Reviews in Mathematical Physics
Vol. 15, No. 4 (2003) 1–47

EXPONENTIALLY SMALL SPLITTING AND ARNOLD DIFFUSION FOR MULTIPLE TIME SCALE SYSTEMS

MICHELA PROCESI

Received 8 May 2002
Revised 10 February 2003

We consider the class of Hamiltonians:

$$\frac{1}{2}\sum_{j=1}^{n-1} I_j^2 + \frac{1}{2}\varepsilon I_n^2 + \frac{p^2}{2} + \varepsilon[(\cos q - 1) - b^2(\cos 2q - 1)] + \varepsilon\mu f(q)\sum_{i=1}^{n}\sin(\psi_i)\,,$$

where $0 \le b < \frac{1}{2}$, and the perturbing function $f(q)$ is a rational function of e^{iq}. We prove upper and lower bounds on the splitting for such class of systems, in regions of the phase space characterized by one fast frequency. Finally using an appropriate Normal Form theorem we prove the existence of chains of heteroclinic intersections.

Keywords:

Contents

1. Presentation of the Model and Main Theorems

The general setting of this paper is the problem of homoclinic splitting and Arnol'd diffusion in *a priori* stable systems with three or more relevant time scales. The general strategy is the one proposed in [1] and [2] and in particular the application to *a priori* stable systems proposed in [3] and further developed in [4]. More precisely we consider a class of close to integrable n degrees of freedom Hamiltonian systems for which one can prove the existence of $(n-1)$-dimensional unstable KAM tori together with their stable and unstable manifolds. We use a perturbative diagrammatic construction (proposed and developed in [3], [4] and [5]) to prove upper bounds on the angles of intersection of the stable and unstable manifolds of a KAM torus (homoclinic splitting). Such bounds are generally exponentially small in the perturbation parameter and depend on the chosen torus and in particular on the number of fast degrees of freedom. For systems with one fast degree of freedom we prove as well lower bounds on the homoclinic splitting through the mechanism of Melnikov dominance. Finally for such systems we prove the existence of "long" chains of heteroclinic intersections; namely we produce a list of unstable KAM tori $\mathcal{T}_1, \dots, \mathcal{T}_h$ such that $\mathcal{T}_1, \mathcal{T}_h$ are at distances of order one in the action variables and the unstable manifold of each $\mathcal{T}_i$ intersects the stable manifold of $\mathcal{T}_{i+1}$. This paper is a generalization of the results of [4], [5], [6], therefore in proving our claims we will rely heavily on intermediate results proved in the latter papers which we will not prove again.

Consider the class of Hamiltonians

$$\frac{1}{2}\sum_{j=1}^{n-1}\tilde{I}^2 + \frac{1}{2}\varepsilon\tilde{I}_n^2 + \frac{\tilde{p}^2}{2} + \varepsilon\left[(\cos\tilde{q}-1) - \frac{1-c^2}{4}(\cos 2\tilde{q}-1)\right] + \varepsilon\mu f(\tilde{q})\sum_{i=1}^{n}\sin\tilde{\psi}_i\,, \tag{1.1}$$

where the pairs $\tilde{I} \in \mathbb{R}^n$, $\tilde{\psi} \in \mathbb{T}^n$ and $\tilde{p} \in \mathbb{R}$, $\tilde{q} \in \mathbb{T}$ are conjugate action-angle coordinates, $0 < c \le 1$, $f(\tilde{q})$ is odd and analytic on the torus and μ, ε are small parameters. We will consider them independent and then prove that one can prove Arnold Diffusion for $\mu \le \varepsilon^P$, for an appropriate P.

This class of Hamiltonians is a model for a near to integrable system close to a simple resonance where the dependence on the hyperbolic variables is not through the standard pendulum, but still maintains various qualitative properties of the pendulum. Namely we have a "generalized pendulum",

$$\frac{\tilde{p}^2}{2} + \varepsilon\left[(\cos\tilde{q}-1) - \frac{1-c^2}{4}(\cos 2\tilde{q}-1)\right]$$

which has an unstable fixed point in $\tilde{p} = \tilde{q} = 0$ with Lyapunov exponent $\lambda = c\sqrt{\varepsilon}$.

Generally one rescales the time and action variables so that the Lyapunov exponent is one:

$$I(t) = \frac{\tilde{I}\left(\frac{t}{c\sqrt{\varepsilon}}\right)}{c\sqrt{\varepsilon}}\,, \quad \psi(t) = \tilde{\psi}\left(\frac{t}{c\sqrt{\varepsilon}}\right), \quad p(t) = \frac{\tilde{p}\left(\frac{t}{c\sqrt{\varepsilon}}\right)}{c\sqrt{\varepsilon}}\,, \quad q(t) = \tilde{q}\left(\frac{t}{c\sqrt{\varepsilon}}\right). \tag{1.2}$$

Such rescaling sends Hamiltonian 1.1 in

$$\frac{(I, A(\varepsilon)I)}{2} + \frac{p^2}{2} + \frac{1}{c^2}\left[(\cos q - 1) - \frac{1-c^2}{4}(\cos 2q - 1)\right] + \mu f(q)\sum_{i=1}^{n}\sin(\psi_i) \qquad (1.3)$$

where $A(\varepsilon)$ is the diagonal matrix with eigenvalues $a_i = 1$ for $i = 1, \ldots, n-1$ and $a_n = \varepsilon$. So from now on we will work on Hamiltonian (1.3) and turn back to Hamiltonian (1.1) only to prove the existence of *heteroclinic chains*. The system (1.3) is integrable for $\mu = 0$. It represents a list of n uncoupled rotators and a generalized pendulum (depending on the parameter c). We will denote the frequency of the rotators (which determines the initial data $I(0)$) by ω so that

$$I(t) = I(0) = A^{-1}\omega\,, \quad \psi(t) = \psi(0) + \omega t\,.$$

The initial data are chosen in an appropriate domain (physically interesting in the variables $\tilde{I}$) so that there are at least three characteristic orders of magnitude for the frequencies of the unperturbed system.

Definition 1.1. *In frequency space we first consider the ellipsoid*

$$\Sigma := \left\{x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i^2/a_i = 2E\right\}$$

where E is an order one constant[a] $E \sim O_\varepsilon(1)$.

For notational convenience we split the frequency ω in two vectorial components: $\omega = (\frac{\omega_1}{\sqrt{\varepsilon}}, \varepsilon^\alpha\omega_2)$ *with* $\omega_1 \in \mathbb{R}^m$, $\omega_2 \in \mathbb{R}^{n-m}$, *and* $0 \le \alpha \le \frac{1}{2}$. *Finally, given two suitable order one constants* R, $r \sim O_\varepsilon(1)$, *we consider the region*

$$\Omega \equiv \{\omega \in \mathbb{R}^n : \sqrt{\varepsilon}\omega \in \Sigma\,, \quad r < |\omega_{1,i}| < R \quad \textit{and} \quad r < |\omega_2| < R\,, \quad \varepsilon^\alpha|\omega_{2,i}| \ge \sqrt{\varepsilon}\,,$$
$$\varepsilon^\alpha|\omega_{2,n-m}| \sim \sqrt{\varepsilon}\}\,.$$

We have chosen the generalized pendulum so that its dynamics on the separatrix is particularly simple,[b] namely

$$q(t) = 2\ \mathrm{arc}\cot g\left(\frac{1}{c}\sinh(\pm t)\right), \quad e^{iq(t)} = \frac{\sinh(\pm t) + ic}{\sinh(\pm t) - ic}\,. \qquad (1.4)$$

There are at least three characteristic time scales $O_\varepsilon(\varepsilon^{-\frac{1}{2}})$, $O_\varepsilon(\varepsilon^\alpha)$, $O_\varepsilon(\sqrt{\varepsilon})$ (coming from the degenerate variable I_n) and 1 which is the Lyapunov exponent of the unperturbed pendulum.

We will call $\psi_1, \ldots, \psi_m$ the fast variables and we will sometimes denote them as $\psi_F \in \mathbb{T}^m$. Conversely we will call $\psi_{m+1}, \ldots, \psi_n$ the slow variables $\psi_S \in \mathbb{T}^{n-m}$.

[a]Now and in the following we will say $a(\varepsilon) \sim O_\varepsilon(f(\varepsilon))$ if $\lim_{\varepsilon\to 0^+}\frac{a(\varepsilon)}{f(\varepsilon)} = L \ne 0$.

[b]The motion on the separatrix can be easily obtained by direct computation; the main feature is that the motion on the separatrix is such that $e^{iq(t)}$ is a rational function of e^t. Here we are considering the simplest class of examples, which contains the standard pendulum $c = 1$.

The perturbing function is a trigonometric polynomial of degree one in the rotators ψ and a rational function[c] in e^{iq}. We have decoupled the dependence of ψ and q only to simplify the computations. For each $\omega \in \mathbb{R}^n$ the unperturbed system has an unstable fixed torus,

$$p(t) = q(t) = 0\,, \quad I(t) = I(0) = A^{-1}\omega\,, \quad \psi(t) = \psi(0) + \omega t\,.$$

The stable and unstable manifolds of such tori coincide and can be expressed as graphs on the angles.

Definition 1.2. *Given any $\gamma \in \mathbb{R}$, $\varepsilon < \gamma \le O(\varepsilon^{\frac{1}{2}})$ and a fixed $\tau > n-1$, we define the set*

$$\Omega_\gamma \equiv \left\{\omega \in \Omega : |\omega \cdot l| > \frac{\gamma}{|l|^\tau}\,, \quad \forall\, l \in \mathbb{Z}^n/\{0\}\right\}$$

of γ, τ Diophantine vectors in Ω. Now we consider

$$\Omega_\gamma^* \equiv \Omega_\gamma \times \left(-\frac{1}{2}, \frac{1}{2}\right)$$

and for all $(\omega, \rho) \in \Omega_\gamma^*$ we set $\omega_\rho = (1+\rho)\omega$.

For all $(\omega, \rho) \in \Omega_\gamma^*$ and for all $l \in \mathbb{Z}^n/\{0\}$ $|\omega_\rho \cdot l| > \frac{\gamma}{2|l|^\tau}$, $\omega \in \Omega_\gamma$ implies that ω_1 and ω_2 are Diophantine as well; we will call τ_F and τ_S their exponents.

KAM like theorems (see [2], [5]) imply that there exists $\mu_0(\varepsilon, \gamma) \sim \varepsilon^2$ such that if $|\mu| \le \mu_0$ and if $(\omega, \rho) \in \Omega_\gamma^*$, there exists one and only one n-dimensional H_μ-invariant unstable torus $T_\mu(\omega, \rho)$ whose Hamiltonian flow is analytically conjugated to the flow $\mathbb{T}^n \ni \vartheta \to \vartheta + \omega_\rho t$. Moreover one can parameterize the stable and unstable manifolds of $T_\mu(\omega, \rho)$ by functions $I^\pm(\omega, \varphi, q, \mu)$, analytic in the last three arguments, with $\varphi, q \in \mathbb{T}^n \times [-\frac{3}{2}\pi, \frac{3}{2}\pi]$. Namely given

$$z^\pm(\omega, \varphi, q, \mu) = (I^\pm(\omega, \varphi, q, \mu), p^\pm(\omega, \varphi, q, \mu), \varphi, q)\,,$$

where the pendulum action is derived by energy conservation, the trajectory[d]:

$$z(\omega, \varphi, q, \mu, t) = \begin{cases} \Phi_H^t z^+(\omega, \varphi, q, \mu) & \text{if } t > 0 \\ \Phi_H^t z^-(\omega, \varphi, q, \mu) & \text{if } t < 0 \end{cases}$$

tends exponentially to a quasi-periodic function of frequency ω.

Remark 1.3. We have introduced the variable ρ in order to fix the energy of the perturbed system,[e] namely given a list of $\omega_i \in \Omega_\gamma$ one can find $\rho(\omega_i, \mu)$ such that all the corresponding whiskered tori are on the same energy surface, see for instance [5].

[c] Actually it is sufficient that the singularity of $f(\psi(t), q(t))$, which is nearest to the real axis is polar and isolated.
[d] Φ_H^t is the evolution at time t of the Hamiltonian flow (1.3).
[e] The final goal is to find heteroclinic intersections on the fixed energy surface, and so "Arnold diffusion", but in the following sections we will discuss only homoclinic intersections and so we will drop the parameter ρ.

Definition 1.4. *We will study the difference between the stable and unstable manifolds on an hyper-plane transverse to the flow* (*a Poincaré section*), *we choose the hyper-plane* $q = \pi$ *and consequently drop the dependence on* q. *We call*

$$G_j^0(\varphi,\omega) = \frac{1}{2} a_j (I_j(\varphi,\omega,0^-) - I_j(\varphi,\omega,0^+))$$

the splitting vector and prove that $G_j^0(\varphi = 0,\omega) = 0$. *A measure of the transversality is*

$$\Delta_{ij}^0 = \partial_{\varphi_j} G_i^0(\varphi)|_{\varphi=0}$$

called splitting matrix.

We will prove the following theorems:

Theorem 1. *The splitting matrix* Δ^0 *satisfies the formal power series relation*[f]:

$$\Delta^0 \sim A D^0 B$$

where A, B *are close to identity matrices and* D^0 *is the "holomorphic part" of the splitting matrix; namely its entries are expressed as integrals over* $\mathbb{R}$ *of analytic functions. Moreover the formal power series involved are all asymptotic.*[g]

This statement was posed as a conjecture in [7] Paragraph 3.

Corollary 1.5. *The preceding Theorem implies that Hamiltonian* (1.3), *in regions of the action variables corresponding to* $m \neq 0$ *fast time scales, has exponentially small upper bounds on the determinant of the splitting matrix*:

$$|\det \Delta^0| \leq C e^{-\frac{c}{\varepsilon^b}}, \quad \textit{with } b = \frac{1}{2m},$$

provided that $\mu < \varepsilon^{1+2\frac{n}{m}}$.

Notice that Theorem 1 *can be proved for much more general systems than model* (1.3).

Theorem 2. *Consider Hamiltonian* (1.3) *in regions of the action variables corresponding to* $m = 1$ *fast variables and for perturbing functions* $f(q)$ *such that the pole* $f(q(t))$ *closest to the imaginary axis, say* $\bar{t}$, *is such that* $|\mathrm{Im}\, \bar{t}| = d \leq \mathrm{arc\,sin}\, c$. *Setting* $\mu \leq \varepsilon^P$ *with* $P = p/2 + 8 + 4n$ *where* p *is the degree of the pole of* $f(q(t))$ *in* $\bar{t}$ *we prove that*

$$C_1 \varepsilon^{-p_1} e^{-\frac{d|\omega_1|}{\sqrt{\varepsilon}}} \leq |\det \Delta^0| \leq C_2 \varepsilon^{-p_2} e^{-\frac{d|\omega_1|}{\sqrt{\varepsilon}}}$$

where C_1, C_2, p_1, p_2 *are appropriate order one constants.*

[f] We denote formal power series identities with the symbol $A \sim B$.
[g] A formal power series $\sum \mu^n a_n(\varepsilon)$ is asymptotic if for all $q > 0$ there exists $Q > 0$ such that for all $n \leq \varepsilon^{-q}$ then $a_n(\varepsilon) \leq \varepsilon^{-Qn}$.

Corollary 1.6. *Under the conditions of Theorem* 2 *the Hamiltonian* (1.1) *has heteroclinic chains, namely a set of* $N \geq 1$ *trajectories* $z^1(t), \ldots, z^N(t)$ *together with* $N+1$ *different minimal sets*[h] $\mathcal{T}_0, \ldots, \mathcal{T}_N$ *such that for all* $1 \leq i \leq N$

$$\lim_{t\to-\infty} \operatorname{dist}(z^i(t), \mathcal{T}_{i-1}) = 0 = \lim_{t\to\infty} \operatorname{dist}(z^i(t), \mathcal{T}_i)\,.$$

Moreover one can construct such chains between tori $\mathcal{T}(\omega^a;\mu)$, $\mathcal{T}(\omega^b;\mu)$ *such that* ω^a, $\omega^b \in \bar{\Omega} \subset \Omega_\gamma$ *and*

$$|\varepsilon^{-\frac{1}{2}}(\omega_n^a - \omega_n^b)| \sim O_\varepsilon(1)\,.$$

The techniques used for proving the Theorems are those proposed in [3] and developed in [4] for partially isochronous three time scale systems with three degrees of freedom. In this paper, particular attention is given to the formalization of the tree expansions and of the "Dyson equation" and relative cancellations proposed in [4]. This enables us to extend Theorem 1 to systems with n degrees of freedom and at least two time scales; moreover the proof is definitely simplified and quite compact. In this article we have considered completely anisochronous systems only to fix an example; generalizing to partially (or totally, thus recovering the results of [8]) isochronous systems is completely trivial. Indeed Theorem 1 and hence Corollary 1.5 can be proved for very general systems, as we will show in a forthcoming paper.

Moreover we have generalized the class of perturbing functions and the "pendulum" (the literature considers only trigonometric polynomials and the standard pendulum); the latter generalizations are quite technical but nevertheless non-trivial and interesting, we think, as the techniques we propose are easily generalizable and give a clear picture of the limits of proving Arnold diffusion via Melnikov dominance.

2. Perturbative Construction of the Homoclinic Trajectories

One can use perturbation theory to find the (analytic for $\mu \leq \mu_0$) trajectories on the S/U manifolds of Hamiltonian[i] (1.3)

$$z(\varphi,\omega,t) = \sum_k (\mu)^k z^k(\varphi,\omega,t)\,.$$

Namely we insert the expansion in μ in the Hamilton equations of system (1.3),

$$\begin{aligned} \dot{I}_j &= -(\mu)\cos\psi_j f(q)\,, & \dot{\psi}_j &= a_j I_j\,, \\ \dot{p} &= \frac{1}{c^2}\sin q(1-(1-c^2)\cos q) - (\mu)\sum_{i=1}^n \sin\psi_i \frac{df}{dq}(q)\,, & \dot{q} &= p\,, \end{aligned} \tag{2.1}$$

[h] A closed subset of the phase space is called minimal (with respect to a Hamiltonian flow ϕ_h^t) if it is non-empty, invariant for Φ_h^t and contains a dense orbit. In our case the minimal sets will be unstable tori $\mathcal{T}(I)$ with $\omega(I)$ Diophantine.

