\alpha := \min_{\Lambda \in \mathcal{L}} \min_{n \in \mathbf{Z}^d, |n| \leqslant N, \Pi^\Lambda n \neq 0} |\Pi^\Lambda n|$ is strictly positive.

We now perform a suitable version of the standard “covering lemma” in which the whole frequency space is covered by nonresonant zones. The fundamental blocks used to construct this covering will be r-neighborhoods of any R^Λ, i.e., $R^\Lambda_r := \{I^* \in \mathbf{R}^d \mid \mathrm{dist}(I^*, R^\Lambda) \leqslant r\}$ for suitable $r > 0$ depending on $\dim \Lambda$. Let $r_d > 0$ be such that $(d + 1) r_d < c_{12}\kappa$, for some c_{12} sufficiently small to be determined. For $1 \leqslant s \leqslant d - 1$ we can define recursively numbers r_s sufficiently small such that $0 < r_s < \alpha r_{s+1}/2N$, verifying[13]

$$\dim \Lambda = \dim \Lambda' = s,\ R^\Lambda \neq R^{\Lambda'} \quad \Rightarrow \quad R^\Lambda_{(s+1)r_s} \cap R^{\Lambda'}_{(s+1)r_s} \subset \bigcup_{i=s+1}^{d} Z^i_{r_i}. \tag{7.18}$$

[12] B and D are thought as complex domains, as in the sequel $\overline{B}$ and $\overline{D}$.

[13] Assumption (7.18) means that, in order to go from a neighborhood of a $(d - s)$-order resonance to a different one, we have to pass through an higher-order dimensional one.

We also define, for $1 \leqslant s \leqslant d-1$,

$$S^0 := \mathbf{R}^d \Big\backslash \left(\bigcup_{i=1}^{d} Z^i_{2r_i} \right) \quad \text{and} \quad S^s := Z^s_{(s+1)r_s} \Big\backslash \left(\bigcup_{i=s+1}^{d} Z^i_{(s+2)r_i} \right),$$

i.e., the s-order resonances minus the higher-order ones. We claim that $\mathbf{R}^d = S^0 \cup \cdots \cup S^{d-1} \cup Z^d_{(d+1)r_d}$ is the covering that we need. We also define

$$S^0 \subset S^0_* := \mathbf{R}^d \Big\backslash \left(\bigcup_{i=1}^{d} Z^i_{r_i} \right) \quad \text{and} \quad S^s \subset S^s_* := Z^s_{(s+1)r_s} \Big\backslash \left(\bigcup_{i=s+1}^{d} Z^i_{(s+1)r_i} \right).$$

If the orbit lies near a certain R^Λ (but far away from higher-order resonances) then the following lemma says that the drift of the actions I^* in the direction which is parallel to R^Λ is small.

Lemma 7.5. *Suppose that $I^*(0) \in S^s$, $I^*(t) \in S^s_*$ and $|I^*(t)| \leqslant \bar{r} + r/2$, $\forall 0 \leqslant t \leqslant T^*$ for some $T^* \leqslant \kappa_0 |\ln \mu|/\mu$ and $0 \leqslant s \leqslant d-1$. Then, if $s \geqslant 1$, there exists a sublattice $\Lambda \subset \mathbf{Z}^{d+1}$, $\dim \Lambda = s$ such that $I^*(t) \in R^\Lambda_{(s+1)r_s} \setminus (\bigcup_{i=s+1}^d Z^i_{(s+1)r_i})$, $\forall 0 \leqslant t \leqslant T^*$. Moreover if κ_0 is sufficiently small*[14]

$$\left| \Pi^\Lambda \big(I^*(t) - I^*(0) \big) \right| \leqslant r_1/2 \quad \forall 0 \leqslant t \leqslant T^* \tag{7.19}$$

and hence, for $s \geqslant 1$, $|I^*(t) - I^*(0)| \leqslant 2(s+1)r_s + r_1/2$. In particular for $I^*(0) \in S^0$ we have that $|I^*(t) - I^*(0)| \leqslant r_1/2$, $\forall 0 \leqslant t \leqslant T^*$.

Proof. In the case $s = 0$ we take $\Lambda = \{0\}$. The existence of Λ is trivial because $I^*(0) \in S^s$ and hence $I^*(0) \in R^\Lambda_{(s+1)r_s}$ for some $\Lambda \in \mathcal{L}$ with $\dim \Lambda = s$. The fact that $I^*(t) \in R^\Lambda_{(s+1)r_s} \setminus (\bigcup_{i=s+1}^d Z^i_{(s+1)r_i})$, $\forall 0 \leqslant t \leqslant T^*$, follows from $I^*(t) \in S^s_*$, $\forall 0 \leqslant t \leqslant T^*$ and (7.18). Now we want to apply Lemma 7.4 with $\beta := \alpha r_1/2$ and $\mathcal{D} := R^\Lambda_{(s+1)r_s} \setminus (\bigcup_{i=s+1}^d Z^i_{(s+1)r_i})$. We have to verify that $\mathcal{D}$ is β-nonresonant mod Λ. Fix $|h_0| \leqslant N$, $h_0 = (l_0, n_0) \notin \Lambda$ (respectively $\neq 0$ for $s = 0$). We first estimate $|l_0 + n_0 \cdot I^*_0|$ for all $I^*_0 \in \mathcal{D}_0 := R^\Lambda \setminus (\bigcup_{i=s+1}^d Z^i_{(s+1)r_i})$. If $\Lambda' := \Lambda \oplus \langle h_0 \rangle$ and $n^*_0 := \Pi^\Lambda n_0$ we have two cases: $n^*_0 \neq 0$ or $n^*_0 = 0$. If $n^*_0 \neq 0$ we can perform the following decomposition: $I^*_0 = I^*_1 + v$ with $I^*_1 \in R^{\Lambda'}$, $v \in L^\Lambda$ and moreover[15] $v = \pm |v| n^*_0 / |n^*_0|$. Since $I^*_0 \notin (\bigcup_{i=s+1}^d Z^i_{(s+1)r_i})$ then $I^*_0 \notin Z^{\Lambda'}_{(s+1)r_{s+1}}$ and, hence $|v| \geqslant (s+1)r_{s+1}$. Using the previous estimate, the fact that $I^*_1 \in \Lambda'$ and $|n^*_0| \geqslant \alpha$, we conclude that

[14] In the case $s = 0$ Π^Λ is simply the identity on $\mathbf{R}^d$.