[i] Notice that the apex k on the functions I, ψ represents the order in the expansion in μ NOT an exponent. To avoid confusion, when we need to exponentiate we always set the argument in parentheses.

and find initial data $I(\omega,\varphi,\mu,0^{\pm})$ (and consequently $p(\omega,\varphi,\mu,0^{\pm})$) such that the solution of (2.1) tends exponentially to a quasi-periodic function of frequency ω. Inserting in the Hamilton equations the convergent power series representation:

$$I(t,\varphi,\mu)=\sum_{k=0}^{\infty}(\mu)^k I^k(t,\varphi)\,,\quad \psi(t,\varphi,\mu)=\sum_{k=0}^{\infty}(\mu)^k\psi^k(t,\varphi)\,,$$
$$p(t,\varphi,\mu)=\sum_{k=0}^{\infty}(\mu)^k p^k(t,\varphi)\,,\quad q(t,\varphi,\mu)=q^0(t)+\sum_{k=1}^{\infty}(\mu)^k\psi_0^k(t,\varphi)$$

we obtain, for $k>0$, the hierarchy of linear non-homogeneous equations,[j]

$$\begin{aligned}&\dot I_j^k=F_j^k(\{\psi_i^h\}_{\substack{i=0,\ldots,n\\ h<k}})\,,\quad \dot\psi_j^k=a_jI_j^k\,,\quad \text{for } j=1,\ldots,n\,;\\ &\dot p^k=\frac{1}{c^2}(\cos(q^0(t))-(1-c^2)\cos(2q^0(t)))\psi_0^k+F_0^k(\{\psi_i^h\}_{\substack{i=0,\ldots,n\\ h<k}})\,,\quad \dot\psi_0^k=p^k\,,\end{aligned}\tag{2.2}$$

where the functions F_i^k are defined as follows. Set: $[\cdot]_k=\frac{1}{k!}\frac{d^k}{d\mu^k}(\,\cdot\,)|_{\mu=0}$; we have

$$\begin{aligned}F_j^k(t)=&-\left[\partial_{\psi_j}f^1\left(\sum_{h=1}^{k-1}(\mu)^h\vec\psi^h(t)\right)\right]_{k-1}\\ &-\delta_{j0}\left[\partial_{\psi_0}f^0\left(\sum_{h=1}^{k-1}(\mu)^h\psi_0^h(t)\right)\right]_k\,,\quad j=0,\ldots,n\end{aligned}$$

where $\vec\psi^h(t)$ is the vector $\psi_0^h(t),\ldots,\psi_n^h(t)$,

$$f^1(\vec\psi)=\sum_{i=1}^{n}\sin\psi_i f(\psi_0)\,,\quad f^0(\psi_0)=\frac{1}{c^2}\left((\cos\psi_0-1)+\frac{1-c^2}{2}\sin^2\psi_0\right),$$

finally δ_{ji} denotes the Kronecker delta. For $k=0$ we obtain the unperturbed homoclinic trajectory:

$$z^0(t)=(A^{-1}\omega,p^0(t),\varphi+\omega t,q^0(t))\,,$$

$(q^0(t),p^0(t))$ is the lower branch of the pendulum separatrix starting at $q=\pi$ written in Eq. (1.4).

For $k>0$ we have a linear non-homogeneous ODE that we can solve by variation of constants. The fundamental solution of the linearized pendulum equation is given by,

$$W(t)=\begin{vmatrix}\dot w_0 & \dot x_0^0\\ w_0 & x_0^0\end{vmatrix}\,,\quad w_0=\frac{1}{2}\sigma(t)x_0^1\quad\text{where}\quad\sigma(t)=\operatorname{sign}(t)$$

[j]When it is not strictly necessary we will omit the prefixed initial data of the angles $\varphi=\psi_1(0),\ldots,\psi_n(0)$; $\psi_0(0)=\pi$.

$$x_0^0 = \frac{c^2 \cosh(t)}{c^2 + \sinh(t)^2},$$

$$x_0^1 = \frac{\sigma(t) x_0^0}{2c^4}(2(-3 + 4\,c^2)\,t + \sinh(2\,t) + 4(-1 + c^2)^2\,\tanh(t))\,. \tag{2.3}$$

It is easily seen (see [3] or [5]) that one can choose an appropriate "primitive" in the right hand side of the first column of Eqs. (2.2) so that the solutions are exponentially quasi-periodic.

2.1. *Whisker calculus, the "primitive"* $\Im^t$

Let us first define the function spaces on which we work, all the definitions and statements of this Subsection and of the following one are proposed and explained in detail in [3], we are simply reformulating them to suit our needs.

Definition 2.1. (i) *H is the vector space (on $\mathbb{C}$) generated by monomials of the form*

$$m = \sigma(t)^a \frac{|t|^j}{j!} x^h e^{i(\varphi+\omega t)\cdot\nu} \quad \textit{where } h \in \mathbb{Z}\,, \quad \nu \in \mathbb{Z}^n\,, \quad j \in \mathbb{N}\,,$$

$$x = e^{-|t|}\,, \quad a = 0, 1\,, \quad \sigma(t) = \operatorname{sign}(t)\,. \tag{2.4}$$

(ii) *Given two positive constants b and d, $H(b,d)$ is the subset of functions $f(t)$ analytic on the real axis in $t \neq 0$ that admit, separately for $t > 0$ and $t < 0$, a (unique) representation,*

$$f(t) = \sum_{j=0}^{k} \frac{|t|^j}{j!} M_j^{\sigma(t)}(x, \varphi + \omega t)\,, \tag{2.5}$$

with $M_j^{\sigma(t)}(x,\varphi)$ trigonometric polynomials in φ and the function $M_k^{\sigma(t)}$ not identically zero.

The Fourier coefficients $M_{j\nu}^{\sigma(t)}(x)$ are all holomorphic in the x-plane in a region

$$\{0 < |x| < e^{-b}\} \cup \{|\arg x| < d\}$$

and have possible polar singularities at $x = 0$. k is called the t degree of f.

In Fig. 1 we have represented a possible domain of analyticity for the $M_{\nu j}$. Notice that H is contained in all the spaces $H(b,d)$; moreover if $|t| > b$, $f(t)$ can be represented as an absolutely convergent series of monomials of the type m, separately for $t > b$ and $t < -b$. One can easily check that the functional that acts on monomials m of the form (2.4) as

$$\Im^t(m) = \begin{cases} -\sigma^{a+1} x^h e^{i(\psi+\omega t)\cdot\nu} \displaystyle\sum_{p=0}^{j} \frac{|t|^{j-p}}{(j-p)!(h - i\sigma\omega\cdot\nu)^{p+1}} & \text{if } |h| + |\nu| \neq 0 \\ -\dfrac{\sigma^{a+1}|t|^{j+1}}{(j+1)!} & \text{if } |h| + |\nu| = 0 \end{cases} \tag{2.6}$$

is a primitive of m.

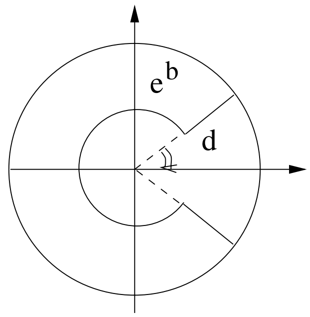

Fig. 1.

We can extend $\Im^t$, with $|t| > b$, to a primitive on functions $f \in H(b,d)$ by expanding f in the monomials m (we obtain absolutely convergent series) and applying (2.6). Then if $|t| \le b$ we set

$$\Im^t \equiv \Im^{2\sigma(t)b} + \int_{2\sigma(t)b}^{t}, \tag{2.7}$$

obviously the choice of $2b$ is arbitrary and this is still the same primitive of f.

In $H(b,d)$ we can extend $\Im^t$ to complex values of t such that $t \in C(b,d)$ where

$$C(b,d) := \{t \in \mathbb{C} : |\mathrm{Im}\ t| \le d,\ |\mathrm{Re}\ t| \le b\} \cup \{t \in \mathbb{C} : |\mathrm{Im}\ t| \le 2\pi,\ |\mathrm{Re}\ t| > b\},$$

is the domain in Fig. 1 in the t variables.

An equivalent (and quite useful) definition of $\Im^t$ is

$$\Im^t f = \oint \frac{du}{2i\pi u} \int_{\sigma(t)\infty + is}^{t} e^{-\sigma(\tau)u\tau} f(\tau)d\tau\,, \tag{2.8}$$

where $\sigma(t) = \mathrm{sign}(\mathrm{Re}\, t)$, $t = t_1 + is$, with t_1, $s \in \mathbb{R}$ and the integral is performed on the line $\mathrm{Im}\ \tau = s$; finally the integrals in u have to be considered to be the analytic continuation on u from u positive and large.

This definition is clearly compatible with the formal definition given above and one easily sees that $H(b,d)$ is closed under the application of $\Im^t$.

Definition 2.2. *$H_0(b,d)$ is the subspace of $H(b,d)$ of functions that can be extended to analytic functions in $C(b,d)$.*

Notice that f is in $H_0(b,d)$ if it is in $H(b,d)$ and $f(t)$ is analytic at $t = 0$.

Remark 2.3. If $f \in H_0(b,d)$ then generally $\Im f \notin H_0(b,d)$ and has a discontinuity in $t = 0$. For instance if $f \in L_1$ is positive, then

$$\Im(f) := (\Im^{0^-} - \Im^{0^+})f = \int_{-\infty}^{\infty} f \ne 0\,.$$

We can construct operators which preserve $H_0(b,d)$; let $\Im = \Im^{0^-} - \Im^{0^+}$ and

$$\Im^t_+ = \begin{cases} \Im^t & \text{if } t \geq 0 \\ \Im^t - \Im & \text{if } t < 0\,, \end{cases} \qquad \Im^t_- = \begin{cases} \Im^t & \text{if } t \leq 0 \\ \Im^t + \Im & \text{if } t > 0\,. \end{cases}$$

The operator

$$\frac{1}{2}\sum_{\rho=\pm 1} \Im^t_\rho = \Im^t - \frac{1}{2}\sigma(t)\Im$$

preserves the analyticity.

Now let us cite two important properties of $H_0(b,d)$, proved in [3].

Lemma 2.4. *In $H_0(b,d)$ we have the following shift of contour formulas*: $\forall f \in H_0(b,d)$ *and for all* $d > s \in \mathbb{R}$,

$$(i)\ \ \Im f(\tau) = \Im f(\tau + is)\,,$$

$$(ii)\ \ \sum_{\rho=\pm 1} \Im^{t+is}_\rho f(\tau) = \oint \frac{dR}{2i\pi R} \sum_{\rho=\pm 1} \int_{\rho\infty}^t e^{-R\sigma(\tau)(\tau+is)} f(\tau+is)d\tau\,.$$

2.2. *The recursive equations*

One can easily verify that $f^1(\psi_0(t), q_0(t))$ and $f^0(q_0(t))$ are in $H_0(a,d)$ (and bounded at infinity) for some "optimal" values a, d corresponding respectively to the maximal distance from the imaginary axis and the minimal distance from the real axis of the poles of such functions. One can prove by induction, see [3] or [5] for the details, that the solutions of Eqs. (2.2) tend to quasi-periodic functions provided that the initial data are chosen to be:

$$I^k_j(\varphi, \omega, 0^\pm) = \sum_k \mu^k \Im^{0^\pm} F^k_j\,, \quad p(\varphi, \omega, 0^\pm) = \sum_k \mu^k \Im^{0^\pm} x^0_0 F^k_0\,.$$

Moreover one can prove that $F^k_j(\varphi, \omega, t)$ has no constant component. Consequently it is convenient to express the trajectories in terms of the "primitives" $\Im^t$ in the form ($a_0 = 1$):

$$(\mu)^k \psi^k_j(\varphi, t) = (\mu)^k a_j Q^t_j F^k_j + x^0_j G^{1k}_j + x^1_j G^{0k}_j$$

where $x^0_j = 1$, $x^1_j = |t|$ for $j \neq 0$ while the x^i_0 are defined in Eq. 2.3,

$$Q^t_j[f] = \frac{1}{2}(\Im^t_+ + \Im^t_-)[(x^0_j(t)\sigma(\tau)x^1_j(\tau) - x^1_j(t)\sigma(t)x^0_j(\tau))f(\tau)]\,, \quad G^{ik}_j = (\mu)^k \frac{1}{2} a_j \Im x^i_j F^k_j\,.$$

For the proofs of these assertions see [3] or [5].

Notice that by our definitions,

$$I_j(\varphi, 0^-) - I_j(\varphi, 0^+) = 2a_j^{-1} \sum_k G^{0k}_j \equiv 2a_j^{-1} G^0_j\,, \quad 2G^0_0 = p(\varphi, 0^-) - p(\varphi, 0^+)\,.$$

We define the formal power series

$$\sum_k G_j^{lk}(\varphi) \equiv G_j^l(\varphi), j = 0, \ldots, n, \quad l = 0, 1.$$

Notice that by the KAM theorem the G_j^0 are convergent series.

Remark 2.5. (i) We will often use formal power series and in particular formal power series identities, namely identities which hold only at each order k in the series expansion in μ; we will mark such identities with the symbol $A \sim B$.

In Sec. 4.5 we will prove that the formal power series we use are "asymptotic". As a definition of asymptotic power series we will assume that a formal power series $\sum \mu^n a_n(\varepsilon)$ is asymptotic if for all $q > 0$ there exists $Q > 0$ such that, for all $n \leq \varepsilon^{-q}$, $a_n(\varepsilon) \leq \varepsilon^{-Qn}$. This implies that we can control the first ε^{-q} terms provided that $\mu < \varepsilon^Q$.

(ii) It should be stressed that we do not need to prove convergence for all the asymptotic power series involved in a given identity to obtain information on those series which are known to be convergent (by the KAM theorem).

The following Proposition contains some important properties of the operators Q_j all proved in [3].

Proposition 2.6 (Chierchia). (i) *The operators Q_j are "symmetric" on $H(a, d)$:*

$$\Im(f\, Q_j g) = \Im(g\, Q_j f).$$

(ii) *$H_0(a, d)$ is closed under the application of Q_j^t.*
(iii) *The operators Q_j preserve parities and if $f \in H_0(a, d)$ is odd then $\Im f = 0$.*
(iv) *If $F, G \in H(a, d)$ are such that the projection on polynomials, $\pi_P F \cdot G$, has no constant component, then*

$$\Im^{0^\sigma} G(\tau)\partial_\tau F(\tau) = F(0^\sigma)G(0^\sigma) - \Im^{0^\sigma} F(\tau)\partial_\tau G(\tau).$$

Proposition 2.6(iii) immediately implies the following (again proved in [3])

Corollary 2.7. *For all $k \in \mathbb{N}$, $j = 0, \ldots, n$, $i = 0, 1$, the function $G_j^{ik}(\varphi)$ is zero for $\varphi = 0$. In particular the splitting vector is zero for $\varphi = 0$ and the system has an homoclinic point.*

Proof. We proceed by induction; by Proposition 2.6(iii) $G_j^{i\,1}(\varphi = 0) = 0$ as it is the integral of an odd analytic function. Consequently $\psi_j^1(\varphi = 0, t)$ is both odd and in $H_0(a, d)$. Now we suppose that $G_j^{i\,h}(\varphi = 0) = 0$ and $\psi_j^h(\varphi = 0, t)$ is odd and in $H_0(a, d)$ for all $h < k$ and $j = 0, \ldots, n$. The function F_j^k is an odd analytic function of the angles ψ_i $(\partial_{\psi_j} f^\delta)$ computed at $\psi = \sum_{h<k} (\mu)^h \psi_j^h(\varphi = 0, t)$ which is again odd and in $H_0(a, d)$. We can apply Proposition 2.6(iii) so $G_j^{i\,k}(\varphi = 0) = 0$ and $\psi_j^k(\varphi = 0, t)$ is both odd and in $H_0(a, d)$. □

3. Proofs of the Theorems

We define the formal power series:

$$\Delta^a_{i,j} = \partial_{\varphi_i} G^a_j \,, \quad \text{for } j = 1, \ldots, n \,, \quad \delta^a_i = \partial_{\varphi_i} G^a_0 \,, \quad \text{for } a = 0, 1 \,.$$

Notice that such series are known as *a priori* to be convergent only for $a = 0$.

Lemma 3.1. *The stable and unstable manifolds are on the same energy surface so that*

$$\sum_{j=1}^{n} G^0_j(\varphi)(I_j(\varphi, 0^+) + I_j(\varphi, 0^-)) = -G^0_0(\varphi)(p(\varphi, 0^+) + p(\varphi, 0^-)) \,, \tag{3.1}$$

this relation implies that at the homoclinic point $\varphi = 0$,

$$\vec{I}(\varphi = 0, 0^+)\Delta^0 = -\delta^0 p(\varphi = 0, 0^+) \,.$$

Proof. Equation (3.1) are simply the energy conservation at time $t = 0$:

$$(I(\varphi, 0^+), AI(\varphi, 0^+)) + p^2(\varphi, 0^+) = (I(\varphi, 0^-), AI(\varphi, 0^-)) + p^2(\varphi, 0^-) \,,$$

the potential part of the Hamiltonian cancels as the perturbation depends only on the angles. Finally we differentiate in φ and compute at the homoclinic point where $G^0_j = 0$ by Corollary 2.7. □

3.1. *The formal linear equation*

In the recursive construction of ψ_j, and consequently of G^i_j, we have distinguished three "blocks":

$$(0) \;\; x^0_j G^{0k}_j \,, \qquad (1) \;\; x^1_j G^{1k}_j \,, \qquad (2) \;\; (\mu)^k a_j Q^t_j(F^k_j) \,, \tag{3.2}$$

as the G^{ih}_j can be brought out of the integral we can say that ψ^k_j and G^{0k}_j ($j = 1, \ldots, n$) are polynomials in the G^{rh}_l with $l = 0, \ldots, n$, $h = 1, \ldots, k-1$, $r = 0, 1$. This can be seen as a formal power series identity:

$$G^0_j(\varphi) \sim J^0_j(\varphi) + \sum_{r=0,1} \left(\sum_{l=1}^{n} N^{[r]}_{jl}(\varphi) G^r_l(\varphi) + n^{[r]}_j G^r_0 \right)$$
$$+ \textit{ quadratic terms } + \cdots \;\; [r] = |r - 1| \,.$$

Following [4] we differentiate this relation in the parameter φ and evaluate it on the homoclinic point where $G^i_j \sim 0$, this leads to a linear formal identity for Δ^0:

$$\Delta^0 \sim D^0 + N^1\Delta^0 + N^0\Delta^1 + n^1\delta^0 + n^0\delta^1 \tag{3.3}$$

where $D^0_{ij} = \partial_{\varphi_j} J^0_i|_{\varphi=0}$.

Notice that we do not have an explicit expression for the matrices N^i and n^i although we have a recursive algorithm for the coefficients of the series expansion; we will use trees to find such explicit expressions. We can notice however that D^0

is the holomorphic part of the splitting matrix, namely it is obtained by using only the holomorphic block (2) of (3.2) in the construction of the homoclinic trajectory.

We insert the energy conservation relations in Eq. (3.3):

$$\left(1 - N^1 + \frac{1}{p(\varphi = 0, 0^+)} n^1 \vec{I}(\varphi = 0, 0^+)\right)\Delta^0 \sim D^0 + N^0\Delta^1 + n^0\delta^1 \,. \tag{3.4}$$

The tree representation of the trajectories leads to the following Propositions all proved in the next Sections:

Proposition 3.2. *The following formal power series relations hold*:

$$\text{(i) } D^0 \sim N^0 \,, \quad \text{(ii) } n^0 \sim \frac{c}{2} D^0 \omega \,,$$

(iii) D_0 *is the Hessian of a function* S^0 *at the homoclinic point:* $D^0_{ij} = \partial_{\varphi_i}\partial_{\varphi_j} S^0(\varphi)|_{\varphi=0}$.

This Proposition generalizes to Hamiltonian (1.3) similar results of [4].

Relations (i) and (ii) inserted in (3.4) directly imply that:

$$\left(1 - N^1 + \frac{1}{p(\varphi = 0, 0^+)} n^1 \vec{I}(\varphi = 0, 0^+)\right)\Delta^0 \sim D^0\left(1 + \Delta^1 + \frac{1}{2}\omega\delta^1\right). \tag{3.5}$$

Proposition 3.3. *One can use the tree representation to find appropriate bounds on the order k terms of the series expansion of the formal power series of Eq.* (3.5). *If we denote by* M^k *the order k term of the* μ *expansion of a formal power series* M, *we have*:

$$\max(|N^{1k}|, |p^k(\varphi = 0, 0^+)|, |I^k_j(\varphi = 0, 0^+)|, |\delta^{1k}|, |\Delta^{1k}|) \le (k!)^{c_1}(C\varepsilon^{-1})^k \,.$$

Moreover the Fourier coefficients of the function S^{0k}:

$$S^{0k}(\varphi) = \sum_{\nu \in \mathbb{Z}^n : |\nu| \le k} e^{i\varphi\cdot\nu} \hat{S}^{0k}(\nu) \,,$$

respect the inequalities

$$|\hat{S}^{0k}(\nu)| \le \begin{cases} (k!)^{c_1}(C\varepsilon^{-\frac{p+7}{2}})^k e^{-|\omega\cdot\nu|d} \\ (k!)^{c_1} C^k \varepsilon^{-k} e^{-|\omega\cdot\nu|c} \end{cases} ,$$

where C, $c < d$ *are appropriate order one constants,* $c_1 = 4\tau + 4$ (τ *is the Diophantine exponent of* ω) *finally* p *is the degree of the pole nearest to the real axis of* $f(q^0(t))$.

Proof of Theorem 1. Proposition 3.3 implies that the formal series of relation (3.5) are asymptotic for $N < \varepsilon^{-q}$ with $|\mu| < \mu_0 = \bar{C}\varepsilon^{(4\tau+4)q+1}$. Then the formal relation (3.5) is an equality for the truncated series $M^{\le N}$ (let us call A the formal matrix on the left of Δ^0 and B the one on the right of D^0):

$$A^{\le N}\Delta^{0\le N} = D^{0\le N} B^{\le N} + o\left(\frac{\mu}{\mu_0}\right)^N \,.$$

Both $A^{\leq N}$ and $B^{\leq N}$ are close to identity and so have order one determinants. This proves Theorem 1 and consequently the conjecture posed in [7], namely that the leading order of the splitting determinant is given by its analytic part $\det D^0$. $\square$

Proof of Corollary 1.5. Let us now set $\bar{m} = \tau_F + 1$, where τ_F is the Diophantine exponent of $\omega_1|\omega_1 \cdot \nu_F| \geq \gamma_F|\nu_F|^{-\bar{m}+1}$ $\bar{m} > m$ (m is the number of fast degrees of freedom). We choose $N = C_1\varepsilon^{-\frac{1}{2\bar{m}}}$ (where $C_1 \leq (\gamma_F/|\omega_2|)^{\frac{1}{2\bar{m}}}$) if $\alpha = 0$) so that we can remove the absolute value in $e^{-c|\omega\cdot\nu|}$ and for all frequencies ν such that $\nu_F \neq 0$:

$$|\hat{S}^{0k}(\nu)| \leq (k!)^{c_1}\varepsilon^{-k}e^{-\frac{c\gamma_F}{\sqrt{\varepsilon}(N)^{\bar{m}-1}}+c|\omega_2||\nu|};$$

we can sum on the frequencies $\nu : \nu_F \neq 0$ in

$$D_{ij}^{0k} = \sum_{|\nu|\leq k} \nu_i\nu_j S^{0k}(\nu)$$

with φ_i or φ_j fast.

$$D_{ij}^{0k} \leq (k!)^{c_1}k^3\varepsilon^{-k}e^{-\tilde{c}\varepsilon^{-1/2\bar{m}}}\left(\sum_{0\leq l\leq k} e^{c|\omega_2|l}\right)^{n-m} \leq (k!)^{c_1}(\tilde{C}\varepsilon^{-1})^k e^{-\tilde{c}\varepsilon^{-1/2\bar{m}}}.$$

So we can sum the asymptotic series D^0 for $k \leq N$ and $\nu < \varepsilon^{1+2(\tau+1)/\bar{m}}$,

$$\min\left(|\det D^{0\leq N}|, \left(\frac{\mu}{\mu_0}\right)^N\right) \leq Ce^{-\tilde{c}\varepsilon^{-1/2\bar{m}}}.$$

Finally we can take any $\bar{m} > m$ and $\tau > n-1$ so we choose $\bar{m}^{-1} = m^{-1} - (\log(\varepsilon^{-\frac{1}{2}}))^{-1}$ (similarly $\tau + 1 = n + (\log(\varepsilon^{-\frac{1}{2}}))^{-1}$) so that $\varepsilon^{-1/2\bar{m}} = e^{-1}\varepsilon^{-1/2m}$ and $\varepsilon^{1+2(\tau+1)/\bar{m}} \geq C\varepsilon^{1+2n/m}$ for some order one C. $\square$

If we have only one fast variable we can give better bounds on $\det D^0$, namely we use

$$|\hat{S}^{0k}(\nu)| \leq (k!)^{c_1}(C\varepsilon^{-\frac{p+7}{2}})^k e^{-|\omega\cdot\nu|d}$$

and the fact that for one fast frequency

$$|\omega\cdot\nu| \geq \frac{|\omega_1|}{\sqrt{\varepsilon}}|\nu_1| - \varepsilon^\alpha|\omega_2||\nu_S|,$$

provided that $\nu_1 \neq 0$ and $N \leq c\varepsilon^{-\frac{1}{2}}$ (with $c < |\omega_1|/|\omega_2|$ if $\alpha = 0$), so by summing up the formal power series we have that:

$$D_{ij}^{0\leq N} = D_{ij}^{01} + \sum_{k=2}^{N} D_{ij}^{0k},$$

where if $|\mu| \leq \tilde{C}\varepsilon^{\frac{p+7}{2}+2n}$:

$$\sum_{k=2}^{N} D_{ij}^{0k} \leq \sum_{k=2}^{N}(\mu\tilde{C}\varepsilon^{-\frac{p+7}{2}-2n})^k \sum_{0<\nu_1<k}[e^{-\frac{|\omega_1|d\nu_1}{\sqrt{\varepsilon}}}] \leq (\mu\tilde{C}\varepsilon^{-\frac{p+7}{2}-2n})^2[e^{-\frac{|\omega_1|d}{\sqrt{\varepsilon}}}] \tag{3.6}$$

(the term in square brackets appears only if $i = 1$ or $j = 1$). So to prove Theorem 2 we have to show that for $\mu \leq \varepsilon^P$ the first order $D_{ij}^{0,1}$ dominates.

3.2. *Lower bounds on the Melnikov term*

The first order of D_0 is

$$D_{ij}^{01} = -\delta_{ij}\ \mathrm{Im}\left[\int_{-\infty}^{\infty} e^{i\omega_j t}(f(q_0(t)) - f(0))\right],$$

namely the integral of an even, analytic, exponentially decreasing function. If $j \neq 1$ we bound this integral with an order one constant.

Lemma 3.4. *The singularities of* $F(t) = f(q^0(t))$ *come in groups of eight* (*in* $|\mathrm{Im}\ t| \leq \pi$); *namely if* t_0 *is a singularity so are*

$$-t_0\,, \quad \pm\bar{t}_0\,, \quad \pm t_0 + i\pi\,, \quad \pm\bar{t}_0 + i\pi\,.$$

The residues of $f(q_0(t))$ *at such points are related in particular if the Laurent series of* F *at* t_i *is*

$$\sum_{k\geq -p} g_k(t_i)(t - t_i)^k\,,$$

then $g_k(t_i) = -(-1)^k \bar{g}_k(-\bar{t}_i)$.

Proof. We are simply using the fact that $f(q)$ is real and odd and that $z = e^{iq(t)}$ has two preimages t and $-t + i\pi$. This implies in particular that $F(t) = -F(-t) = \bar{F}(t)$.[k] □

In the assumptions of Theorem 2 we have imposed that there is one[l] couple $(t_0, -\bar{t}_0)$ of poles closest to the imaginary axis coming from $f(q_0(t))$ rather than $f^0(q_0(t))$. Then $f(0) = 0$ as f is odd, and by definition $f(q(t))$ has two poles on the line $|\mathrm{Im}\ t| = d$, so if $\omega_1 > 0$ we shift the integration to a line $\mathrm{Im}\ t = l > d$, not ε close to any singularity (if $\omega_1 < 0$ we shift to $\mathrm{Im}\ t = -l < -d$).

$$\begin{aligned}\left|\mathrm{Im}\left[\int_{-\infty}^{\infty} e^{i\omega_1 t} f(q_0(t))\right]\right| &\geq 2\pi|\mathrm{Re}[\mathrm{Res}(e^{i\frac{\omega_1}{\sqrt{\varepsilon}}t}(f(q_0(t)), t_0) \\ &\quad + \mathrm{Res}(e^{i\frac{\omega_1}{\sqrt{\varepsilon}}t}(f(q_0(t)), -\bar{t}_0)]| \\ &\quad - e^{-\frac{|\omega_1|}{\sqrt{\varepsilon}}l}\left|\mathrm{Im}\left[\int_{-\infty}^{\infty} e^{i\frac{\omega_1}{\sqrt{\varepsilon}}t}(f(q_0(t + il)) - f(0))\right]\right|,\end{aligned} \tag{3.7}$$

the last integral is again the integral of a bounded ε independent function so we bound it by an order one constant. The residue at the poles can be computed:

$$\sum_{k=1,p} \frac{(i\omega_1)^{k-1}}{(k-1)!\varepsilon^{(k-1)/2}}(g_k(t_0) - (-1)^k \bar{g}_k(t_0))$$

[k] The symbol $\bar{f}(z) := \overline{f(\bar{z})}$.

[l] Naturally we could deal with any finite number of poles with this property.

which is real and generally greater than

$$Ce^{-\frac{|\omega_1|}{\sqrt{\varepsilon}}d}\varepsilon^{-\frac{p-1}{2}}\,. \tag{3.8}$$

Proof of Theorem 2. We choose $|\mu| \le \varepsilon^{p/2+8+4n}$ so that (3.8) dominates on (3.6). $\square$

3.3. *Heteroclinic intersection for systems with one fast frequency*

In the following we will consider systems with one fast frequency and in the *a priori* stable variables of Hamiltonian (1.1). We can fix $\mu = \varepsilon^P$ and ensure Melnikov dominance, as discussed in the previous sections. This means that we have lower and upper bounds on the splitting determinant (and on the eigenvalues of the splitting matrix) of the type:

$$a\varepsilon^p e^{-c\varepsilon^{-\frac{1}{2}}} \le \det \Delta^0(\omega) \le b\varepsilon^{-p}e^{-c\varepsilon^{-\frac{1}{2}}}\,.$$

The coefficients p, a, b, c depend on the perturbing function f.

We consider the function:

$$\begin{aligned} F(\varphi,\omega_0,\omega) &= \tilde{I}^-_\mu(\varphi,\omega,\rho(\omega)) - \tilde{I}^+_\mu(\varphi,\omega_0,\rho(\omega_0)) \\ &\equiv c\sqrt{\varepsilon}(I^-_\mu(\varphi,\omega,\rho(\omega)) - I^+_\mu(\varphi,\omega_0,\rho(\omega_0))) \end{aligned}$$

where ω, $\omega_0 \in \Omega_\gamma$. Notice that

$$F(0,\omega_0,\omega_0) = 0\,, \quad \det \frac{\partial F}{\partial \varphi}(0,\omega_0,\omega_0) = 2^n\varepsilon^{n/2}\det\Delta^0(\omega_0)\,.$$

Hence from the implicit function theorem there exists a function $\varphi(\omega,\omega_0,\varepsilon)$ for which

$$F_\mu(\varphi(\omega,\omega_0,\varepsilon),\omega,\omega_0) \equiv 0\,,$$

provided $|\omega-\omega_0|$ is small enough. Fixed ω_0 standard computations (see [5]) show that the smallness condition is

$$|\omega-\omega_0| \le C\varepsilon^{-2p}e^{-2c\varepsilon^{-\frac{1}{2}}}\,.$$

To prove the existence of heteroclinic intersections, we have to prove the existence of a chain of KAM tori at distances of order $B = O_\varepsilon(e^{-C\varepsilon^{-\frac{1}{2}}})$ for some $C > 2c$, namely we have to adapt to our anisotropic setting (one fast and many slow time scales) the classical techniques discussed in detail in [2] or [5].

Proposition 3.5. *There exists a list of Diophantine frequencies* $\omega_1,\ldots,\omega_h \in \Omega_\gamma$ *such that*:

$$\text{(i) } \sqrt{\varepsilon}|\omega_i-\omega_{i+1}| \le e^{-C_1\varepsilon^{-\frac{1}{2}}} \quad \text{(ii) } \varepsilon^{-\frac{1}{2}}|\Pi_n(\omega_1-\omega_h)| \sim O_\varepsilon(1)\,, \tag{3.9}$$

where Π_n *is the projection on the nth component. To each of the frequencies* ω_i *is associated a preserved unstable invariant torus of Hamiltonian* (1.1), $\mathcal{T}(\omega_i,\rho_i)$

(with $\rho_i \in [-\frac{1}{2}, \frac{1}{2}]$*) of frequency* $\sqrt{\varepsilon}\rho_i\omega_i$*. The scaling factor* ρ_i *is chosen so that all the invariant tori are on the same energy surface, as explained in Remark* 1.3.

To prove the Proposition we proceed in two steps:

(1) Define an appropriate set $\bar{\Omega}$ of Diophantine frequencies respecting condition (3.9).
(2) Prove the existence of unstable KAM tori of frequency: $\sqrt{\varepsilon}\rho\omega$ for $\rho \in [-\frac{1}{2}, \frac{1}{2}]$ and $\omega \in \bar{\Omega}$. We will only sketch the proof of this second point.

Definition 3.6. *Given an order one* $C_1 > 2c$*, set* $A_1 = e^{-C_1\varepsilon^{-\frac{1}{2}}}$ *and consider the set*:

$$\bar{\Omega} := \left\{ \omega \in \Omega : \begin{cases} (a) \quad \sqrt{\varepsilon}|\omega \cdot l| \geq \dfrac{A_1}{|l|^\tau} & \forall\, l \in \mathbb{Z}^n/\{0\} : l_1 \neq 0 \\ (b) \quad \sqrt{\varepsilon}|\omega \cdot l| \geq \dfrac{\varepsilon^2}{|l|^\tau} & \forall\, l \in \mathbb{Z}^n/\{0\} : l_1 = 0 \end{cases} \right\}.$$

As there is only one fast time scale the condition $\omega \in \Omega$ can be given only on the slow variables, while the fast variable is obtained by "energy conservation" $\omega \in \Sigma$ (Σ is the ellipsoid of Definition 1.1), namely we consider a function $F : \mathbb{R}^{n-1} \to \Sigma$:

$$F(x) := \left\{ \sqrt{2E - \sum_{i=2}^{n-1} x_i^2 - \varepsilon^{-1}x_n^2}\,, \quad x_2, \ldots, x_n \right\},$$

so that given $\beta = \frac{1}{2} + a$ ($\frac{1}{2} \leq \beta \leq 1$) and R, r, R_1, r_1, r_2, appropriate order one constants[m] and defining:

$$\tilde{\Omega} := \{\tilde{\omega} \in \mathbb{R}^n : \tilde{\omega}\varepsilon^{-\frac{1}{2}} \in \Omega\}\,, \quad \text{we have } \tilde{\Omega} = F(B(R,r) \cap M)$$

where $B(R,r) \subset \mathbb{R}^{n-1}$ is the spherical shell[n] of radiuses $\varepsilon^\beta R$, $\varepsilon^\beta r$ and

$$M := \{\omega \in \mathbb{R}^{n-1} : \varepsilon r_1 \leq \omega_n \leq \varepsilon R_1\,, \quad \omega_i > r_2\varepsilon^\beta\,, \quad i = 2, \ldots, n-1\}\,.$$

As we always deal with $\tilde{\omega} = \sqrt{\varepsilon}\omega$ we will omit the tilde rescaling all the relations. The Jacobian of F in $B(R,r) \cap M$ is bounded from above and below by order one constants so that given a measurable set[o] $S \subset \Omega$ $\text{meas}(F^{-1}(S)) \sim \text{meas}(S)$.