[15] We observe that $\operatorname{dist}(I^*_0, R^{\Lambda'}) = |v|$.

$$\begin{aligned}\left|l_0 + n_0 \cdot I_0^*\right| &= \left|\left(l_0 + n_0 \cdot I_1^*\right) + n_0 \cdot v\right| = |n_0 \cdot v| = \left|n_0^* \cdot v\right| = |v|\left|n_0^*\right| \\ &\geqslant \alpha(s+1) r_{s+1}.\end{aligned} \tag{7.20}$$

Now we consider the case in which $n_0^* = 0$. In this case it is simple to see that $h_0 = (l', 0) + h$ where $h \in \Lambda$ and $l' \in \mathbf{Z} \setminus \{0\}$. So $|l_0 + n_0 \cdot I_0^*| = |l'| \geqslant 1$. Now we can prove that $|l_0 + n_0 \cdot I^*| \geqslant \beta$ for all $I^* \in \mathcal{D}$. In fact $I^* = I_0^* + u$ with $I_0^* \in \mathcal{D}_0$ and $|u| \leqslant (s+1) r_s$. Using (7.20) and $r_s < \alpha r_{s+1}/2N$, we have

$$\begin{aligned}|l_0 + n_0 \cdot I^*| &\geqslant \left|l_0 + n_0 \cdot I_0^*\right| - |n_0 \cdot u| \geqslant \alpha(s+1) r_{s+1} - N(s+1) r_s \\ &\geqslant \alpha(s+1) r_{s+1}/2 \geqslant \beta,\end{aligned}$$

proving that $\mathcal{D}$ is β-nonresonant mod Λ. Finally we can verify (7.15) if μ_8 is sufficiently small. Now we are ready to apply Lemma 7.4 in order to prove (7.19). Using (7.13), the fact that f^* contains only the Λ-resonant Fourier coefficients, (7.17) and Hamilton's equation for $\overline{H}$ we have:

$$\begin{aligned}\left|\Pi^{\Lambda}\left(I^*(t) - I^*(0)\right)\right| &\leqslant c_2 T_V \mu + c_{13}\mu^2\left(\kappa_0 |\ln \mu|/\mu\right) + c_{14} i_0 \mu \\ &\leqslant c_2 c_7 \kappa_0 + c_{13}\mu\kappa_0 |\ln \mu| + c_{14} c_6 \kappa_0 \leqslant r_1/2\end{aligned}$$

if κ_0 and μ_8 are sufficiently small. □

Proof of Lemma 7.2. Suppose first that $|I^*(t)| \leqslant \bar{r} + r/2$ $\forall 0 \leqslant t \leqslant \kappa_0 |\ln \mu|/\mu$. If $I^*(0) \in Z^d_{(d+1)r_d}$ and $I^*(t) \in Z^d_{(d+1)r_d}$ $\forall 0 \leqslant t \leqslant \kappa_0 |\ln \mu|/\mu$ then $|I^*(t) - I^*(0)| \leqslant 2(d+1) r_d$ and the lemma is proved if $c_{12} < 1/4$. Otherwise we can suppose that $I^*(0) \in S^s$ for some $0 \leqslant s \leqslant d-1$. If $I^*(t) \in S^s_*$ $\forall 0 \leqslant t \leqslant \kappa_0 |\ln \mu|/\mu$ then we can apply Lemma 7.5 proving the lemma for c_{12} small enough. Suppose that $\exists\, 0 < T^* < \kappa_0 |\ln \mu|/\mu$ such that $I^*(t) \in S^s_*$ $\forall 0 \leqslant t < T^*$ but $I^*(T^*) \notin S^s_*$. We will prove that

$$I^*(T^*) \in S^0 \cup \cdots \cup S^{s-1} \tag{7.21}$$

that means that the orbit can only enter in zones that are "less" resonant. In fact by Lemma 7.5 we see that $I^*(T^*) \notin \bigcup_{i=s+1}^d Z^i_{(s+1)r_i}$, moreover, since $I^*(T^*) \notin S^s_*$, we have that $I^*(T^*) \notin Z^s_{(s+1)r_s}$ and hence $I^*(T^*) \notin \bigcup_{i=s}^d Z^i_{(s+1)r_i}$. If $I^*(T^*) \in S^0$ we have finished. If $I^*(T^*) \notin S^0$ then $I^*(T^*) \in \bigcup_{i=1}^{s-1} Z^i_{2r_i} \subseteq \bigcup_{i=1}^{s-1} Z^i_{(i+1)r_i}$. If $I^*(T^*) \in S^1$ we have finished. If $I^*(T^*) \notin S^1$ then $I^*(T^*) \notin Z^1_{2r_1} \setminus \bigcup_{i=2}^d Z^i_{3r_i}$ and hence $I^*(T^*) \in \bigcup_{i=2}^{s-1} Z^i_{(i+1)r_i}$. Iterating this procedure we prove (7.21).

The conclusion is that if the order of resonance changes along the orbit, it can decrease only so that the orbit may eventually arrive in the completely nonresonant zone S^0 where there is stability. Considering the "worst" case, i.e., when $I^*(0) \in Z^d_{(d+1)r_d}$ and the orbit arrives in S^0, summing all the contributions from Lemma 7.5, we have that, if c_{12} is sufficiently small,

$$\begin{aligned}\left|I^*(t)-I^*(0)\right| &\leqslant 2(d+1)r_d+\sum_{s=1}^{d-1}\bigl(2(s+1)r_s+r_1/2\bigr)+r_1/2\\ &=\sum_{s=1}^{d}2(s+1)r_s+dr_1/2\leqslant \kappa/2. \end{aligned}\tag{7.22}$$

In order to conclude the proof of the lemma we have only to prove that if $|I^*(0)|\leqslant \bar r$ then $|I^*(t)|\leqslant \bar r+r/2\ \forall 0\leqslant t\leqslant \kappa_0|\ln\mu|/\mu$. This is an immediate consequence of (7.22) and of the fact that $\kappa\leqslant r$. □

7.4. Stability in the region $\mathcal{E}_2^-$

If, for all $t\geqslant 0$ $(p(t),q(t))\in\mathcal{E}_2^-$, then it follows easily that $|p(t)|,|q(t)-\pi|=\mathrm{O}(\mu^{c_d/2})$. Then, defining $f_1(I,\varphi):=f(I,\varphi,0,\pi)$ and $f_2(I,\varphi,t):=\mu^{-c_d/2}[f(I,\varphi,p(t),q(t))-f_1(I,\varphi)]$, it results that $|\partial_I f_2(I,\varphi;t)|,|\partial_\varphi f_2(I,\varphi;t)|\leqslant const$. Clearly if $(I(t),\varphi(t),q(t),p(t))$ is a solution of (7.1) then $(I(t),\varphi(t))$ is solution of Hamiltonian

$$H_1:=H_1(I,\varphi;t):=h(I)+\mu f_1(I,\varphi)+\mu^{1+(c_d/2)}f_2(I,\varphi;t).$$