Condition (b) naturally defines subsets of $B(R,r) \cap M$. Moreover we can project the set respecting condition (a) on the subspace of the slow variables. Call this set $\bar{\Omega}_4 \subset B(R,r) \cap M$.

Let us call $S(x)$ the $(n-2)$-dimensional sphere centered in the origin and of radius $\varepsilon^\beta x$. We take $2r < R$ and consider $\bar{R}$ so that

$$R_1/2 < \bar{R} < R_1\,, \quad \frac{r}{R} > \frac{r_1}{\bar{R}}\,. \tag{3.10}$$

[m] This condition automatically imply $\bar{r} \leq \sqrt{\varepsilon}\omega_1 \leq \bar{R}$, notice that we are not using the same notation as in (1.1), here ω_i is always the ith component of ω.
[n] We call spherical shell of radiuses b, a the $(n-1)$-dimensional domain $\{x \in \mathbb{R}^{n-1} : a \leq |x| \leq b\}$.
[o] The symbol $\sim$ means that the two measures are of the same order in ε.

Definition 3.7. *Consider the sets*

$$S_2 := \{\omega \in S(R) : \varepsilon(R_1 - (R_1 - \bar{R})/4) \le \omega_n \le \varepsilon(\bar{R} + (R_1 - \bar{R})/4),\ \omega_i \ge r_2\varepsilon^\beta\,, \\ \forall\, i \neq n\}\,,$$

$$S_3 := \{\omega \in S(R) : \varepsilon R_1 \le \omega_n \le \varepsilon \bar{R}\,, \quad \omega_i \ge r_2\varepsilon^\beta\,, \quad \forall\, i \neq n\}\,.$$

$M \cap S(R) \supset S_3 \supset S_2$; and the sets all have measure of order $\varepsilon^{(n-3)\beta+1}$.

Given a set $X \in S(R)$, its cone $\mathcal{C}(X)$ is the set of semilines stemming from the origin and reaching points of X. We consider truncated cones $T(X) := \mathcal{C}(X) \cap B(R, r)$, and, for any $r < a < b < R$, $T_{a,b}(X) = T(X) \cap B(b, a)$.

Notice that by (3.10) *if $X \in S_3$, then $T(X) \in M \cap B(R, r)$.*

Remark 3.8. Recall that given a measurable set $X \in S(R)$, the cone of X is measurable and meas$T(X) \sim \varepsilon^\beta$ meas(X), meas $T_{a,b}(X) \sim \varepsilon^\beta(b-a)$ meas(X).

Definition 3.9. *Given $A_2 = e^{-C_2\varepsilon^{-\frac{1}{2}}}$ with $2c < C_2 < C_1$ and for all $s \in \mathbb{R}$, $1 < s < 4R/r$, we consider the sets*:

$$\bar{\Omega}_2(s) = \left\{\omega \in B(R, r) : |\omega \cdot l| \ge \frac{s\varepsilon^2}{|l|^\tau} \quad \forall\, l \in \mathbb{Z}^{n-1}/\{0\}\ |l| \le A_2^{-1}\right\},$$

$$\Omega_3(s) = \left\{\omega \in B(R, r) : |\omega \cdot l| \ge \frac{s\varepsilon^2}{|l|^\tau} \quad \forall\, l \in \mathbb{Z}^{n-1}/\{0\}\right\}.$$

Remark 3.10. Standard measure theoretic arguments imply that the sets $(\bar{\Omega}_i(s) \cap S(R))^C \cap S(R)$ all have measure of order $\varepsilon^{(n-3)\beta+2}$; this implies as well that $(\bar{\Omega}_i(s) \cap S_2)^C \cap S_2$ has measure of the same order and the same holds for intersections with S_3 and for $(\bar{\Omega}_2(s) \cap \bar{\Omega}_3(s) \cap S_2)^C \cap S_2$. We will repeatedly use such relations.

Lemma 3.11. (i) *Given a point $\omega \in \bar{\Omega}_2(2R/r) \cap S_2$, the whole solid ball $B_\rho(\omega)$ of center ω and radius $\rho = \varepsilon^2 A_2^{1+\tau}$ is contained in $\bar{\Omega}_2(R/r)$ and its intersection with $S(R)$ is contained in S_3.*

(ii) *The whole truncated cone $T(\bar{\Omega}_2(R/r) \cap S_3)$ is in $\bar{\Omega}_2(1)$, same for $\bar{\Omega}_3$.*

Proof. (i) First notice that any $(n-2)$-dimensional "ball", $B_\rho(x) \cap S(R) \in S_3$ if $x \in S_2$. Now consider $\omega \in \Omega_2(2R/r) \cap S_2$ and a vector $x \in \mathbb{R}^{n-1}$ on the unit sphere:

$$|(\omega + \rho x) \cdot l| \ge ||\omega \cdot l| - |l|\rho| \ge |\omega \cdot l| \left(\left|1 - \rho\frac{|l|}{|\omega \cdot l|}\right|\right), \quad \text{as } \frac{|l|}{|\omega \cdot l|} \le \frac{r|l|^{\tau+1}}{2R\varepsilon^2}$$

and $|l| \le A_2$, setting $\rho = \varepsilon^2 A_2^{1+\tau}$ we have $0 < \rho\frac{|l|}{|\omega \cdot l|}|)\frac{1}{2}$.

(ii) Given a point $x \in \Omega_3(R/r) \cap S(R)$ (or in $x \in \Omega_2(R/r) \cap S(R)$) then $rx/R \in S(r)$. Moreover for $r/R \le t \le 1$:

$$|tx \cdot l| = t|x \cdot l| \ge r/R\frac{R\varepsilon^2}{r|l|^\tau} = \frac{\varepsilon^2}{|l|^\tau}\,. \qquad \square$$

Lemma 3.12. *The set $\bar{\Omega}_2(R/r) \cap S(R)$ is union of a finite number of disjoint convex domains. Each domain is contained in a $(n-2)$-dimensional "ball" of radius $C_3\varepsilon^\beta A_2$ for an appropriately fixed order one C_3.*

Proof.

$$
\begin{aligned}
&(\bar{\Omega}_2(R/r) \cap S(R)) \\
&\quad \equiv S(R) \bigcap_{\substack{l \in \mathbb{Z}^{n-1} \\ |l| \le A_2}} \left(\left\{ x \in \mathbb{R}^{n-1} : (x \cdot l) > \frac{R\varepsilon^2}{r|l|^\tau} \right\} \cup \left\{ x \in \mathbb{R}^{n-1} : (x \cdot l) < -\frac{R\varepsilon^2}{r|l|^\tau} \right\} \right),
\end{aligned}
$$

now the intersection of sets such that each connected component is convex has the same property. Suppose, by contradiction, that there are points x_1, $x_2 \in \Omega_2(R/r) \cap S(R)$ such that the arc $\widehat{x_1 x_2}$ is all in $\Omega_2(R/r) \cap S(R)$ and has length greater than $2R^{-1}\sqrt{n}\varepsilon^\beta A_2$. Let $\langle x_1, x_2 \rangle$ be the plane generated by the vectors x_1, x_2, and on it consider the sector $\mathcal{S}$ of unit vectors orthogonal to $\widehat{x_1 x_2}$, this sector has angle $\vartheta = 2\sqrt{n}A_2$. The product space of $\langle x_1, x_2 \rangle^\perp$ with the sector $\mathcal{S}$ is a multi-cylinder in which there cannot be entire vectors $l \in \mathbb{Z}^{n-1}$ with $|l| \le A_2^{-1}$.

Now we consider the intersection of the multi-cylinder with the sphere $|x| = A_2^{-1} - 2\sqrt{n}$, on $\langle x_1, x_2 \rangle$ it is an arc of length greater than $2\sqrt{n}$ so that a ball of radius $\sqrt{n}$ is contained in the multi-cylinder. Now in each ball of radius $\sqrt{n}$ there is at least one entire vector. Namely let x be the center of the ball then $[x]$ (entire part of each component) is entire and $|x - [x]|_\infty \le 1$. □

Let N be the number of connected domains of $\bar{\Omega}_2(R/r) \cap S(R)$ contained in S_3. Each domain contains an $(n-2)$-dimensional "ball" of radius $\rho = \varepsilon^2 A_2^{1+\tau}$, so that $N \le A_2^{-(n-2)(\tau+1)}\varepsilon^{\beta(n-2)-2n+5}$.

Let us now consider the Cantor set $\bar{\Omega}_3(R/r) \cap S_3$, by Remark 3.10 we have that $(\bar{\Omega}_3(R/r) \cap S_3)^C \cap S_3$ has measure of order $\varepsilon^{(n-3)\beta+2}$. This implies that $\bar{\Omega}_3(R/r) \cap S_3 \cap \bar{\Omega}_2(R/r)$ is not empty and the measure of $(\bar{\Omega}_3(R/r) \cap S_3 \cap \bar{\Omega}_2(R/r))^C \cap S_3$ is of order $\varepsilon^{(n-3)\beta+2}$.

Lemma 3.13. *There exists a connected domain D of $\Omega_2(R/r) \cap S_3$ such that*

$$
\mathrm{meas}(D \cap \bar{\Omega}_3(R/r)) \ge A_2^{(n-2)(\tau+1)+1} \,.
$$

Proof. Suppose the assertion to be false, then calling D_i, $i = 1, \ldots, N$ the connected domains:

$$
\mathrm{meas}\, S_3 \sim \mathrm{meas}(\bar{\Omega}_2(R/r) \cap S_3 \cap \bar{\Omega}_3(R/r)) = \sum_{i=1}^{N} \mathrm{meas}(D_i \cap \bar{\Omega}_3) \le A_2^{(n-2)(\tau+1)+1} N
$$

which is absurd. □

Then we can use Lemma 3.11(ii) and consider the truncated cone $T(D) \subset \bar{\Omega}_2(1)$, by Lemma 3.13 $P = T(D) \cap \bar{\Omega}_3(1)$ has measure of order $A_2^{(1+\tau)(n-2)+1}\varepsilon^\beta$; namely the Cantor set P contains all radial segments having an endpoint in $D \cap \bar{\Omega}_3(R/r)$ and the other on $S(r)$.

Consider an $(n-1)$-dimensional ball of radius $\rho \sim \varepsilon^\beta A_2$ centered on a point $x \in D$ and which contains D (such ball exists by Lemma 3.13). Given $h = [\frac{2(R-r)}{3\rho R}]$, consider the points $x_i = t_i x$ with $t_i = 1 - 3/2i\rho$ $h \geq i \in \mathbb{N}_0$ and let us cover $T(D)$ with a finite number of balls B_i of radius ρ and centered on points x_i.

Setting $\rho = 2C_3\varepsilon^\beta A_2$ we have that $B_i \cap B_j$ is empty if $|i-j| > 1$ and each $B_i \cap B_{i+1}$ contains a truncated cone $T_{a_i,b_i}(D)$ with $b_i - a_i \geq \rho/4$. We consider the sets $P_i = T_{a_i,b_i}(D) \cap \Omega_3(1)$, by Lemma 3.13 each P_i has measure of order $\varepsilon^\beta A_2^{(1+\tau)(n-2)+2}$.

Now we consider the Cantor set $\bar{\Omega}_4$ whose complementary set in $M \cap B(R,r)$ has measure of order $\varepsilon^{(n-2)\beta+1}A_1$. Its intersection with P_i has measure of order $\varepsilon^\beta A_2^{(1+\tau)(n-2)+2}$, provided that $A_1 < A_2^{(\tau+1)(n-2)+3}$. Consider a list $\omega_i \in P_i \cap \bar{\Omega}_4$; for each i we have that ω_i, $\omega_{ai+1} \in B_{i+1}$ so the list respects condition 3.9(i) moreover

$$\min_{y \in B_0} y_n \geq \bar{R} - 2C\varepsilon^\beta A_2 \quad \text{and} \quad \max_{y \in B_h} y_n \leq \frac{r}{R}R_1 + 2C\varepsilon^\beta A_2$$

for some order one C so the list respects condition 3.9(ii).

In the Appendix A.2 we have proved, generalizing similar results of [9], that there exists a symplectic transformation, well defined in a region W of the phase space $(\tilde{I}, \psi)$, which sends Hamiltonian (1.1) in the local normal form:

$$\frac{1}{2}(J, AJ) + \sqrt{\varepsilon}G_1(PQ, \sqrt{\varepsilon}) + \mu g_1(\phi_S, J, P, Q) + \alpha f_1(\phi, J, P, Q) \tag{3.11}$$

where $\alpha = O_\varepsilon(e^{-C\varepsilon^{-\frac{1}{2}}})$ for any order one C. W is of order one in the actions both in the fast direction J_1 and in the degenerate one J_n, namely there exists points w_1, $w_2 \in W$ such that $|\Pi_{J_n}(w_1 - w_2)| = O_\varepsilon(1)$. We can then prove a KAM theorem for the Hamiltonian (3.11) for $\mu < \varepsilon^4$ with the frequencies ω in $\bar{\Omega}$ by choosing $(A_1)^2 \ll \alpha$. Roughly speaking, KAM theorems are proved by performing an infinite sequence of symplectic transformations defined in a set of nested domains whose intersection is not trivial. Each approximation step reduces the order of the perturbation quadratically and is well defined provided an appropriate smallness condition is verified. Roughly speaking, such condition is of the type: $\mu\gamma^{-2} \ll 1$ where μ is the small parameter and γ is the Diophantine constant of the frequency ω of the preserved torus. To apply this scheme to Hamiltonian (3.11), we first perform a finite number of approximation steps on the slow variables with J_1 as a parameter; the small denominators involved are $|\omega_S \cdot l|$ on which we have the stronger Diophantine condition so that the approximation scheme works provided that $\mu\varepsilon^{-4} \ll 1$. Eventually we will reduce the μ perturbation to order α and then continue with the classical KAM scheme on all the variables, now the smallness condition is $\alpha A_1^{-2} \ll 1$. This completes the proof of Proposition 3.5.

4. Tree Representation

4.1. *Definitions of trees*

We briefly review the tree representation of the homoclinic trajectory. The definitions contained in this Subsections are all adapted from [10].

Definition 4.1. *A* **graph** *G consists of two sets $V(G)$ (vertices), $\mathcal{E}(G)$ (edges) such that $\mathcal{E}(G)$ is a subset of the unordered pairs of distinct elements of $V(G)$. We will always consider finite graphs, i.e. graphs such that $N(G) = |V(G)|$ is finite. Two vertices i, $j \in V(G)$ are said to be adjacent if $(i,j) \in \mathcal{E}(G)$. It is customary to write $n \in G$ in place of $n \in V(G)$ and $(i,j) \in G$ in place of $(i,j) \in \mathcal{E}(G)$.*

Two graphs G_1, G_2 are equal if and only if they have the same vertex set and the same edge set.

Definition 4.2. *A* **path** *joining the vertices i, $j \in G$ is a subset $\mathcal{P}_{ij}$ of $\mathcal{E}(G)$ of the form*

$$\mathcal{P}_{ij} := \{(i,v_1),(v_1,v_2),\ldots,(v_k,j)\}\,.$$

A graph G is **connected and without loops** *if for all i, $j \in G$, there exists one and only one path that connects them. Such graphs are called* **trees***. Their vertices are called nodes and their edges are called branches.*

A tree T such that the set $V(T) = \{1,2,\ldots,N(T)\}$ is called a numbered tree.

Definition 4.3. *A* **labeled** *tree is a tree A plus a label $\mathcal{L}_A(v) \geq 0$ which is generally a set of functions $f^i_A(v)$ defined on the nodes.*

When possible we will omit the subscript A in the functions f^i.

Definition 4.4. *Two labeled trees X, Y are isomorphic if there is a bijection, say h, from $V(X)$ to $V(Y)$ such that for all $a \in V(X)$, $\mathcal{L}_X(a) \equiv \mathcal{L}_Y(h(a))$. Moreover $(a,b) \in \mathcal{E}(X)$ if and only if $(h(a),h(b)) \in \mathcal{E}(Y)$. We say that h is an isomorphism from X to Y. Notice that since h is a bijection h^{-1} is well defined and is an isomorphism from Y to X. We will call* **symmetries** *or automorphisms of X, the isomorphisms from X to X.*

It is often convenient and more compact to represent a tree by a diagram, with points for the nodes and lines for the branches, as in Fig. 2.

In this diagrams the positions of the points and lines do not matter — the only information it conveys is which pairs of nodes are joined by a branch. This means that the two diagrams in Fig. 2 are equal by definition.

Fig. 2.

Strictly speaking these diagrams do not define graphs, since the set V is not specified. However, if the diagram has N points, we may assign distinct natural numbers $1, 2, \ldots, N$ to the points (which we still call nodes), so obtaining a labeled numbered tree. Then it is easily seen that the two trees in Fig. 2 are isomorphic.

Definition 4.5. *We will call diagrams the equivalence classes of labeled trees via the relation $A \cong B$ if and only if A and B are isomorphic.*

An obvious consequence of this definition is that, $\mathcal{L}_A(v)$ and $N(A)$ are well defined on the equivalence classes.

We can choose a representative A' of the equivalence class A by giving a numbering $1, 2, \ldots, N(A)$ to the nodes of A.

Remark 4.6. Given an equivalence class of labeled trees A and a numbering A', the group of automorphisms of A' can be identified with a subgroup of the group of permutations on $N(A)$ elements $S_{N(A)}$; we denote such subgroup by $\mathcal{S}(A')$.

$\mathcal{S}(A')$ is the subgroup of the permutations $\sigma \in S_{N(A)}$ which fix both $\mathcal{E}(A')$ and the labels $\mathcal{L}(A')$. Namely[p] $\sigma \in \mathcal{S}(A) \to \sigma\mathcal{E} = \mathcal{E}$ and $\mathcal{L}(n) = \mathcal{L}(\sigma(n))$ for all $n \leq N(A)$.

Given two isomorphic trees A' and A'', representatives of A, let h be a bijection such that $\mathcal{E}(A') = \sigma\mathcal{E}(A'')$. The groups $\mathcal{S}(A')$ and $\mathcal{S}(A'') = h^{-1}\mathcal{S}(A')h$ are isomorphic. We will improperly call the equivalence classes via this relation the **symmetry** group $\mathcal{S}(A)$ of the diagram A.

Using standard notation (see for instance [11]) we denote by $a := (i_1, i_2, \ldots, i_m)$ with $\mathbb{N} \ni i_j \leq N(A)$ the permutation such that $a(i_h) = i_{h+1}$, $a(i_m) = i_1$, and $a(n) = n$ for all $\mathbb{N} \ni n \leq N(A)$ such that $n \notin \{i_1, i_2, \ldots, i_m\}$. Moreover we denote by ab the composition of a and b. As an example in Fig. 3 consider the numbered tree A ($N(A) = 6$), its symmetries are the identity and $a := (1, 4)$; $b := (2, 3)$; $c \equiv a \circ b$; $d \equiv (5, 6)(1, 2)(4, 3)$, $e := (5, 6)(1, 3)(2, 4)$; $f := (5, 6)(1, 2, 4, 3)$; $g := f \circ a$. Clearly any other numbering on A would give an isomorphic symmetry group.

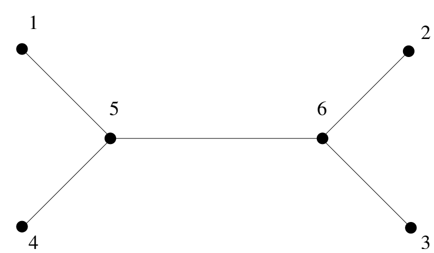

Fig. 3.

[p]With standard abuse of notation we denote $\sigma\mathcal{E}(A')$ the function such that $\sigma(a, b) = (\sigma a, \sigma b)$ for all (a, b) in $\mathcal{E}(A')$.

Definition 4.7. *Given a tree A and a node $v \in A$, we define its* **orbit**:

$$[v] := \{w \in A : w = g(v) \text{ for some } g \in \mathcal{S}(A)\}\,,$$

i.e. the list of nodes obtained by applying the whole group $\mathcal{S}(A)$ to v, notice that this is an equivalence relation (a proof of this statement is in [10]).

In the example of Fig. 3 there are two orbits, which in the chosen numbering are:

$$[1] \equiv \{1, 2, 3, 4\} \text{ and } [5] \equiv \{5, 6\}\,.$$

Remark 4.8. The orbits are well defined on the equivalence classes of labeled trees, it should be clear, for instance, that the nodes signed in black in the diagram of Fig. 4 are an orbit.

Definition 4.9. *A* **rooted** *labeled tree is a labeled tree A plus one of its nodes called the first node (v_A or v_0); this gives a partial ordering to the tree, namely we say that $i > j$ if $\mathcal{P}_{v_0 j} \subset \mathcal{P}_{v_0 i}$. Moreover choosing a first node induces a natural ordering on the couples of nodes representing the branches namely $(a, b) \in \mathcal{E}(A)$ implies that $a < b$. We recall some definitions on rooted trees*:

(a) *the level of v $l(v)$ is the cardinality of $\mathcal{P}_{v_0 v}$*;
(b) *the nodes subsequent to v, $s(v)$, are the nodes adjacent to v and of higher level*; *the node preceding v is the only node adjacent to v and of lower level*;
(c) *given v node of A, we call $A^{\geq v}$ the rooted tree (with first node v) of the nodes $w \geq v$; we call $A^{\setminus v}$ the remaining part of the tree A.*

An isomorphism between rooted trees (A, v_A), (B, v_B) is an isomorphism between A and B which sends v_A in v_B. The symmetries of a rooted labeled tree (A, v_A), which we denote again by $\mathcal{S}(A, v_A)$ are the subgroup of the symmetries of the corresponding unrooted tree that fix the first node v_A. As done for trees, we can represent the equivalence classes of rooted trees with diagrams, representing by convention the first node on the left and all the nodes of the same level aligned vertically (it should be obvious that the definitions $v > w$, $A^{\setminus v}$ and $A^{\geq v}$ are well posed on the equivalence classes).

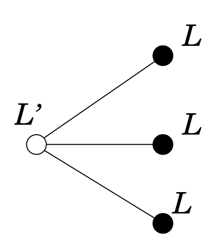

Fig. 4.

4.2. *Admissible trees*

Definition 4.10. *We consider rooted labeled trees such that some nodes are distinguished by having a different set of labels.*[q] *An* **admissible** *tree is a symbol*:

$$A, \{v_A\}, \{v_1, \dots, v_m\}, \{w_1, \dots, w_h\}$$

such that A is a tree, all the v_i, w_j and v_A are nodes of A, the v_i are all end-nodes,

$$\{v_i\}_{i=1}^m \cap \{w_j\}_{j=1}^h = \varnothing$$

and the v_i are all different.

We call $\{v_i\}_{i=1}^m \equiv \mathcal{F}(A)$ the fruits of A, $\{w_j\}_{j=1}^h \equiv \mathcal{M}(A)$ the marked[r] *nodes of A and the set*

$$\overset{0}{A}\colon \{v \notin \mathcal{F}(A)\}$$

the free nodes of A. Finally $s_0(v)$ are the free nodes in $s(v)$.

The labels are distributed in the following way:

(a) *For each node $v \neq v_A$, one angle label $j_v \in \{0, \dots, n\}$ (remember that we are considering a system with $n+1$ degrees of freedom).*
(b) *For each node v, one order label $\delta_v = 0, 1$ if $v \in \overset{0}{A}$ and $\delta_v \in \mathbb{N}$ otherwise.*
(c) *For each node $v \in \mathcal{M}(A)$, one angle-marking $J = 0, \dots, n$ and one function-marking $h(t) \in H$.*
(d) *For each node $v \in \mathcal{F}(A)$, one type label $i = 0, 1$.*

We set a grammar on the so defined labeled rooted trees, namely:

$$\delta_v = 0 \to \{j_v = J_v = 0,\ |s(v)| \geq 2,\ j_{v'} = 0\ \forall\, v' \in s(v)\}\,.$$

To draw the diagrams without writing down the labels we give a color to each $j = 1, n$ (which forces $\delta = 1$) and two different colors for the couples of labels $j = 0$, $\delta = 1$ and $j = 0$, $\delta = 0$.

In all the pictures we will set $n = 1$ and choose the colors gray, black and white, see Fig. 5. The fruits $\mathcal{F}(A)$ will be represented as "bigger" end-nodes colored with the color corresponding to their angle label and with their order and type written on a side. The marked nodes will be distinguished by a box of the color corresponding to their angle-marking and with their function-marking written on a side. If the function marking is $h(t) = 1$ we will omit the function marking. By convention the first node is set on the left, and the nodes of the same level are aligned vertically.

Definition 4.11. (1) *We will call fruitless trees the (labeled rooted trees) A such that $\mathcal{F}(A)$ is empty. We will say that a fruit v stems from w if $v \in s(w)$.*

(2) *We will call $\mathcal{T}$ the set of equivalence classes (as in Definition 4.5) of admissible trees, $\overset{0}{\mathcal{T}}$ the subset of $\mathcal{T}$ of trees with at least a free node and $\mathcal{A}$ the subset of*

[q] The dynamical meaning of the labels will be clear when we define the "value" of a tree.
[r] A node v can appear many times in $\mathcal{M}(A)$ we will say it carries more than one marking.

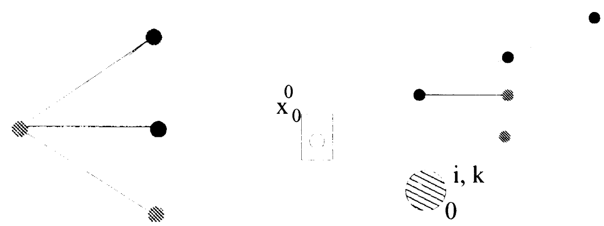

Fig. 5. Examples of trees in $\mathcal{A}^5$ and in $(\mathcal{T}_0^0)^5$ (see Definition 4.13).

$\overset{0}{\mathcal{T}}$ of "fruitless" trees. Finally we will call $\overset{m}{\mathcal{A}}$ the subset of $\mathcal{A}$ of fruitless trees with no marking.

(3) We will call $\mathcal{F}_j^{ik}$ the "tree" composed of one fruit of order k, angle j and type i; clearly

$$\mathcal{T} \equiv \mathcal{T} \bigcup_{\substack{i=0,1 \\ j=0,\dots,n \\ k>0}} \mathcal{F}_j^{ik} .$$

Notational Convention 1. *Using standard notation we represent the equivalence classes by $[A]$ where A is an admissible tree.*

Moreover given a tree A we will write $A \in \mathcal{T}$ if it is a representative of an equivalence class in $\mathcal{T}$.

Definition 4.12. *The order of a tree $A \in \mathcal{T}$ is:*

$$o(A) = \sum_{v \in A} \delta_v .$$

The order of a node v of A is $o(v) = o(A^{\geq v})$.

Given a tree $A \in \overset{0}{\mathcal{T}}$ and one of its nodes v we call $A^{\geq v}$ the tree composed of the nodes greater or equal to v; if $A^{\geq v}$ is not a fruit then it is not admissible as it carries a label j in the first node. In such case, we conventionally set $A^{\geq v} \in \mathcal{T}$ by setting a mark $J(v) = j_v$, $h(v,t) = 1$ on v and subsequently "forgetting" the label j_v.

It is easily seen that $o(A) > 0$ for all $A \in \mathcal{T}$ and that

$$\mathcal{T}^k \equiv \{A \in \mathcal{T} \text{ t.c. } o(A) = k\}$$

is a finite set; clearly the same is true in $\overset{0}{\mathcal{T}}$ and in $\mathcal{A}$.

Notational Convention 2. *In all our sets an apex k means we consider the subset of trees of order k.*

We list here the subsets of $\overset{0}{\mathcal{T}}$ and $\mathcal{A}$ that we will need in the following sections.

Definition 4.13. (a) *$\mathcal{A}_j^a$ $(\mathcal{T}_j^a)$ with $j = 0, \dots, n$, $a = 0, 1$, is the subset of $\mathcal{A}$ $(\overset{0}{\mathcal{T}})$ such that $\mathcal{M}(A) \equiv \{v_A\}$ and $J(v_A) = j$, $h(v_A, t) = x_j^a(t)$.*

(b) $\mathcal{A}_{ij}^{ab}$ $(\mathcal{T}_{ij}^{ab})$, *with* $i,j = 0,\dots,n$, $a,b = 0,1$, *is the subset of* $\mathcal{A}$ $(\overset{0}{\mathcal{T}})$ *such that* $\mathcal{M}(A) \equiv \{v_A, v\}$ *for some* $v \in A$ *moreover* $J(v_A) = i$, $h(v_A,t) = x_i^a$, $J(v) = j$, $h(v,t) = x_j^b$.

Given a set S one can consider a vector space on $\mathbb{Q}$ generated by formal linear combinations of the elements of the set; we represent it by $\mathbb{V}(S)$.

Definition 4.14. $\mathbb{V}(S)$ *is the vector space of linear combinations of elements of* S *with rational coefficients.*

$$[A] \in S \to [A] \in \mathbb{V}(S)\,, \quad [A],[B] \in \mathbb{V}(S) \to q_1[A] + q_2[B] \in \mathbb{V}(S)\,, \quad \forall q_1, q_2 \in \mathbb{Q}\,.$$

We construct $\mathbb{V}(S)$ for the sets in Definition 4.13, we obtain infinite dimensional vector spaces that can be expressed as direct sum of finite dimensional spaces generated by the sets S^k (we call these spaces $\mathbb{V}^k(S)$).[s]

Definition 4.15. *In particular, we will be interested in the following vectors:*

$$\mho^k = \frac{1}{k} \sum_{\substack{A \in (\overset{m}{\mathcal{A}})^k \\ \delta_{v_A} = 1}} \frac{A}{|\mathcal{S}(A)|}\,, \quad \Lambda_i^{ak} = \sum_{A \in (\mathcal{T}_i^a)^k} \frac{A}{|\mathcal{S}(A)|}\,,$$

$$\mho_i^{ak} = \sum_{A \in (\mathcal{A}_i^a)^k} \frac{A}{|\mathcal{S}(A)|}\,, \quad \mho_{ij}^{abk} = \sum_{A \in \mathcal{A}_{ij}^{abk}} \frac{A}{|\mathcal{S}(A)|}\,,$$

where the sum $A \in S^k$ *means choosing one representative* A *for each equivalence class* (*diagram*) *of the set* S^k. *Clearly the vectors are determined only up to isomorphisms. The same vectors without the apex* k *will represent the formal series*[t]: $V = \sum_{k=1}^{\infty} V^k$.

4.3. *Values of trees*

We link the vectors defined in Definition 4.15 to the dynamics by defining an appropriate tree "value" $\mathcal{V}(A)$ where $A \in \mathcal{T}$. This definition can be extended to diagrams provided that $\mathcal{V}(A) = \mathcal{V}(B)$ if A and B are isomorphic, moreover we can uniquely extend $\mathcal{V}$ to a linear function on $\mathbb{V}(\mathcal{T})$. The presentation is very schematic as this definitions can be found in [3] and following papers; let us only write the F_j^k explicitly (using well known formulas on the derivatives of composite functions), e_j the vectors of the canonical basis:

$$F_j^k = -\sum_{\delta=0,1} \sum_{\vec{m} \in \mathbb{N}_0^n} (\nabla^{\vec{m}+e_j} f^\delta(t)) \sum_{\{p_j^h\}_{\vec{m},k-\delta}} \prod_{\substack{j=0 \\ h=1}}^{n,k-1} \frac{1}{p_j^h!} (\psi_j^h)^{p_j^h}$$

[s]We are using the fact that the sets are disjoint union of the corresponding "fixed order" sets S^k.
[t]Remember that the apex k is NOT an exponent.

where $\{p_j^h\}_{\vec{m},k}$ is a list of numbers in $\mathbb{N}_0 \equiv \mathbb{N} \cup \{0\}$ which respect the relations

$$\sum_h p_j^h = m_j \,, \quad \sum_{j,h} h p_j^h = k \,, \quad \text{finally we define } \nabla^{\vec{m}} f(t) = \left[\prod_{j=0}^{n} \partial_{\psi_j}^{m_j} f(\psi) \right]_{\substack{\psi_i = \varphi_i + \omega_i t \\ \psi_0 = q_0(t)}} .$$

So we define

$$\mathcal{V}_\varphi(A) = \prod_{v > v_0} (\Im_+^{\tau_w} + \Im_-^{\tau_w}) \Psi_\varphi(A)$$

where

$$\Psi_\varphi(A) = \prod_{\substack{v \in \mathcal{A}^0 \\ v > v_0}} w_{j_v}(\tau_w, \tau_v) \prod_{v \in A^0} \left(-\frac{1}{2} a_{j_v} \right) \mu^{\delta_v} \nabla^{\sum_{j=0}^n m_v(j) e_j} f^{\delta_v}$$
$$\times \prod_{\alpha \in \mathcal{F}(v)} x_{j_\alpha}^{[i_\alpha]} \prod_{\beta \in \mathcal{M}(v)} h_\beta(v, \tau_v) \prod_{\alpha \in \mathcal{F}(A)} G_{j(\alpha)}^{o(\alpha), i(\alpha)} .$$

$\mathcal{F}(v)$ are the fruits stemming from v, $\mathcal{M}(v)$ is the list of markings of the node v, w is the node preceding v and finally $m_v(j)$ is the number of elements in $\{v, s_0(v), \mathcal{F}(v), \mathcal{M}(v)\}$ having angle label (or angle marking) equal to j. We write $s_0(v)$, $\mathcal{F}(v)$ instead of $s(v)$ to remark that the fruits are not considered proper nodes. Notice that $\Psi_\varphi(A)$ contains the kernels of the integral operators Q_j so that $\mathcal{V}$ is obtained by "integrating" on the times τ_v $v > v_0$; clearly the integrations must be performed in the correct order, first the end-nodes .... The following proposition is standard (it is proved in [3] for numbered trees instead of equivalence classes), we sketch the proof in the Appendix.

Proposition 4.16. *The value of the splitting vectors $G_j^{ik}(\varphi)$ is*

$$G_j^{ik}(\varphi) = \Im \mathcal{V}_\varphi(\Lambda_j^{ik}) \,.$$

The value of the homoclinic trajectory ψ_j^k is

$$(\mu)^k \psi_j^k(t, \varphi) = (\Im_+^t + \Im_-^t) w_j(t, \tau_0) \mathcal{V}_\varphi(\Lambda_j^{ik}) + \sum_{a=0,1} x_j^{[a]} G_j^{ak} \,.$$

Definition 4.17 (Equivalent trees). *We are mainly interested in the splitting vectors and splitting matrix so we will consider two trees to be equal if they have the same value in the computation of the G_j^a.*

$$A \cong B \text{ iff } \Im \mathcal{V}_\varphi(A) = \Im \mathcal{V}_\varphi(B) \; \forall \, \varphi \in \mathbb{T}^n \,;$$

such identity can hold only for some initial data $\bar{\varphi}$, in such case we write

$$(A \cong B)_{\bar{\varphi}} \,.$$

4.4. *Tree identities*

4.4.1. *Mark adding functions*

We can define linear functions[u] on $\mathbb{V}(\mathcal{T})$, for instance we can add markings to a tree; given $A \in \overset{0}{\mathcal{T}}$ the symbol

$$h(v,t)\partial_l^v A$$

represents the application of an angle-marking $J(v) = l$ and a function-marking $h(v,t)$ in the node v; formally

$$A, \{v_A\}, \{v_i\}_{i=1}^m, \{w_j\}_{j=1}^h \to A, \{v_A\}, \{v_i\}_{i=1}^m, \{\{w_j\}_{j=1}^h \cup \{v\}\}\,,$$

notice that given two nodes v, w in the same orbit $[v]$ $\partial_l^v A$ is isomorphic to $\partial_l^w A$. We can define the linear function:

$$M_j(h(t))[A] := \sum_{v \in \mathring{A}} h(v,t)\partial_j^v A\,. \tag{4.1}$$

Particularly interesting mark adding functions are $M_j^b \equiv M_j(x_j^b(t))$.