Now[16] one can construct, in the standard way, an analytic symplectic map $\Phi:(\bar I,\bar\varphi)\to(I,\varphi)$ with $|\bar I-I|=\mathrm{O}(\mu/\beta)$, and two analytic functions $\bar h,\bar f$ such that $[h+\mu f_1]\circ\Phi(\bar I,\bar\varphi)=\bar h(\bar I)+\bar f(\bar I,\bar\varphi)$ with $\|\bar f\|=\mathrm{O}(\mu^2)$. Defining $f_3:=f_3(\bar I,\bar\varphi;t):=f_2(\Phi(\bar I,\bar\varphi);t)$ we also get that $|\partial_{\bar I}f_3(\bar I,\bar\varphi;t)|,|\partial_{\bar\varphi}f_3(\bar I,\bar\varphi;t)|\leqslant const./\beta$. The solutions of the Hamiltonian H_1 are symplectically conjugated, via Φ^{-1}, to the solutions of the Hamiltonian

$$H_2:=H_2\bigl(\bar I,\bar\varphi;t\bigr):=\bar h\bigl(\bar I\bigr)+\bar f\bigl(\bar I,\bar\varphi\bigr)+\mu^{1+(c_d/2)}f_3\bigl(\bar I,\bar\varphi;t\bigr)$$

for which we obtain, directly from Hamilton's equations, the estimates:

$$\left|\bar I(t)-\bar I(0)\right|\leqslant const.\mu^{c_d/4},\qquad \forall |t|\leqslant const.\mu^{-1-c_d/4}.$$

It follows that, if $(I(0),\varphi(0),p(0),q(0))\in\mathcal{E}_2^-$, then

$$\begin{aligned}\left|I(t)-I(0)\right|&\leqslant\left|I(t)-\bar I(t)\right|+\left|\bar I(t)-\bar I(0)\right|+\left|\bar I(0)-I(0)\right|\\ &\leqslant const.\mu^{c_d/4},\quad \forall|t|\leqslant const.\mu^{-1-c_d/4}\end{aligned}$$

(if at some instant t the solution $z(t)$ escapes outside $\mathcal{E}_2^-$ it is exponentially stable in time).

Finally, from the previous steps, we can conclude that there exists $\mu_1>0$ such that $0<\mu\leqslant\mu_1$ Theorem 1.2 holds.

[16] For brevity we prove only the case in which $I(0)$ is in a nonresonant zone. The resonant case can be treated as in $\mathcal{E}_2^+$.

Appendix A

Proof of Lemma 2.1. We shall use the following lemma:

Lemma A.1. *There exists $T_0 > 0$ such that, $\forall T \geqslant T_0$, for all continuous*

$$f : [-1, T+1] \to \mathbf{R},$$

there exists a unique solution h of

$$-\ddot{h} + \cos Q_T(t) h = f, \quad h(0) = h(T) = 0. \tag{A.1}$$

The Green operator $\mathcal{G} : C^0([-1, T+1]) \to C^2([-1, T+1])$ defined by $\mathcal{G}(f) := h$, satisfies

$$\max_{t \in [-1, T+1]} \left|h(t)\right| + \left|\dot{h}(t)\right| \leqslant C \max_{t \in [-1, T+1]} \left|f(t)\right| \tag{A.2}$$

for some positive constant C independent of T.

Proof. We first note that the homogeneous problem (A.1) (i.e., $f = 0$) admits only the trivial solution $h = 0$. This immediately implies the uniqueness of the solution of (A.1). The existence result follows by the standard theory of linear second-order differential equations. We now prove that any solution h of (A.1) satisfies (A.2). It is enough to show that $\max_{t \in [-1, T+1]} |h(t)| \leqslant C' \max_{t \in [-1, T+1]} |f(t)|$. Indeed we obtain by (A.1) that $\max_{t \in [-1, T+1]} |h(t)| + |\ddot{h}(t)| \leqslant (2C' + 1) \max_{t \in [-1, T+1]} |f(t)|$ and, by elementary analysis, this implies (A.2) for an appropriate constant C.

Arguing by contradiction, we assume that there exist sequences $(T_n) \to \infty$, (f_n), (h_n) such that

$$-\ddot{h}_n + \cos Q_{T_n}(t) h_n = f_n, \quad h_n(0) = h_n(T_n) = 0,$$
$$|h_n|_n := \max_{t \in [-1, T_n+1]} \left|h_n(t)\right| = 1, \quad |f_n|_n \to 0.$$

By the Ascoli–Arzela Theorem there exists $h \in C^2([-1, \infty), \mathbf{R})$ such that, up to a subsequence, $h_n \to h$ in the topology of C^2 uniform convergence in $[-1, M]$ for all $M > 0$. Since $Q_{T_n} \to q_0 - 2\pi$ uniformly in all bounded intervals of $[-1, \infty)$, we obtain that

$$-\ddot{h} + \cos q_0(t) h = 0, \quad h(0) = 0, \quad \sup_{t \in [-1, \infty)} \left|h(t)\right| \leqslant 1. \tag{A.3}$$

Now the solutions of the linear differential equation in (A.3) have the form $h = K_1 \xi + K_2 \psi$, where $(K_1, K_2) \in \mathbf{R}^2$, $\xi(t) = \dot{q}_0(t) = 2/\cosh t$ and $\psi(t) = \frac{1}{4}(\sinh t + t/\cosh t)$ satisfies $\dot{\psi}\xi - \dot{\xi}\psi = 1$. The bound on h implies that $K_2 = 0$ and $h(0) = 0$ implies that $K_1 = 0$. Hence $h = 0$. In the same way we can prove that $h_n(\cdot - T_n) \to 0$ uniformly in every bounded subinterval of $(-\infty, 1]$.