Lemma 4.18. *The vector* $\mho_{ij}^{ab}$ *is obtained from* $\mho_i^a$ *by the mark adding function*

$$M_j^b[\mho_i^a] = \mho_{ij}^{ab}\,.$$

Proof. We need to show that

$$\sum_{A\in\mathcal{A}_i^a} \sum_{v\in\overset{0}{A}} \frac{1}{|\mathcal{S}(A)|}\partial_j^v A = \sum_{A\in\mathcal{A}_i^a} \sum_{[v]\in\overset{0}{A}} \frac{m[v]}{|\mathcal{S}(A)|}\partial_j^{[v]} A = \sum_{B\in\mathcal{A}_{ij}^{ab}} \frac{B}{|\mathcal{S}(B)|}\,,$$

in the second equality $m[v]$ is the cardinality of the orbit of v and the sum over $[v]$ means we choose one representative from each equivalence class; similarly the symbol $\partial_j^{[v]}$ is the application of the angle marking j to one of the nodes of the orbit $[v]$. We are simply grouping the isomorphic trees $\partial_j^w A$ with $w \in [v]$ and choosing a representative of the equivalence class. Given each tree $B \in \mathcal{A}_{ij}^{ab}$ there is one and only one couple $A \in \mathcal{A}_i^a$, $[v] \in A$ such that $\partial_j^{[v]} A = B$ (there is a common representative). The symmetry group of B fixes both the marked nodes so $|\mathcal{S}(A)| = m[v]|\mathcal{S}(B)|$ by the Lagrange theorem.[v] □

Lemma 4.19. *The function* M_j^0 *with* $j = 1, \ldots, n$ *is a function on the values of trees. Given a fruitless tree* $A \in \mathcal{A}$ *the mark-adding function* M_j^0 *with* $j = 1, \ldots, n$ *acts as the derivative on the angle* φ_j:

$$\partial_{\varphi_j} \Im\mathcal{V}_\varphi(A) = \Im\mathcal{V}_\varphi(M_j^0[A])\,.$$

[u]We always define functions F on trees. Then one should verify that $F(A)$ and $F(B)$ are isomorphic if A, B are so. This implies that one can uniquely extend the functions on the vector spaces by linearity.

[v]We refer to the Lagrange theorem which states that the order of a group G acting on a set V is the order of the orbit of a point $v \in V$ times the order of the subgroup of G which fixes v.

Proof. Adding an angle marking j to the node v is equivalent to adding e_j to m_v in $\nabla^m f^\delta$, so we add a derivative in ψ_j to the function $f^{\delta_v}(\psi)$ which is to be evaluated in $\psi_j = \varphi_j + \omega_j \tau_v$, $\psi_0 = q_0(t)$. If $j \neq 0$ this is equivalent to applying a φ_j derivative to the node v. As the dependence on φ comes only from the functions f^1 we have proved our assertion. □

4.4.2. *Fruit adding functions*

Remark 4.20. Notice that by our definition of equivalent trees adding a fruit of order k, type i and angle j in the free node v of a tree $A \in \overset{0}{\mathcal{T}}$ is equivalent to adding a mark $x_j^{[i]}(t)\partial_j^v$ to the node v and multiplying by the φ dependent function G_j^{ik}.

As we have seen in Eq. (3.4) the only contributions to Δ_{ij}^{0k} come from the parts of G_i^{0k} which are at most linear in the G_l^{mh} with $l = 0, \ldots, n$ $h < k$ $m = 0, 1$. In tree representation we can say that the only contribution comes from trees with one fruit. So to find the matrices N^a and n^a $(a = 0, 1)$ we have to understand how to pass from fruitless trees to trees with one fruit. First of all let us notice that the fruitless contribution to Λ_j^0 is clearly $\mho_j^0$ so that

$$D_{ij}^0 = \Im \mathcal{V}_{\varphi=0} \mho_{ij}^{00} \,.$$

Now we can add a fruit $\mathcal{F}_j^{ik}$ to the node v of a tree $A \in \overset{0}{\mathcal{T}}$ by adding a node y labeled (i, k, j) to the list $\mathcal{F}(A)$ and setting $y \in s(v)$, given a tree a we apply this function to each node $v \in A$ then sum on the nodes v. By Remark 4.20 this is equivalent to applying the function $G_j^{ik}(\varphi) M_j^{[i]}$, where $[i] = |i - 1|$, to A.

Proof of Proposition 3.2(i). If $j \neq 0$ we can obtain each tree with one fruit by adding the fruit to a node of a fruitless tree as described above; so that by Lemma 4.18,

$$N_{ij}^{ak} = \Im \mathcal{V}_{\varphi=0}(\mho_{ij}^{0a})$$

and consequently $N^0 = D^0$. □

If $j = 0$ we have trees with one fruit attached to nodes with $\delta_v = 0$, so that detaching the fruit we do not obtain an acceptable tree (the node has only one successive free node). We construct such trees from fruitless ones by using a different function: given a tree $A \in \mathcal{A}$ and a node $v \in A$ $v \neq v_A$ and $j_v = 0$ we attach the node y of the tree in Fig. 6 to v and w (by convention the node preceding v). Formally we set

$$l^i(A, v) = \mathcal{E}(A) \setminus (w, v) \cup (w, y) \cup (y, v) \,;$$

then $G_0^{[i]h} l^i(A, v)$ is a tree with one fruit, stemming from y $(\delta_y = 0)$ and y has only one successive free node.

We apply $l^i(A, v)$ to the nodes of A, and set $l^i(A, v) = 0$ if $v = v_A$ or if $j_v \neq 0$.

$$L^i(A) = \sum_{v \in A} l^i(A, v) \,;$$

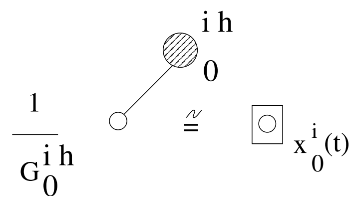

Fig. 6.

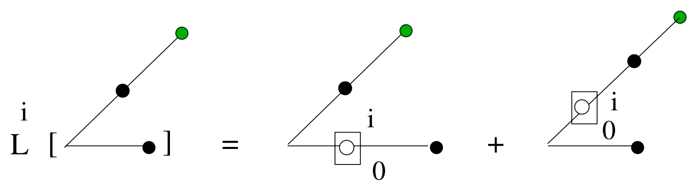

Fig. 7. The fruit adding functions.

notice that this is NOT well defined as a function $\mathcal{A} \to \mathcal{A}$. However $G_0^{ih} l^{[i]}(A)$ is well defined and $\mathcal{A} \to \overset{0}{\mathcal{T}}$ and so we can define the "value" of $L^i(A)$; in the next Subsection we will prove that $L^i(A)$ is equivalent to an acceptable tree.

Lemma 4.21. *Calling* $\overset{0}{\mathcal{T}}{}^{1F}$ *the set of trees with one fruit and*

$$\Lambda_j^{0(1F)} = \sum_{A \in \overset{0}{\mathcal{T}}{}_j^{1F}} \frac{A}{|\mathcal{S}(A)|}\,,$$

we have that:

$$\Lambda_j^{0(1F)} = \sum_{l=0,1} \sum_{i=0}^{n} G_i^l (M_i^{[l]}(\mho_j^0) + \delta_{i0} L^{[l]}(\mho_j^0))\,,$$

and consequently

$$n_j^0 = \Im \mathcal{V}_{\varphi=0}(\mho_{j0}^{00} + L^0(\mho_j^0))\,.$$

Proof. Consider a tree B with one fruit, of angle i, order k and type l attached to a node v'. If such node has more than one successive or $\delta_v \neq 0$, then it can be obtained by applying $G_i^{[l]h} x_i^l \partial_i^v$ to a tree $A \in \mathcal{A}_j^0$. If the node has $\delta_v = 0$ and only one successive then there exists one and only one couple A, v with $A \in \mathcal{A}_j^0$, v node of A such that $G_0^{[l]h} l^l(A, [v]) = B$ (as usual the symbol $[v]$ means choosing one representative for the equivalence class). The symmetry group of B fixes both the first node and the fruit (and so consequently all the path joining the fruit to the first node), so if we divide by G_i^{lk} we obtain a tree with two marked nodes which again fixes the first node and the node v' where the fruit was attached; if v'

has only one successive free node, say v, then that is fixed as well. This proves the proposition as given $A \in \mathcal{A}_j^0$,

$$L^a(A) = \sum_{[v]\in A} m[v] l(A,[v]) \quad \text{and moreover} \quad |\mathcal{S}(A)| = m[v]|\mathcal{S}(l(A,[v]))|\,.$$

Finally n_j^0 is the linear term in G_0^1 in the expansion of G_j^0, so it is given by trees with one fruit of angle $j = 0$ and type $l = 0$. □

Remark 4.22. As $f^\delta(t) = F(\psi_i(0) + \tilde{\omega}_i t, \psi_0(t))$ and $\dot{\psi}_0(t) = -\frac{2}{c} x_0^0(t)$, we have that:

$$\partial_{\tau_v} \nabla^{\vec{m}} f^\delta(\tau_v) = \sum_{j=1,\ldots,n} \omega_j \nabla^{\vec{m}+e_i} f^\delta(\tau_v) - \frac{2}{c} x_0^0 \nabla^{\vec{m}+e_0} f^\delta(\tau_v)\,.$$

For notational convenience we define a symbol[w] $\partial_t^y A$, where A is a fruitless tree and v is one of its nodes, by setting

$$\Psi_\varphi(\partial_t^y A) = \prod_{\substack{v\in A\\ v>v_0}} w_{j_v}(\tau_w, \tau_v) \prod_{v\in A} \left(-\frac{1}{2} a_{j_v}\right) \mu^{\delta_v} \prod_{\beta\in\mathcal{M}(v)} h_\beta(v,\tau_v)$$
$$\times \prod_{\substack{v\in A\\ v\neq y}} \nabla^{\sum_{j=0}^n m_v(j) e_j} f^{\delta_v} \partial_{\tau_y}(\nabla^{\sum_{j=0}^n m_y(j) e_j} f^{\delta_y})\,.$$

This definition implies that[x]

$$\sum_{v\in A} \partial_t^v A \cong \sum_{j=1,\ldots,n} \omega_j M_j^0(A) - \frac{2}{c} M_0^0(A)\,.$$

Lemma 4.23. *Given an odd function $G \in H_0$ the following relation holds*:

$$\partial_t Q_j^t(G) = Q_j^t\left[\partial_\tau G(\tau) + \frac{2}{c}\delta_{j0} x_0^0(\tau) \partial_0^3 f^0(\tau) Q_0^\tau(G)\right]\,.$$

The proof of this Lemma (proposed in [4]) is straightforward but quite long, we report it in the Appendix.

Lemma 4.24. *Given a tree $A \in \mathcal{A}_i^0$, $i = 1,\ldots,n$ then*

$$\mathcal{V}_{\varphi=0}\left(\sum_{v\in A} \partial_t^v A - l^0(A,v)\right) = \partial_t \mathcal{V}_{\varphi=0}(A)\,.$$

Proof. We drop the $\varphi = 0$ in $\mathcal{V}$ for notational convenience. The assertion is trivially true for trees with only one node, so we prove it by induction on the order of the trees. Let us define $\mathcal{A}_j^h$ as the set of fruitless trees of order h with only one marking,

[w] We could define $\partial_t^v(A)$ to be a special marked tree.
[x] Remember that $A \cong B$ means that $\Im\mathcal{V}(A) = \Im\mathcal{V}(B)$.

placed on the first node, $J_{v_0} = j$ and $h(v_0, t) = 1$; for $j \neq 0$, $\mathcal{A}_j^h \cong \mathcal{A}_j^{0h}$. Suppose Lemma 4.24 holds for all trees in $\mathcal{A}_j^h$, $h < k$ for $j = 0, \ldots, n$, then for[y] $A \in \mathcal{A}_i^{0k}$,

$$\partial_{\tau_0}\mathcal{V}(A) = -\frac{1}{2}(\partial_{\tau_0}\nabla^{\vec{m}(v_0)} f^{\delta_{v_0}}) \prod_{v\in S(v_0)} Q_{j_v}[\mathcal{V}(A^{\geq v})]$$

$$+ \sum_{v\in S(v_0)} \mathcal{V}(A^{/v})\partial_{\tau_0}[Q_{j_v}\mathcal{V}(A^{\geq v})]) \,.$$

Now we set $\mathcal{V}(A^{\geq v}) = F$ (which is odd when $\varphi = 0$) and apply Lemma 4.23 to $F \in H_0$:

$$\partial_{\tau_0} Q_{j_v}(F) = Q_{j_v}(\partial_{\tau_v} F) + \frac{2}{c}\delta_{j0} Q_0(x_0^0(\tau_y)\partial_0^3 f^0(\tau_y) Q_0(F)) \,,$$

clearly

$$\partial_{\tau_v} F = \partial_{\tau_v}[\mathcal{V}(A^{\geq v})] \quad \text{and } \delta_{j0} Q_0(x_0^0(\tau_y)\partial_0^3 f^0(\tau_y) Q_0(F)) = -\mathcal{V}(l^0(A, v)) \,.$$

So we obtain

$$\partial_{\tau_0}\mathcal{V}(A) = \mathcal{V}(\partial_t^{v_0} A) - \sum_{v\in S(v_0)} \frac{2}{c}\mathcal{V}(l^0(A, v)) + \sum_{v\in S(v_0)} \mathcal{V}(A^{/v})[Q_{j_v}\partial_{\tau_v}\mathcal{V}(A^{\geq v})])$$

by definition $A^{\geq v} \in \mathcal{A}_j^h$ for some j, h. So we consider trees of lower order for which the Proposition is true by the inductive hypothesis. $\square$

Proof of Proposition 3.2(ii). By Lemma 4.21 we must show that

$$n_i^0 = \Im\mathcal{V}_{\varphi=0}(\mho_{i0}^{00} + L^0(\mho_i^0)) = \frac{c}{2}\Im\mathcal{V}_{\varphi=0}\left(\sum_{j=1}^{n} \mho_{ij}^{00}\omega_j\right) . \tag{4.2}$$

Now for $j \neq 0$, $\Im\partial_t\mathcal{V}_{\varphi=0}(\mho_j^0) = 0$ as the integrand has no constant component. So we can use Lemma 4.24 and Remark 4.22 to obtain Eq. 4.2. $\square$

4.4.3. *Changing the first node*

Another way of manipulating trees is to change the first node (which is distinguishable as it does not have the label j). Generally one can obtain various trees in $\overset{0}{\mathcal{T}}$ by simply changing the uncolored node (for example one can shift the angle labels down along a path joining any node v to the uncolored one v_A). However not all the trees obtained in such a way are in $\mathcal{T}$.

Definition 4.25. *Given a tree* $A \in \overset{0}{\mathcal{T}}$, *let* v_A *be the first node and* v *a free node; the change of first node* $P(A, v) : \overset{0}{\mathcal{T}} \to \overset{0}{\mathcal{T}}$ *is so defined*:

[y] We recall that $\mathcal{V}(A) = \mathcal{V}(A^{/v})Q_{j_v}\mathcal{V}(A^{\geq v})$.

Let $v_A = v_0, v_1, \ldots, v_m = v$ *be the nodes of the path* $\mathcal{P}_{v_A,v}$. $P(A,v)$ *is obtained from* A, $\{v_A\}$, $\{v_i\}_{i=1}^m$, $\{w_j\}_{j=1}^h$ *by shifting only the* j *labels of the nodes of* $\mathcal{P}_{v_A,v}$ *in the direction of* v_A. *This automatically implies that* v *is left* j*-uncolored and is the first node of* $P(A,v)$. *If we obtain a tree not in* $\mathcal{T}$ *we set* $P(A,v) = 0$.

$P : \mathbb{V}(\mathcal{T}) \to \mathbb{V}(\mathcal{T})$ *is the linear function such that* $\forall\, A \in \mathcal{T}$, $P(A) = \sum_{v\in \overset{0}{A}} P(A,v)$.

Lemma 4.26. $P(A,v) = 0$ *if and only if* $\delta_{v_A} = 0$, $|s(v_A)| = 2$. *This means that the possibility of applying the change of first node does not depend on the chosen* $v \neq v_A$.

Proof. Consider the trees A and $P(A,v)$ and the nodes $v_A = v_0, v_1, \ldots, v_m = v$ of the path $\mathcal{P}_{v_A,v}$. For each $i = 0, m-1$, v_i precedes v_{i+1} in A and follows it in $P(A,v)$. So for each node $w \neq v_A, v$, the number of following nodes $s(w)$ is the same in A and $P(A,v)$; $s(v_A)$ decreases by one and $s(v)$ consequently increases by one. This implies that all trees A with $\delta_{v_A} = 0$ and $|s(v_A)| = 2$ have $P(A,v) = 0$ for all v. Moreover if v_i has $\delta = 0$, then it has $j = 0$ as well as all the nodes (including v_{i+1}) following it. This means that in $P(A,v)$, it will still have $\delta = j = 0$, the same $s(v_i) \geq 2$; moreover v_{i-1} that follows v_i in $P(A,v)$ has $j = 0$. □

We will call $\overset{r}{\mathcal{T}}$ the trees whose first node can be changed.

Lemma 4.27. *By Proposition* 2.6(*a*), *we have*:

$$\begin{gathered}\forall\, A \in \overset{r}{\mathcal{T}},\ \forall v \in A : P(A,v) - A \in \ker \Im\mathcal{V}_\varphi \\ \forall\, A \in \overset{r}{\mathcal{T}}_{(j,f)(i,h)}\colon P_1(A) - A \in \ker \Im\mathcal{V}_\varphi\end{gathered} \tag{4.3}$$

Proof. Notice that given a tree A and one of its nodes v, if $w \in \mathcal{P}(v_A, v)$ then $P(A,v) = P(P(A,w),v)$, so we only need to prove the assertion for $v \in s(v_A)$. Given $A \in \overset{r}{\mathcal{T}}$ and $v \in s(v_A)$ such that $j_v = j$, we compare $\Im\mathcal{V}(A)$ and $\Im\mathcal{V}(B)$ with $B = P(A,v)$, so B has first node v (no label j_v) and a node v_A in $s(v)$ with $j_{v_A} = j$.

$$\begin{aligned}\Im\mathcal{V}(A) = &-\frac{1}{2} a_{j_{v_a}}(\mu)^{\delta_{v_A}} \Im\nabla^{\sum_j m_{v_A}(j)e_j} f^{\delta_{v_A}} \prod_{\substack{w\in s(v_A)\\ w\neq v}} Q_{j_w}[\mathcal{V}(A^{\geq w})] \\ &\times Q_j\left[(-\mu)^{\delta_v}\nabla^{\sum_j m_v(j)e_j} f^{\delta_v} \prod_{w_1\in s(v)} \mathcal{V}(A^{\geq w_1})\right],\end{aligned}$$

which by the symmetry of Q_j is equal to

$$\begin{aligned}&\Im\nabla^{\sum_j m_v(j)e_j}(-\mu)^{\delta_v} f^{\delta_v} \prod_{w_1\in s(v)} \mathcal{V}(A^{\geq w_1}) Q_j \\ &\times\left[(-\mu)^{\delta_{v_A}}\nabla^{\sum_j m_{v_A}(j)e_j} f^{\delta_{v_A}} \prod_{\substack{w\in s(v_A)\\ w\neq v}} Q_{j_w}[\mathcal{V}(A^{\geq w})]\right].\end{aligned}$$

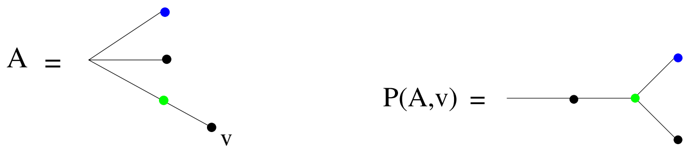

Fig. 8. An example of trees that are equivalent by changing the first node.

This is the value of B, namely, both in A and in B, $m_v(i)$ with $i \neq j$ is the number of elements in $(s(v), \mathcal{M}(v), \mathcal{F}(v))$ having label i and $m_v(j) - 1$ is the number of elements in $(s(v), \mathcal{M}(v), \mathcal{F}(v))$ having label j. □

Lemma 4.28. *For each $i = 1, n$, we have*

$$\mho_i^0 \cong M_i^0(\mho)\,.$$

Proof. The proof of this statement is in [7], we report it here for completeness. By Lemma 4.27 we have that for A in $\mathcal{A}_j^k$,

$$A \cong \frac{1}{k} \sum_{[v]:\delta_v = 1} m[v] P(A, v) \quad \text{so} \quad \sum_{A \in \mathcal{A}_j^k} \frac{A}{|\mathcal{S}(A)|} = \frac{1}{k} \sum_{A \in \mathcal{A}_j^k} \sum_{[v]:\delta_v = 1} \frac{m[v]}{|\mathcal{S}(A)|} P(A, v)\,,$$

now there exists one and only one couple $B \in \overset{m}{\mathcal{A}}$, $[v_A] \in B$ such that $\delta_{v_B} = 1$ and $\partial_j^{v_A} B = P(A, v_B)$. Finally by the Lagrange Theorem, $(m[v_B])^{-1} |\mathcal{S}(A)| = (m[v_A])^{-1} |\mathcal{S}(B)|$. □

This completes the proof of Proposition 3.2.

4.5. *Upper bounds on the values of trees*

Given a fruitless tree $A \in \overset{m}{\mathcal{A}}$ of order k (so with at most $2k - 1$ nodes), its value through $\Im \mathcal{V}_\varphi^1$ is of the form:

$$\left(-\frac{1}{2}\right)^{N(A)} \left(\prod_{v \geq v_0} a_{j_v}\right) \Im \prod_{v > v_0} (\Im_+^{\tau_w} + \Im_-^{\tau_w})(\mu)^{\delta_{v_0}} \nabla^{\sum_j m_{v_0}(j) e_j} f^{\delta_{v_0}}$$
$$\times \prod_{v > v_0} (\mu)^{\delta_v} \nabla^{\sum_j m_v(j) e_j} f^{\delta_v} w(\tau_w, \tau_v)\,. \qquad \text{(a)}$$

We expand f^1 in Fourier series in the rotator angles,

$$f^1(\psi, q) = \sum_{|\nu| = 1} e^{i\nu \cdot \psi} f_\nu(q)\,,$$

so that each node has one more label $\nu_v \in \mathbb{Z}^n$. We will represent as $A(\nu)$ a tree A with labels ν_v such that

$$\sum_{v \in A} \nu_v = \nu\,.$$

In each node v with $\delta = 1$ we have as factor the function $d^{n_v} f_{\nu_v}(q(t))$ where $n_v = m_v(0)$.

The functions $f_\nu(q)$ and $q(t)$ are such that $f_\nu(q(t)) = F_\nu(e^t) \in H_0(a,d)$. Naturally by our analyticity assumptions $f_\nu(q(t))$ is limited for $|t| \to \infty$ in $|\mathrm{Im}\ t| < 2\Pi$. We are considering rational functions $F_\nu(e^t)$, let us call t^i_ν their (finite number of) poles in $|\mathrm{Im}\ t| \le \Pi$ (all with $\mathrm{Im}\ t \ne 0$) then

$$d = \min_{\nu,i} |\mathrm{Im}\ (t^i_\nu)|\,; \quad a = \max_{\nu,i} |\mathrm{Re}(t^i_\nu)|\,. \tag{4.4}$$

Moreover the following proposition holds.

Lemma 4.29. *The functions $\partial^k_0 f_\nu(q(t)) = F^k_\nu(e^t)$ are all limited rational functions of e^t, whose poles are the same as those of $F^{\,0}_\nu(e^t)$; moreover*

$$\max_{t\in C(a+2,d-\sqrt{\varepsilon})} |F^k_\nu(e^t)| \le Ck!\varepsilon^{\frac{p+k}{2}}\,. \tag{4.5}$$

Proof. We can use Cauchy estimates on $\partial^k_0 f_\nu(q)$ provided that the images in the q variables of $C(a+2, d-\sqrt{\varepsilon})$ and of $C(a+1, d-\frac{1}{2}\sqrt{\varepsilon})$ via the function $q_0(t)^{-1}$, have distance of the order of $c\sqrt{\varepsilon}$ for some order one c. This can be verified by direct computation or proved using simple geometric arguments. $\square$

Having fixed $\nu = \sum_v \nu_v$, in integral (a) we shift the integration to $\mathbb{R} + i\sigma(\omega_\nu)d'$ where $d' < d$ (we will then fix $d' = d - \sqrt{\varepsilon}$ to obtain optimal estimates and $d' = c \le d/2$ to obtain simply exponentially small estimates), $\omega_\nu = \omega\cdot\nu$ and $\sigma(x)$ is the sign of x. As the functions are all analytic in $|\mathrm{Im}(t)| \le d'$ the integral (a) is unchanged. Notice that in integral (a) we cannot choose the sign of the shift in the single node integrals and so we need to work in the (symmetric) domains $C(a, d')$ to guarantee the indifference of extending in the lower or upper half-plane. To simplify the notation we set

$$\sigma(\omega_\nu) = + \quad \text{and define} \quad E(d',\nu) = e^{-|\omega_\nu|d'}\,.$$

If A has k nodes with $\delta = 1$, let $\{\nu_v\}^k_\nu$ be the lists of k vectors $\nu_v \in \mathbb{Z}^n$ such that $\sum \nu_v = \nu$. The value of $A(\nu)$ (tree $A \in \overset{m}{\mathcal{A}}$ with total frequency ν) in integral (a) is:

$$\begin{aligned}
&\left(-\frac{1}{2}\right)^{N(A)} e^{i\nu\cdot\varphi} E(d',\nu) \sum_{\{\nu_v\}^k_\nu} \left[\prod_{\substack{s=1,\dots,n\\ \delta_v=1,\,v\ge v_0}} (i\nu_{vs})^{m_v(s)}\right] \oint \frac{dR_{v_0}}{2i\pi R_{v_0}} \int_{-\infty}^{\infty} d\tau_{v_0} e^{-\sigma(\tau_{v_0})R_{v_0}} \\
&\quad\times [d^{n_{v_0}} f^{\delta_v}_{\nu_{v_0}}(q(\tau_{v_0} + id'))] e^{i\omega_v\tau_{v_0}} \prod_{v>v_0} \oint \frac{dR_v}{2i\pi R_v}\left(\int_{-\infty}^{\tau_w} d\tau_v + \int_{\infty}^{\tau_w} d\tau_v\right) \\
&\quad\times e^{-\sigma(\tau_v)R_v(\tau_v+id')} w_{j_v}(\tau_w + id', \tau_v + id') \prod_{v\ge v_0} [d^{n_v} f^{\delta}_{\nu_v}(q(\tau_v + id'))] e^{i\omega_v\tau_v}\,;
\end{aligned} \tag{a}$$

naturally $f^0_\nu = 0$ for all non-zero ν. As usual w is the node preceding v, $m_v(s)$ is the number of nodes in the list v, $s(v)$ with label $j = s$, $n(v)$ the number of those with label $j = 0$ and $\omega_v = \omega_{\nu_v}$.

The residues in R are introduced by using the Definition 2.8. The factors $(i\nu_{vs})^{m^s_v}$ come only from nodes with $\delta_v = 1$ and their product is bounded by 1. Now we want estimates on the integrals that depend only on the order k; we start by splitting the sums in monomials.

(1) Split $w_j(\tau_w + id, \tau_v + id)$ into 6 terms if $j = 0$ or 2 terms if $j \neq 0$: so we obtain 6^{3k-1} terms. Each of this terms is of the form

$$\tau_v^h x_v^{-l} y(x_v) \tau_w^{h'} x_w^{-l'} y'(x_w)\,,$$

where $x_v = e^{-|\tau_v|}$, $0 \le h$, h', l', $l \le 1$ and both $y(x)$, $y'(x)$ are analytic in $|x| \le 1$ (we will call this the limited x dependent part of the Wronskian).

(2) Separate $\int_{-\infty}^{\tau_w} d\tau_v + \int_{\infty}^{\tau_w} d\tau_v$, and $\Im d\tau_{v_0}$ in integral (a). We get other 2^k terms like

$$\prod_{v \ge v_0} \oint \frac{dR_v}{2i\pi R_v} \left(\int_{\rho_v \infty}^{\tau_w} d\tau_v e^{-\sigma(\tau_v) R_v(\tau_v + id)} e^{i\omega_v \tau_v} (\tau_v)^{h_v} x^{l_v} \prod_{j=1}^{|s(v)|+2} y_v^j(x_v) \right),$$

where $0 \le l_v$, $h_v \le |s(v)| + 1$. Notice that ρ_v is not the sign of τ_v but an extra label. The functions y_v^j are chosen in the following way:

(i) One of the y_v^j is either coming from $\partial_0^{n_v} f^0$, (i.e. it is in the list $\cos(mq(\tau_v + id))$, $\sin(mq(\tau_v + id))$ with $m = 1, 2$) or is one of the $F^k_{\nu_v}$.
(ii) One is the limited x_v dependent part of a term from the Wronskian at the node v.
(iii) For each node v' following v there is one function y_v^j which is the x_v dependent part of a term coming from the Wronskian $w(\tau_v, \tau_{v'})$.

Notice that the functions y are by definition in $H(a, d)$ and respect condition 4.29.

(3) Given a node $v \in s(v_0)$ split the integral $\int_{\rho_v \infty}^{\tau_{v_0}} d\tau_v$ as $\int_{\rho_v \infty}^{0} d\tau_v - \int_{\rho_{v_0} \infty}^{0} d\tau_v + \int_{\rho_{v_0} \infty}^{\tau_{v_0}} d\tau_v$ and proceed recursively for all nodes (other 3^{2k+1} terms). We consider first the contributions from the term with $\int_{\rho_{v_0} \infty}^{\tau_w} d\tau_v$ for all nodes (the others will be expressed as products of the same kind of integrals).

Set $\rho_{v_0} = -1$, we want to estimate:

$$I_-(A) = \prod_{v \ge v_0} \oint \frac{dR_v}{2i\pi R_v} \left(\int_{-\infty}^{\tau_w} d\tau_v e^{R_v(\tau_v + id)} e^{i\omega_v \tau_v} (\tau_v)^{h_v} x_v^{-l_v} \prod_{j=1}^{|s(v)|+2} y_j^v(\tau_v) \right). \tag{4.6}$$

Finally we split the first integral $\int_{-\infty}^{0} = \int_{-\infty}^{-a_0} + \int_{-a_0}^{0}$, where $a_0 > 0$ is suitably large ($a_0 = a + 2\log 2$). We set $y_j^v(\tau_v) = \sum_{r=0} y_j^{v,r} x^r$ and $C_{\{r_v\}} = \prod_v y_j^{vr_v}$. The integral is

$$I_m^{a_0} = \mathrm{Res} \sum_{\{r_v\}} C_{\{r_v\}} \prod_v \frac{\partial^{n_v}}{\partial E_v^{h_v}} \prod_v \int_{-\infty}^{\tau_w} d\tau_v (e^{R_v(\tau_v+id)+E_v\tau_v} e^{i\omega_v\tau_v} x_v^{r_v}) \tag{4.7}$$

with $\tau_{w_0} = -a_0$. Starting from the end-nodes we now perform the integrals in $d\tau_v$ then the derivatives in E_v and finally the residues in R_v, we do this first for all the end-nodes and then proceed to the inner nodes hierarchically.

Lemma 4.30. *Integral* (4.7) *produces the bounds*

$$I_m^{a_0} \le \varepsilon^{-m}(m!)^{2\tau+2} C_1^k \prod_v \left[\prod_{j=1}^{|s(v)|+2} \left(\sum_h |y_j^{v,h}||x_0^h| \right) \right];$$

$x_0 = e^{-a_0}$, *m is the number of nodes* ($\le 2k-1$), $|s(v)|$ *the number of nodes following v and C_1 is some order one constant. Finally τ is the Diophantine exponent of* $\frac{\omega}{\sqrt{\varepsilon}}$,

$$|\omega \cdot n| > \varepsilon^{\frac{1}{2}} \gamma |n|^{-\tau} \quad \textit{for some } \gamma = O_\varepsilon(1)\,.$$

If we choose $a_0 > a$ the series are all convergent (*by the analyticity of the y_j's in x_0*).

We choose $x_0 = \frac{e^{-a}}{8}$ and estimate the coefficients of the Taylor series in the ball $|x| \le e^{-a-2}$:

$$\sum_{k=0}^{\infty} |y_j^{v,k}| x_0^k \le 4 \max_{|x|\le 2x_0} (y_j^v(x))\,.$$

Proof of Lemma 4.30. This is taken from [3].

$$\text{The integral } \int_{-\infty}^{t} x^K e^{iA\tau} e^{B\tau} = \frac{x^K e^{(iA+B)t}}{K+B+iA}\,,$$

so the E_v derivatives in the end-node v give 2^{h_v} terms of the form:

$$h_1^v! \frac{x_w^{r_v} e^{idR_v} e^{(i\omega_v+R_v)\tau_w}}{r_v+R_v+i\omega_v} (\tau_w)^{h_2^v} \quad h_1^v + h_2^v = h_v\,. \tag{4.8}$$

The residue of R_v^{-1} times (4.8) is (4.8) if $|r_v| + |\omega_v| \neq 0$ and

$$\frac{h_2^v!}{(h_2^v+1)!} (\tau_w)^{h_1^v} (\tau_w + id)^{h_2^v+1} \quad \text{if } |r_v| + |\omega_v| = 0\,.$$

Developing the binomial we obtain other 2^{h_v+1} terms, all of the type

$$G^{h_v+1} \bar{m}! x_w^{r_v} e^{i\omega_v\tau_w} (\tau_w)^{\tilde{h}_v}\,.$$

The constant G is the maximum between one ($r_v \neq 0$), $(\min_{|\nu|\le N} |\omega\cdot\nu|)^{-1}$ or $(\frac{\Pi}{2})$ (we use that $d < \frac{\Pi}{2}$). After integrating all the end-nodes following a node w we can integrate in $d\tau_w$ a sum of $2^{2\sum_{v\in s(w)} h_v+1}$ terms of the type

$$G^{\bar{h}} \bar{h}! x_w^{\tilde{r}_w} e^{i\Omega_v\tau_w} (\tau_w)^{\hat{h}}$$

where $\tilde{r}_v = \sum_{v\in s(w)} r_v$, $\Omega_v = \sum_{v\in s(w)} \omega_v$ and $\bar{h} + \hat{h} \leq \sum_{v\in s(w)} h_v + 1$. We have proved that the integrals derivatives and residues correspond to calculating the integrands in (4.7) at the limiting point a_0, ignoring the oscillating factors $e^{i\Omega a_0}$, substituting the Taylor coefficients with their moduli and multiplying by a factor bounded by

$$2^{6k-3}(k!)^4 \max_{0<|\nu|<kN} (|\omega\cdot\nu|)^{-2\tau(2k-1)} \leq C^k(k!)^{4\tau+4}\,. \qquad \square$$

We now consider the "left out part" $\int_{-a_0}^{t} d\tau_{v_0}$ (we will set $t=0$ in integral (a)). Let v_1 be a node of level one.