Now let us fix $\bar{t}$ such that for all n large enough, for all $t \in [\bar{t}, T_n - \bar{t}]$, $\cos Q_{T_n}(t) \geqslant 1/2$ ($\bar{t}$ does exist because of (2.4)). By the previous step, for n large enough, there exists a maximum point $t_n \in (\bar{t}, T_n - \bar{t})$ of $h_n^2(t)$, i.e., $h_n^2(t_n) = |h_n|_n^2 = 1$. Then $(\dot{h_n^2})(t_n) = 2h_n(t_n)\dot{h}_n(t_n) = 0$ and $(\ddot{h_n^2})(t_n) = 2\ddot{h}_n(t_n)h_n(t_n) + 2\dot{h}_n^2(t) \leqslant 0$. By the differential equation satisfied by h_n, we can derive from the latter inequality that $\cos Q_{T_n}(t_n)h_n^2(t_n) \leqslant f_n(t_n)h_n(t_n)$, i.e., $\cos Q_{T_n}(t_n) \leqslant f_n(t_n)$, which, for n large enough, contradicts the property of $\bar{t}$ and the fact that $|f_n|_n \to 0$. □

Now we can deal with the existence result of Lemma 2.1. Let $T := (\theta^- - \theta^+)$, $\omega = (\varphi^- - \varphi^+)/T$, $\overline{\varphi}(t) := \omega(t - \theta^+) + \varphi^+$. In the following we call c_i constants depending only on f. We are searching for solutions (φ, q) of (2.1) with $\varphi(\theta^\pm) = \varphi^\pm$, $q(\theta^\pm) = \mp\pi$, in the following form:

$$\begin{cases} \varphi(t) = \omega(t - \theta^+) + \varphi^+ + v(t - \theta^+), \\ q(t) = Q_T(t - \theta^+) + w(t - \theta^+). \end{cases}$$

Hence we need to find a solution, in the time interval $I := [-1, T + 1]$, of the following two equations:

$$\begin{cases} \ddot{v}(t) = -\mu\big[F_\varphi(v, w)\big](t), & v(0) = v(T) = 0, \\ \big[L(w)\big](t) = \big[G(v, w)\big](t) := -\big[S(w)\big](t) + \mu\big[F_q(v, w)\big](t), & w(0) = w(T) = 0, \end{cases} \tag{A.4}$$

where

$$\big[F_\varphi(v, w; \lambda, \mu)\big](t) := \partial_\varphi f\big(\omega t + \varphi^+ + v(t), Q_T(t) + w(t), t + \theta^+\big),$$
$$\big[F_q(v, w; \lambda, \mu)\big](t) := \partial_q f\big(\omega t + \varphi^+ + v(t), Q_T(t) + w(t), t + \theta^+\big),$$
$$\big[S(w)\big](t) := \sin\big(Q_T(t) + w(t)\big) - \sin\big(Q_T(t)\big) - \cos\big(Q_T(t)\big)w(t),$$
$$\big[L(w)\big](t) := -\ddot{w}(t) + \cos Q_T(t)w(t).$$

We want to solve (A.4) as a fixed point problem. By Lemma A.1, the second equation of (A.4) can be written $w = K := \mathcal{G}(-S + \mu F_q)$. Moreover the first equation (A.4) can be written

$$v(t) = J(t) := \big[J(v, w; \lambda, \mu)\big](t) := \overline{J}(t) - \frac{\overline{J}(0)(T - t) + \overline{J}(T)t}{T}, \tag{A.5}$$

where, setting $F_\varphi(s) = F_\varphi(v(s), w(s))$,

$$\big[\overline{J}(v, w; \lambda, \mu)\big](t) := -\mu \int_{T/2}^{t} \int_{T/2}^{x} F_\varphi(s)\,\mathrm{d}s\,\mathrm{d}x.$$

Let us consider the Banach space $Z = V \times W := \mathcal{C}^1(I;\mathbf{R}^d) \times \mathcal{C}^1(I;\mathbf{R})$, endowed with the norm $\|z\| = \|(v,w)\| := \max\{\|v\|_V,\ \|w\|_W\}$, defined by:

$$\|v\|_V := \sup_{t\in I}\Big[\big|v(t)\big|\big(1+c_1\mu T^2\big)^{-1}\beta^2 + \big|\dot v(t)\big|\beta\Big],$$
$$\|w\|_W := \sup_{t\in I}\Big[\big|w(t)\big| + \big|\dot w(t)\big|\Big]. \tag{A.6}$$

A fixed point of the operator $\Phi : Z \to Z$ defined $\forall z \in Z$ as $\Phi(z) := \Phi(z;\lambda,\mu) := (J(z), K(z))$ is a solution of (A.4). We shall prove in the sequel that Φ is a contraction in the ball[17] $D := B_{\bar c\mu(Z)}$ for an appropriate choice of $\bar c, c_1, C_0$, provided μ is small enough.

We have $|[S(w)](t)| \leqslant w^2(t)$, so that $\forall t,\ |[G(v,w)](t)| \leqslant \bar c^2\mu^2 + c_4\mu$. Now, choosing first $\bar c$ sufficiently large and then μ sufficiently small, we can conclude using (A.2) that, if $z\in D$, $\|K(z)\|_W \leqslant \bar c\mu/4$. Now we study the behavior of J. Let us first consider $\overline{J}$. We define:

$$f_{nl}(t) := f_{nl}\big(Q_T(t)+w(t)\big), \qquad g_{nl}(t) := f'_{nl}\big(Q_T(t)+w(t)\big),$$
$$\alpha_{nl} := n\cdot\varphi^+ + l\theta^+, \qquad \beta_{nl} := n\cdot\omega + l.$$

For $t\in[-1, T+1]$, $z\in D$, we want to estimate:

$$\dot{\overline{J}}(t) = -\mu\int_{T/2}^{t} F_\varphi = -\mu\sum_{|(n,l)|\leqslant N} \mathrm{i}n\,\mathrm{e}^{\mathrm{i}\alpha_{nl}}\int_{T/2}^{t} f_{nl}(s)\,\mathrm{e}^{\mathrm{i}n\cdot v(s)}\,\mathrm{e}^{\mathrm{i}\beta_{nl}s}\,\mathrm{d}s.$$

Integrating by parts, we obtain:

$$-\mathrm{i}\beta_{nl}\int_{T/2}^{t} f_{nl}(s)\,\mathrm{e}^{\mathrm{i}n\cdot v(s)}\mathrm{e}^{\mathrm{i}\beta_{nl}s}\,\mathrm{d}s$$
$$= f_{nl}(T/2)\,\mathrm{e}^{\mathrm{i}n\cdot v(T/2)}\,\mathrm{e}^{\mathrm{i}\beta_{nl}T/2} - f_{nl}(t)\,\mathrm{e}^{\mathrm{i}n\cdot v(t)}\,\mathrm{e}^{\mathrm{i}\beta_{nl}t} \tag{A.7}$$
$$+ \int_{T/2}^{t} g_{nl}(s)\dot Q_T(s)e^{\mathrm{i}n\cdot v(s)}\,\mathrm{e}^{\mathrm{i}\beta_{nl}s}\,\mathrm{d}s \tag{A.8}$$
$$+ \int_{T/2}^{t}\big(g_{nl}(s)\dot w(s) + f_{nl}(s)\mathrm{i}n\cdot\dot v(s)\big)\,\mathrm{e}^{\mathrm{i}n\cdot v(s)}\,\mathrm{e}^{\mathrm{i}\beta_{nl}s}\,\mathrm{d}s. \tag{A.9}$$

By (2.4), the term (A.8) is bounded by $c_5\max\{\mathrm{e}^{-K_2 t},\ \mathrm{e}^{-K_2(T-t)}\}$. Hence, for $z\in D$,

[17] If X is a Banach space and $r>0$ we define $B_r(X) := \{x\in X \mid \|x\|\leqslant r\}$.

$$\int_{T/2}^{t} F_\varphi = u(t) - u(T/2) + R(t),$$

$$\text{with } \left|R(t)\right| \leqslant \frac{c_6}{\beta}\left[\max\left\{e^{-K_2 t}, e^{-K_2(T-t)}\right\} + \bar{c}\left(\mu + \frac{\mu}{\beta}\right)T\right], \tag{A.10}$$

where $u(t) = \sum (n/\beta_{nl})\, e^{i\alpha_{nl}} f_{nl}(t)\, e^{in\cdot v(t)}\, e^{i\beta_{nl}t}$.

So we can write $\overline{J}(t) = j(t) + \mu(t - T/2)u(T/2)$, where

$$j(t) = \int_{T/2}^{t} -\mu u(s)\,ds + \int_{T/2}^{t} -\mu R(s)\,ds.$$

By the bound of $R(t)$ given in (A.10), the second integral can be bounded by $c_7(\mu/\beta)[1 + \bar{c}T^2\mu/\beta]$. Integrating once again by parts as above, we find that the first integral is bounded by $c_8(\mu/\beta^2)[1 + \bar{c}(\mu T/\beta)]$, hence, by the condition imposed on μT, it can be bounded by $\mu\bar{c}/8\beta^2$, provided that C_0 has been chosen small enough and $\bar{c}$ is large enough. Hence

$$\left|j(t)\right| \leqslant \frac{\mu\bar{c}}{\beta^2}\left[\frac{c_7}{\bar{c}} + c_7\mu T^2 + \frac{1}{8}\right].$$

In addition

$$\left|\frac{d}{dt} j(t)\right| = \mu\left|u(t) + R(t)\right| \leqslant c_{10}\frac{\mu\bar{c}}{\beta}\left(\frac{1}{\bar{c}} + \frac{\mu T}{\beta}\right).$$

As a result $\|j\|_V \leqslant \mu\bar{c}/4$, provided $\bar{c}$ and c_1 have been chosen large enough, C_0 small enough.

Now $\overline{J}(t) = j(t) + at + b$, where $a, b \in \mathbf{R}$, so that we may replace $\overline{J}$ with j in (A.5). Since $|J(t)| \leqslant |j(t)| + \max\{|j(0)|, |j(T)|\}(T+2)/T$ and $|\dot{J}(t)| \leqslant |dj(t)/dt| + (1/T)\int_1^{T+1} |dj(s)/ds|\,ds$, we obtain $\|J\|_V \leqslant 3\|j\|_V \leqslant \mu 3\bar{c}/4$. We have finally proved that Φ maps D into itself (in fact into $B_{3\bar{c}\mu/4}$).

Now we must prove that Φ is a contraction. Φ is differentiable and for $z = (v, w) \in D$, $(D\Phi(z)[h, g])(t) = (r(t), s(t))$, r and $s : [-1, T+1] \to \mathbf{R}$ being defined by:

$$\begin{aligned} &\ddot{r}(t) = a_1(t).h(t) + b_1(t)g(t), \quad r(0) = r(T) = 0, \\ &L(s)(t) = a_2(t).h(t) + b_2(t)g(t), \quad s(0) = s(T) = 0, \end{aligned} \tag{A.11}$$

where

$$\begin{aligned} a_1(t) &= -\mu\partial_{\varphi\varphi} f\left(\omega t + \varphi^+ + v(t), Q_T(t) + w(t), t + \theta^+\right), \\ b_1(t) &= -\mu\partial_{\varphi q} f\left(\omega t + \varphi^+ + v(t), Q_T(t) + w(t), t + \theta^+\right), \quad a_2(t) = -b_1(t), \\ b_2(t) &= \cos\left(Q_T(t) + w(t)\right) - \cos Q_T(t) \\ &\quad + \mu\partial_{qq} f\left(\omega t + \varphi^+ + v(t), Q_T(t) + w(t), t + \theta^+\right). \end{aligned}$$

By the same arguments as above $(A, B) \in V_1 \times V$ (where $V_1 := C^1(I, \mathbf{R}^{d^2})$) defined by:

$$\ddot{A}(t) = a_1(t), \quad A(0) = A(T) = 0, \qquad \ddot{B}(t) = b_1(t), \quad B(0) = B(T) = 0$$

satisfy $\|A\|_{V_1} + \|B\|_V \leqslant c_{11}\bar{c}\mu$ ($\| \ \|_{V_1}$ being defined in the same way as $\| \ \|_V$).

Using an integration by parts, we can derive from (A.11) and the bound on $\|A\|_{V_1} + \|B\|_V$ that

$$\left|\dot{r}(t)\right| \leqslant c_{12}\bar{c}\frac{\mu}{\beta}\left[\left(\frac{1+c_1\mu T^2}{\beta^2}\|h\|_V + \|g\|_W\right) + T\left(\frac{\|h\|_V}{\beta} + \|g\|_W\right)\right]. \tag{A.12}$$

Therefore, for C_0 small enough, $|\beta\dot{r}(t)| \leqslant 1/8 \max\{\|h\|_V, \|g\|_W\}$. We derive also from (A.12) that

$$\left|r(t)\right| \leqslant c_{13}\bar{c}\left[\frac{\mu T}{\beta^3} + \frac{c_1\mu^2T^3}{\beta^3} + \frac{\mu T^2}{\beta^2}\right]\max\left\{\|h\|_V, \|g\|_W\right\},$$

which yields

$$\beta^2\left(1+c_1\mu T^2\right)^{-1}\left|r(t)\right| \leqslant c_{14}\bar{c}\left(\mu T/\beta + \frac{1}{c_1}\right)\max\left\{\|h\|_V, \|g\|_W\right\} \leqslant \frac{\max\{\|h\|_V, \|g\|_W\}}{8},$$

provided C_0 is small enough and $c_1/\bar{c}$ is large enough. Finally,

$$\|r\|_V \leqslant \frac{\max\{\|h\|_V, \|g\|_W\}}{4}.$$

Using the properties of L and the fact that

$$\left|a_2(t).h(t) + b_2(t)g(t)\right| \leqslant c_{15}\mu\left(1+c_1\mu T^2\right)/\beta^2\|h\|_V + c_{15}\left(|w(t)|+\mu\right)\|g\|_W$$

we easily derive $\|s\|_W \leqslant \max\{\|h\|_V, \|g\|_W\}/4$ (again provided that C_0, more precisely C_0c_1 is small enough). We have proved that for a good choice of $\bar{c}, c_1, C_0$, $\|D\Phi(z)[h, g]\| \leqslant \|(h, g)\|/2$ for $z \in D$. Hence Φ is a contraction. As a result, it has a unique fixed point z_λ in D (which in fact belongs to $B_{3\bar{c}\mu/4}$). This proves existence.

Now there remains to prove that $\varphi_{\mu,\lambda}(t), q_{\mu,\lambda}(t)$ are C^1 functions of (λ, t). Let (θ_0^+, θ_0^-) be fixed with $T_0 := \theta_0^- - \theta_0^+$ and let $\Lambda = \{\lambda \mid |\theta^+ - \theta_0^+| \leqslant 1/4, \ \ |\theta^- - \theta_0^-| \leqslant 1/4\}$. For $\lambda \in \Lambda$ $I_0 := [-1/2, T_0 + 1/2] \subset [-1, \theta^- - \theta^+ + 1]$, hence the restrictions v_λ^0 and w_λ^0 of v_λ and w_λ to I_0 are well defined.

Let $V_0 \times W_0 := C^1(I_0, \mathbf{R}^n) \times C^1(I_0, \mathbf{R})$ be endowed with the norm $\| \ \|_0$ as defined in (A.6). Define $\Psi : \Lambda \to V_0 \times W_0$ by $\Psi(\lambda) = z_\lambda^0$. We shall justify briefly that Ψ is differentiable and that $\|D\Psi\| \leqslant c_{16}\mu$. z_λ^0 is the unique solution in $B_{\bar{c}\mu}$ of (A.4) (with $T = \theta^- - \theta^+$), which is equivalent to $(v_\lambda, w_\lambda) = \Phi(z_\lambda; \theta^+, \theta^-, \varphi^+, \varphi^-, \mu)$, where $\Phi : B_{\bar{c}\mu} \times \Lambda \times (0, \mu_2) \to V_0 \times W_0$ is smooth. Now, by the previous step, $\|D_z\Phi\| \leqslant 1/2$ everywhere, so that $I - D_z\Phi$ is invertible. Therefore, by the Implicit Function Theorem, Ψ is C^1. This proves that $(\lambda, t) \mapsto \varphi_{\mu,\lambda}(t)$ (respectively $(\lambda, t) \mapsto q_{\mu,\lambda}(t)$) and

$(\lambda, t) \mapsto \dot{\varphi}_{\mu,\lambda}(t)$ (respectively $(\lambda, t) \mapsto \dot{q}_{\mu,\lambda}(t)$) have continuous partial derivatives w.r.t. λ in the set $\{(\lambda, t) | -1/2 + \theta^+ < t < 1/2 + \theta^-\}$, and by the standard theory of differential equations, these partial derivatives have continuous extensions on $\{(\lambda, t) | -1 + \theta^+ < t < 1 + \theta^-\}$. Finally, by (2.1), $\ddot{\varphi}_{\mu,\lambda}$ and $\ddot{q}_{\mu,\lambda}$ depend continuously on (λ, t). □

Appendix B

Proof of Theorem 4.2. In order to prove Theorem 4.2 we need a preliminary lemma. Observe that Λ_R^* is a finite set which is symmetric with respect to the origin. Hence, if it is not empty there exists $p \in \Lambda_R^*$ such that $p \cdot \Omega = \alpha(\Lambda, \Omega, R)$.

Lemma B.1. *Assume that $\Lambda_R^* \neq \emptyset$ and let $p \in \Lambda_R^*$ be such that $p \cdot \Omega = \alpha := \alpha(\Lambda, \Omega, R)$. Assume moreover that $\alpha > 0$ and define $E := [p]^\perp$. Then $\Lambda_0 := \Lambda \cap E$ is a lattice of E. In addition*:

(i) $\alpha/\beta|p| \leqslant 2/R$, *where* $\beta = \inf\{|q \cdot \Omega| \mid q \in (\Lambda_0)^*_{\sqrt{3}R/2}\}$,
$(\Lambda_0)^* = \{q \in E \mid \forall x \in \Lambda_0 \; q \cdot x \in \mathbf{Z}\}$.

In particular $\alpha \leqslant 2\beta$.

(ii) $\alpha(\Lambda, \Omega, \sqrt{7}R/2) \leqslant \beta$.

Proof. Since Λ is a lattice, it is not contained in E. Hence $p \cdot \Lambda$ is a nontrivial subgroup of $\mathbf{Z}$, $p \cdot \Lambda = m\mathbf{Z}$ for some integer $m \geqslant 1$, which implies that $p/m \in \Lambda^*$. But $p/m \cdot \Omega = \alpha/m$ and $|p/m| \leqslant R$, hence by the definition and the positivity of α, $m = 1$. As a result there exists $\bar{x} \in \Lambda$ such that $p \cdot \bar{x} = 1$. Obviously $\Lambda_0 + \mathbf{Z}\bar{x} \subseteq \Lambda$. On the other hand, all $x \in \Lambda$ can be written as $x = (x \cdot p)\bar{x} + y$, where $y \in \Lambda$, $y \cdot p = 0$, i.e., $y \in \Lambda_0$. So the reverse inclusion holds and we may write $\Lambda = \Lambda_0 + \mathbf{Z}\bar{x}$. As a consequence Λ_0 is a lattice of E and

$$\Lambda^* = \left\{r \in \mathbf{R}^l \mid r \cdot \Lambda_0 \subset \mathbf{Z} \text{ and } r \cdot \bar{x} \in \mathbf{Z}\right\} = \left\{q + ap \,\middle|\, q \in \Lambda_0^*,\ a \in \mathbf{Z} - q \cdot \bar{x}\right\},$$
$$\Lambda_R^* = \left\{q + ap \,\middle|\, q \in \Lambda_0^*,\ a \in \mathbf{Z} - q \cdot \bar{x},\ 0 < |q|^2 + a^2|p|^2 \leqslant R^2\right\}.$$

If $\beta = +\infty$ there is nothing more to prove. If $\beta < +\infty$, let $q \in (\Lambda_0)^*_{\sqrt{3}R/2}$ be such that $q \cdot \Omega = \beta$. Let

$$S = \left\{a \in \mathbf{R} \,\middle|\, q + ap \in \Lambda_R^*\right\} = \left\{a \in \mathbf{R} \,\middle|\, a \in \mathbf{Z} - q \cdot \bar{x},\ |a| \leqslant \left(R^2 - |q|^2\right)^{1/2}/|p|\right\}.$$

Since $|q|^2 \leqslant 3R^2/4$, $S \supseteq S' := (\mathbf{Z} - q \cdot \bar{x}) \cap [-R/2|p|, R/2|p|]$. Hence by the definition of α, for all $a \in S'$, $|(q + ap) \cdot \Omega| = |\beta + a\alpha| \geqslant \alpha$, i.e., $\beta/\alpha \notin (-1 - a, 1 - a)$.

As $|p| \leqslant R$, the interval $[-R/2|p|, R/2|p|]$ has length $\geqslant 1$ and must intersect $(\mathbf{Z} - q \cdot \bar{x})$. Therefore $S' \neq \emptyset$, more precisely $S' = \{u, u+1, \ldots, u+K\}$, for some integer $K \geqslant 0$, where $u = \inf S'$. As a result,

$$\beta/\alpha \notin \bigcup_{k=0}^{K} (-1-u-k, 1-u-k) = (-1-u-K, 1-u).$$

Now $S' \cap [-1/2, 1/2] \neq \emptyset$, hence $u + K \geqslant -1/2$ and $-1-u-K < 0$. As a consequence $\beta/\alpha \geqslant 1-u$. Since $[-R/2|p|, -R/2|p|+1] \subseteq [-R/2|p|, R/2|p|]$ intersects $\mathbf{Z} - q \cdot \bar{x}$, $u \leqslant -R/2|p| + 1$. Therefore $\beta/\alpha \geqslant R/2|p|$, which is (i). In particular, since $|p| \leqslant R$, $\alpha \leqslant 2\beta$.

Finally there exists $a \in [-1, 0) \cap (\mathbf{Z} - q \cdot \bar{x})$; $q + ap \in \Lambda^*$, and $|q + ap|^2 = |q|^2 + a^2|p|^2 \leqslant 3R^2/4 + R^2 = 7R^2/4$. Hence $q + ap \in \Lambda^*_{\sqrt{7}R/2}$. We have $|(q + ap) \cdot \Omega| = |\beta + a\alpha| \leqslant \beta$, because $-1 \leqslant a \leqslant 0$ and $\alpha \leqslant 2\beta$. This proves (ii). □

Now we turn to the proof of Theorem 4.2. We first prove that the statement is true for $l = 1$, with $a_1 = 1/2$. Here $\Lambda = \lambda_0 \mathbf{Z}$ for some $\lambda_0 > 0$, and $\Lambda^* = (\lambda_0)^{-1}\mathbf{Z}$. We can assume without loss of generality that $\Omega > 0$. If $\lambda_0 < 2\delta$, then for all $x \in \mathbf{R}$, $d(x, \Lambda) < \delta$. Hence $T(\Lambda, \Omega, \delta) = 0$.

If $\lambda_0 \geqslant 2\delta$, then it is easy to see that $T(\Lambda, \Omega, \delta) = (\lambda_0 - 2\delta)/\Omega \leqslant \lambda_0/\Omega$. On the other hand, $1/\lambda_0 \in \Lambda^*_{1/2\delta}$ and $\alpha(\Lambda, \Omega, 1/(2\delta)) = \Omega/\lambda_0$. The result follows.

Now we assume that the statement holds true up to dimension $l - 1$ ($l \geqslant 2$). We shall prove it in dimension l.

Fix $R > 0$ and define $\delta_R = (4a_{l-1}^2/3 + 4)^{1/2}/R$. We claim that:

(a) If $\Lambda^*_R = \emptyset$ then $T(\Lambda, \Omega, \delta_R) = 0$.
(b) If $\Lambda^*_R \neq \emptyset$, let $p \in \Lambda^*_R$ be such that $p \cdot \Omega = \alpha := \alpha(\Lambda, \Omega, R)$, and define β as in Lemma B.1. Then

$$T(\Lambda, \Omega, \delta_R) \leqslant \max\left\{\alpha^{-1}, \beta^{-1}\right\}.$$

Postponing the proof of (a) and (b), we show how to define a_l. In the case (b), by Lemma B.1(ii), $T(\Lambda, \Omega, \delta_R) \leqslant \alpha(\Lambda, \Omega, \sqrt{7}R/2)^{-1}$. This estimate obviously holds in the case (a) too. Hence for all $R > 0$,

$$T\left(\Lambda, \Omega, \left(4a_{l-1}^2/3 + 4\right)^{1/2}/R\right) \leqslant \alpha\left(\Lambda, \Omega, \sqrt{7}R/2\right)^{-1}.$$

As a consequence, the statement of Theorem 4.2 holds with $a_l = (\sqrt{7}(4a_{l-1}^2/3 + 4)^{1/2}/2)$.

There remains to prove (a) and (b). First assume that $\Lambda^*_R = \emptyset$. Let $p \in \Lambda^* \setminus \{0\}$ be such that for all $p' \in \Lambda^* \setminus \{0\}$, $|p| \leqslant |p'|$. Then $|p| > R$. Let E, Λ_0 be defined from p as in Lemma B.1.

Arguing by contradiction, we assume that $(\Lambda_0)^*_{\sqrt{3}R/2} \neq \emptyset$. By the same arguments as previously there exist $q \in (\Lambda_0)^*_{\sqrt{3}R/2}$ and $a \in [-1/2, 1/2]$ such that $q + ap \in \Lambda^*$. But

$$|q + ap|^2 = |q|^2 + a^2|p|^2 \leqslant (3/4)R^2 + |p|^2/4 < |p|^2$$

and this contradicts the definition of p. Hence $(\Lambda_0)^*_{\sqrt{3}R/2} = \emptyset$ and by the iterative hypothesis, all points of E lies at a distance from Λ_0 less than $2a_{l-1}/\sqrt{3}R$.

From the proof of Lemma B.1, there exists $\bar{x} \in \Lambda$ such that $p \cdot \bar{x} = 1$ and $\Lambda = \Lambda_0 + \mathbf{Z}\bar{x}$. Therefore for all $x \in \mathbf{R}^l$, there is $x' \in x + \Lambda$ such that $|x' \cdot p| \leqslant 1/2$. This implies that $d(x', E) \leqslant 1/(2|p|) \leqslant 1/(2R)$ and hence that $d(x', \Lambda_0) \leqslant (4a_{l-1}^2/3 + 1/4)^{1/2}/R \leqslant \delta_R$. Hence the distance from any point of $\mathbf{R}^l$ to Λ is not greater than δ_R. This completes the proof of (a).

Next assume that $\Lambda_R^* \neq \emptyset$ and let p be as in Lemma B.1. Define α and β in the same way as in Lemma B.1. Let $x \in \mathbf{R}^l$. Again $\Lambda = \Lambda_0 + \mathbf{Z}\bar{x}$ for some $\bar{x} \in \Lambda$ such that $p \cdot \bar{x} = 1$, hence there exists $x' \in x + \Lambda$ such that $p \cdot x' \in [0, 1)$. We have:

$$x' = y + \frac{w}{|p|^2}p, \qquad \Omega = U + \frac{\alpha}{|p|^2}p,$$

with $y, U \in E = [p]^\perp$, $w = p \cdot x' \in [0, 1)$. We shall assume that $\alpha > 0$ (if $\alpha = 0$, there is nothing to prove). Let $\bar{t} = w/\alpha$, and consider the time interval defined by

$$J = [0, 1/\beta] \quad \text{if } \bar{t} < 1/\beta, \qquad J = [\bar{t} - 1/\beta, \bar{t}] \quad \text{if } \bar{t} \geqslant 1/\beta.$$

$J \subset [0, \max\{1/\beta, 1/\alpha\}]$, and it is enough to prove that there exists $t \in J$ such that $d(x', t\Omega + \Lambda_0) \leqslant \delta_R$. The length of J is not less than $1/\beta$. Hence by the iterative hypothesis, there exists $t \in J$ such that $d(y, tU + \Lambda_0) \leqslant 2a_{l-1}/(\sqrt{3}R)$ (notice that for all $q \in \Lambda_0^*$, $q \cdot U = q \cdot \Omega$, so that the linear flow (tU) creates a $2a_{l-1}/(\sqrt{3}R)$-net of E/Λ_0 in time β^{-1}). We have:

$$d(x', t\Omega + \Lambda_0)^2 = \left(\frac{(t - \bar{t})\alpha}{|p|}\right)^2 + d(y, tU + \Lambda_0)^2 \leqslant \left(\frac{\alpha}{\beta|p|}\right)^2 + \frac{4a_{l-1}^2}{3R^2}.$$

Hence, by Lemma B.1(i), $d(x', t\Omega + \Lambda_0) \leqslant (4a_{l-1}^2/3 + 4)^{1/2}/R$. This completes the proof of (b). □

Note added in proof

After this paper was accepted we learned of the preprints:

D. Treshev, Evolution of slow variables in a priori unstable Hamiltonian systems, Preprint.

A. Delshams, R. de la Llave, T.M. Seara, A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model, Preprint.

References

[1] A. Ambrosetti, M. Badiale, Homoclinics: Poincaré–Melnikov type results via a variational approach, Ann. Inst. H. Poincaré Anal. Non Linéaire 15 (2) (1998) 233–252.

[2] V.I. Arnold, Instability of dynamical systems with several degrees of freedom, Soviet Math. Dokl. 6 (1964) 581–585.

[3] M. Berti, L. Biasco, P. Bolle, Optimal stability and instability results of a class of nearly integrable Hamiltonian systems, Rend. Mat. Accad. Naz. Lincei 13 (2002) 77–84.

[4] M. Berti, P. Bolle, Diffusion time and splitting of separatrices for nearly integrable isochronous Hamiltonian systems, Rend. Mat. Accad. Naz. Lincei Ser. (9) 11 (4) (2000) 235–243.

[5] M. Berti, P. Bolle, A functional analysis approach to Arnold diffusion, Ann. Inst. Poincaré Anal. Non Linéaire 19 (2002) 395–450.

[6] M. Berti, P. Bolle, Fast Arnold's diffusion in systems with three time scales, Discrete Contin. Dynam. Systems Ser. A 8 (3) (2002) 795–811.

[7] U. Bessi, An approach to Arnold diffusion through the calculus of variations, Nonlinear Anal. 26 (1996) 1115–1135.

[8] U. Bessi, Arnold's example with three rotators, Nonlinearity 10 (1997) 763–781.

[9] U. Bessi, L. Chierchia, E. Valdinoci, Upper bounds on Arnold diffusion time via Mather theory, J. Math. Pures Appl. 80 (1) (2001) 105–129.

[10] L. Biasco, L. Chierchia, On the stability of some properly-degenerate Hamiltonian systems, Discrete Contin. Dynam. Systems Ser. A 9 (2) (2003) 233–262.

[11] J. Bourgain, F. Golse, B. Wennberg, On the distribution of free path lengths for periodic Lorentz gas, Comm. Math. Phys. 190 (1998) 491–508.

[12] L. Chierchia, G. Gallavotti, Drift and diffusion in phase space, Ann. Inst. H. Poincaré Phys. Théor. 60 (1994) 1–144; see also Erratum: Ann. Inst. H. Poincaré Phys. Théor. 68 (1998) 135.

[13] J. Cresson, Conjecture de Chirikov et optimalité des exposants de stabilité du théorème de Nekhoroshev, Preprint, Univ. Besançon.

[14] J. Cresson, C. Guillet, Periodic orbits and Arnold diffusion, Discrete Contin. Dynam. Systems 9 (2) (2003) 451–470.

[15] P. Lochak, Arnold diffusion: a compendium of remarks and questions, in: C. Simó (Ed.), Hamiltonian Systems With Three More Degrees of Freedom, Proceedings of the NATO Advanced Study Institute, S'Agaró, Spain, 1995, in: NATO ASI Ser. Ser. C, Math. Phys. Sci., Vol. 533, Kluwer, Dordrecht, 1999, pp. 168–183.

[16] J.P. Marco, Transitions le long des chaînes de tores invariants pour les systèmes hamiltoniens analytiques, Ann. Inst. H. Poincaré Phys. Théor. 64 (1996) 205–252.

[17] N.N. Nekhoroshev, An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems, Russian Math. Survey 32 (1977).

[18] N.N. Nekhoroshev, An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems, II, Trudy Sem. Petrovsk. 5 (1979) 5–50; in: O.A. Oleinik (Ed.), Topics in Modern Mathematics, Petrovskii Semin., Vol. 5, Consultants Bureau, New York, 1985.

[19] J. Pöschel, Nekhoroshev estimates for quasi-convex Hamiltonian systems, Math. Z. 213 (1993) 187–216.