We break the integral $\Im^{\tau_{v_0}} d\tau_{v_1}$ as $\Im^{-a_0} d\tau_{v_1} + \int_{-a_0}^{\tau_{v_0}} d\tau_{v_1}$. If we choose the first term and m_1 is the number of nodes of $A^{\geq v_1}$, the integral on $A^{\geq v_1}$ can be bounded by $I^{a_0}_{m_1}$ and we are left with the problem of bounding the "left out part" $\int_{-a_0}^{t} d\tau_{v_0}$ on the remaining subtree $A^{/v_1}$. We repeat the procedure hierarchically and we end up with 2^m terms of the form:

$$I^{a_0}_{m_1}\cdots I^{a_0}_{m_p}\prod_{v\in\vartheta}\int_{-a_0}^{\tau_w} d\tau_v \mathcal{V}(\vartheta)$$

where the subtree ϑ has $\tilde{m}$ nodes and $\tilde{m}+\sum m_j = m$. We bound the last integral by the maximum of the integrand. Let us now examine the 3^m-1 integrals left aside in the analysis of item 3. Starting from the end-nodes we cut off all the subtrees ϑ that contribute a definite integral $\Im^0_\rho$. Such integrals are of the type $I_\rho(\vartheta_i)$ that we have already considered. We are left with an integral again of the type $I_{\rho_0}(\vartheta_0)$ where ϑ_0 is the tree deprived of the ϑ_i. The total number of nodes of the ϑ_i $i=0,\ldots,h$ is m.

Now we only have to compute the maxima of the $|y^v_j(x)|$, that means the maxima of the moduli of the terms form the derivatives of f^0 from the Wronskian and from all the F^k_ν in the regions $C(a=2,d')$. To bound the functions $F^k_\nu(e^t)$ we use the fact that $d' = d-\sqrt{\varepsilon}$ and Proposition 4.29(ii) .

As we are not interested in optimality,[z] we will estimate the maximum of a pole of order k by $\varepsilon^{-\frac{k}{2}}$.

Lemma 4.31. *The $|y^v_j(\tau_v)|$ non coming from f^1 contribute at most a factor ε^{-k-2k_0+1} where k_0 is the number of nodes with $\delta_v = 0$.*

Proof. There are $k_0 \leq k-1$ nodes with $\delta_v = 0$ carrying at most a double pole. Then each of the $k+k_0-1$ nodes $v\neq v_0$ carries a summand of

$$\max_{t\in C(d-\sqrt{\varepsilon},a+2)}(|x^0_j|)\max_{t\in C(a+2,d-\sqrt{\varepsilon})}(|x^1_j|)$$

from the Wronskian. So it is another double pole. $\square$

[z]Notice that if the perturbing function is not a trigonometric polynomial then we do not approach simultaneously both the singularities of f^0 and f^1.

The functions F_ν^n appear exactly k times. Moreover $n = \sum_{i=1}^k n_{v_i}$, so we count each node with $\delta_v = 1$ plus all its successive nodes. As each node with $\delta_v = 0$ has $s(v) \geq 2$,

$$\sum_{i=1}^k n_{v_i} \leq \sum_v n_v - 3k_0 = 2k - k_0 - 1\,.$$

We can bound the maxima of the F_ν^n in $C(a+2, d-\sqrt{\varepsilon})$ using Lemma 4.29, so we have a factor $\sqrt{\varepsilon}^{-(p+2)k+k_0}$.

Finally we notice that $E(d-\sqrt{\varepsilon},\nu) \sim E(d,\nu)$ and we sum on all the trees of order k using the known bound:

$$\sum_{A \in (\overset{m}{\mathcal{A}})^k} \frac{\prod_{\substack{v\in A\\ \delta_v = 1}} n(v)!}{|\mathcal{S}(A)|} \leq (4n)^k\,.$$

This proves Proposition 3.3(ii). To prove Proposition 3.3(i) we set $d' = 0$ so we do not have any divergent contribution from the integrals in $(-a+2, a+2)$. Moreover we can add fruits and markings by simply using the mark adding functions; see [3] for full details.

Appendix A.

A.1. *Proof of Proposition* 4.16

We give values to trees recursively; namely given a tree with fruits $A \in \mathcal{T}_j$ we define its value as:

$$\mathcal{V}(A) = -\frac{1}{2}\mu^{\delta_{v_0}} a_j \nabla^{\vec{m}+e_j} f^{\delta_{v_0}}(t)) \prod_{v\in s(v_0)} \mathcal{W}(A^{\geq v})\,, \quad \text{where}$$

$$\mathcal{W}(A) = (\Im_+^t + \Im_-^t) w_j(t,\tau)\mathcal{V}(A)$$

if A is not a fruit, and

$$\mathcal{W}(\mathcal{F}_j^{ak}) = x_j^{[a]}(t)\Im x_j^a \mathcal{V}(\Lambda_j^k)$$

otherwise. Finally we set $\mathcal{V}(\Lambda_j^1) = -\frac{1}{2}\mu a_j \nabla^{e_j} f^1(t)$. This is clearly the same function $\mathcal{V}$ we defined in Subsec. 4.3. We consider a multi-linear function Γ_j^δ on trees $\mathcal{T}_j$ so define $\Gamma_j^\delta(A_1,\ldots,A_n)$ attaches the fist nodes of the trees A_i to the tree α_j^δ with one marked node $\delta_v = \delta$, $J_v = j$. By convention if we have n copies of the tree A in the list $\{A_i\}$ we will write $\Gamma_j^\delta(A^n)$. Remembering that

$$F_j^k = -\sum_{\delta=0,1} \sum_{\{p_j^h\}_{\vec{m},k-\delta}} \nabla^{\vec{m}+e_j} f^\delta(t) \prod_{\substack{j=0\\ h=1}}^{n,k-1} \frac{(\psi_j^h)^{p_j^h}}{p_j^h!}\,.$$

We define recursively:

$$\Lambda_j^k = \frac{1}{2} a_j \sum_{\delta=0,1} \sum_{\{p_j^h\}_{\vec{m},k-\delta}} \frac{\mu^\delta}{P\{p_j^h\}} \Gamma_j^\delta((\bar{\Lambda}_0^1)^{p_0^1}, \dots, (\bar{\Lambda}_n^{k-1})^{p_n^k-1}) , \quad \bar{\Lambda}_j^k = \Lambda_j^k + \sum_{a=0,1} \mathcal{F}_j^{ak} ,$$

where given a list $\{a_i\}$, $P\{a_i\} = \prod_i a_i!$. By definition, $\mathcal{W}(\bar{\Lambda}_j^k) = \psi_j^k$. Let us prove that $\Lambda_j^k = \sum_{A \in \mathcal{T}_j^k} |\mathcal{S}(A)|^{-1} A$. As in both definitions, Λ_j^k is a sum over all trees in $\mathcal{T}_j^k$. This is equivalent to showing that in the two expressions, each tree A has the same coefficient. We proceed by induction as the statement is trivially true for Λ_j^1. Given a tree $A \in \mathcal{A}_j^k$, let v_1 v_m be its level one nodes and $A^1, \dots A^m$ its level one subtrees; we need that

$$\frac{1}{|\mathcal{S}(A)|} = \frac{N(A^1, \dots, A^m)}{P\{p_i^h(A)\}} \prod_{i=1}^m \frac{1}{|\mathcal{S}(A^i)|}$$

where $\{p_i^h(A)\}$ is the number of trees $\{A^j\}$ in $\mathcal{A}_i^h$ and $N(A^1, \dots, A^m)$ is the number of ways in which one can choose one summand from each $\mho_0^1, \dots, \mho_n^{k-1}$ and obtain the unordered list $(A^1, \dots, A^m)$. Now if $m[v_i]$ is the cardinality of the orbit of v_i (so there are $m[v_1]$ subtrees equal to $A^1 \dots$),

$$N(A^1, \dots, A^m) = \frac{P\{p_i^h(A)\}}{\prod_{[v]_1} m[v]!} \quad \text{and} \quad \prod_{i=1}^m \frac{1}{|\mathcal{S}(A^i)|} = \prod_{[v] \in s(v_0)} \frac{1}{|\mathcal{S}(A^{\geq [v]})^{m[v]}|} .$$

This proves the assertion by the Lagrange Theorem:

$$|\mathcal{S}(A)| = \prod_{[v] \in s(v_0)} m[v]! |\mathcal{S}(A^{\geq v})|^{m[v]} .$$

A.2. *Normal form theorem*

We perform a symplectic change of variables that brings Hamiltonian (1.1) in local "normal form". We will use the standard notations (see [2], [4] or [12], [13]) and the existence of the fast time scale. For systems with one fast time scale, this provides a symplectic change of variables defined in a region W such that $\Pi_I W = O_\varepsilon(1)$, that sends the perturbing terms depending on the fast angle to order $e^{-\frac{1}{\varepsilon^B}}$ for some $B(n) < 1$. This will be the basis for proving Arnold diffusion for systems with one fast variable. For completeness we state the theorem for m fast variables. The first step is to set the pendulum in local hyperbolic normal form (see [2]), we obtain the local Hamiltonian:

$$\frac{1}{2}(I, AI) + \sqrt{\varepsilon} G(pq, \sqrt{\varepsilon}) + \mu f(p, q, \psi) , \tag{A.1}$$

where the function $G(J, \sqrt{\varepsilon})$ is analytic for $|J| < \tilde{k}_0^2 \sim \sqrt{\varepsilon}$ and will be written as Taylor series: $G(J) = \sum_{k \geq 1} J^k G_k$.

The perturbing term $f(p, q, \psi)$ is a trigonometric polynomial of degree N in the rotator angles and an analytic function of p, $q \leq k_0$. So we consider the domain:

$$W(k_0, s_0) \equiv W_0 := \{|p|, |q| \leq k_0,\ I \in V_0(\varepsilon) \subset \mathbb{C}^n,\ \psi \in \mathbb{T}^n \times (-is_0, is_0)\} ,$$

where $V_0(\varepsilon)$ is some n-rectangle such that $\Pi_{I_j} V_0(\varepsilon) = O(\frac{\sqrt{\varepsilon}\omega_j}{a_j})$.

We write f in Taylor series: $f(p,q,\psi) = \sum f_{\nu,k,h} p^k q^h e^{i\nu\cdot\psi}$. For all $s < s_0$, $k < k_0$, we use the weighted norm:

$$|f|_{k,s} \equiv |f|_{W(k,s)} = \sum e^{s|\nu|} |f_{\nu,l,h}| k^{2(l+h)} e^{i\nu\cdot\psi}\,.$$

Definition A.1. *Given a sub-lattice $\Lambda \in \mathbb{Z}^n$ and a point set $D \in V_0(\varepsilon)$, we say that D is $K-\beta$ non-resonant modulo Λ if for all $I \in D$,*

$$|\omega(I)\cdot\nu| \geq \beta,\ \forall \nu : \nu \notin \Lambda \cap |\nu| \leq K\,.$$

If Λ_0 is the lattice generated by the N frequencies ($\nu_i \in \mathbb{Z}^n$) of f, we set $\Lambda \in \Lambda_0$ to be the sub-lattice orthogonal to the fast components. We choose a point set D in the following manner: let P be the set of vectors ω such that[aa] $\omega\varepsilon^{-\frac{1}{2}} \in \Omega$ such that $|\omega_1 \cdot \nu_F| \geq \frac{\gamma}{|\nu_F|^{\tau_F}}$ for an order one γ.

Given $r_0 \in \mathbb{R}^+$, the domain $D(r_0)$ is a thickening of P such that $\forall\, I \in D(r_0)$, there exists $\omega \in P$ such that

$$|AI - \omega| \leq \varepsilon^{\alpha+\frac{1}{2}} r_0$$

for $r_0 < R$; in the following we will set $b = \frac{1}{2} + \alpha$.

Lemma A.2. *$D_0 \equiv D(r_0)$ is $\beta - K$ non-resonant modulo Λ with*

$$K = \left(\frac{\gamma}{4R}\varepsilon^{-b}\right)^{\frac{1}{1+\tau_F}}\,, \quad \beta = (\gamma)^{\frac{1}{1+\tau_F}} (4R\varepsilon^b)^{\frac{\tau_F}{1+\tau_F}}\,.$$

Proof. Given $I \in D(r_0)$ $\omega(I) = AI$ is $\varepsilon^b r_0$-close to an $\omega \in P$, so

$$|\omega(I)\cdot\nu| \geq |\omega_1\cdot\nu_F| - (\varepsilon^b|\omega_2||\nu| + \varepsilon^b r_0|\nu|)$$

with $r < |\omega_2| < R$. Thus we set

$$\varepsilon^b|\omega_2|\gamma^{-1}|\nu|^{\tau_F+1},\ \varepsilon^b r_0|\nu|\gamma^{-1}|\nu|^{\tau_F+1} < \frac{1}{4}\,. \qquad \square$$

We construct an analytic symplectic transformation (μ-close to identity) of the form:

$$Id + \mu S(I',p',\psi,q) = Id + \sum_{1<l\leq\frac{K}{N}} \mu^l \sum_{\nu\neq\Lambda}^{|\nu|\leq lN} S^{(l)}_{\nu,k,h}(p')^k q^h e^{i\nu\cdot\psi}\,,$$

that brings the Hamiltonian A.1 in the normal form[bb]

$$(I',AI') + \sqrt{\varepsilon}G_1(pq,\sqrt{\varepsilon}) + \mu g_1(\psi'_S, I', p', q', \varepsilon, \mu) + \mu^{\frac{K}{N}} f_1(\psi', I', \varepsilon, \mu)$$

[aa]Recall that we are now working on Hamiltonian (1.1) so that all quantities must be appropriately rescaled by $\sqrt{\varepsilon}$.

[bb]The separation between the integrable G_1 and the non-integrable g_1 is kept only because we will eventually set up a KAM scheme for the slow variables, so we need to estimate the size of the integrable part.

in a suitable domain $D'(r_1) \times \mathbb{T}^n_{s_1} \times B^2_{k_1}$, where

$$D'(r) = D(r) \cap \{I : \exists \omega \in P \text{ such that } |a_n I_n - \omega_n| \le r_0 \varepsilon\}\,.$$

The Hamilton–Jacobi equations are

$$\mu A I' \cdot S_\psi + \frac{1}{2}\mu^2 |AS_\psi|^2 + \sqrt{\varepsilon} G(qp' + \mu q S_q)) = \sqrt{\varepsilon} G_1(p'q + p'S_{p'}, \mu)$$
$$+ \mu g_1(\psi_S + \mu S_{I'}, I', p', q + \mu S_{p'}, \varepsilon, \mu) - \mu f(p' + S_q, q, \psi) + o(\mu^K) \tag{A.2}$$

and we assume that we can find some domain $D'(r) \times \mathbb{T}^n_s \times B^2_k$ such that the functions in A.2 are evaluated inside their domain of analyticity. We will call Π_Λ the natural projection on functions NOT depending on the fast angles: $\Pi_\Lambda f(\psi, p, q) = g(\psi_S, p, q)$ and Π_J the natural projection on functions depending only on $J = pq$:

$$F = \sum F_{\nu,k,h} p^k q^h e^{i\nu\cdot\Phi} \; \Pi_J F = \sum F_{0,h,h} (pq)^h\,.$$

We are looking for a symplectic transformation such that $(\Pi_\Lambda) S = 0$. We will solve the Hamilton–Jacobi equations recursively and determine the functions $G_1(J,\mu) = \sum_{i\ge 0} \mu^i G_1(J;i)$ and $\mu g_1(\psi_S, I, p, q, \mu) = \sum_{i\ge 1} \mu^i g_1(\psi_S, I, p, q; i)$. The first order leads to[cc]

$$G_1(J,0) = G(J)\,, \quad G_1(J,1) = \frac{1}{\sqrt{\varepsilon}} \Pi_J f\,, \quad g_1(\psi_{1S}, I', p', q', 1) = (\Pi_\Lambda - \Pi_J) f\,,$$

$$S^{(1)}_{\nu,k,h} = -\frac{f_{\nu,k,h}}{i[I' \cdot \nu] + (k-h)\sqrt{\varepsilon} G_J(p'q)}\,.$$

The term $i[I' \cdot \nu] + (k-h)\sqrt{\varepsilon} G_J(0) = D(\nu,k,h)$ is the "small denominator" that in our case (i.e. up to order $\frac{K}{N}$) admits the lower bound $D(\nu,k,h) \ge \beta$ provided that $I' \in D'(r_0)$. The higher order terms are determined recursively; we set $\mu S^{<l} = \sum_{h=1}^{l-1} \mu^h S^{(h)}$ and $[f(\mu)]_l = \frac{1}{l!} \partial^l_\mu f|_{\mu=0}$.

$$G_1(J,l) = \frac{1}{\sqrt{\varepsilon}} \Pi_J \left[\left(\mu^2 \frac{1}{2} |AS^{<l}_\psi|^2 + \sqrt{\varepsilon} G(qp' + \mu q S^{<l}_q) \right) - \sqrt{\varepsilon} G_1(p'q + p'S^{<l}_{p'}, \mu) \right.$$
$$\left. \left. - \mu g_1(\psi_S + \mu S^{<l}_{I'}, I', p, q + \mu S^{<l}_{p'}, \varepsilon, \mu) + \mu f(p' + S^{<l}_q, q, \psi) \right]_l \right)\,,$$

the remaining resonant terms are in $\mu g_1 = \sum_{m=1}^\infty \mu^m g_1(\psi_s, I', p', q; m)$:

$$g_1(\psi_S, I', p', q; l) = (\Pi_\Lambda - \Pi_J) \left[\left(\frac{1}{2}\mu^2 |AS^{<l}_\psi|^2 + \sqrt{\varepsilon} G(qp' + \mu q S^{<l}_q) \right) \right.$$
$$- \sqrt{\varepsilon} G_1(p'q + p'S^{<l}_{p'}, \mu) - \mu g_1(\psi_S + \mu S^{<l}_{I'}, I', p, q$$
$$\left. \left. + \mu S^{<l}_{p'}, \varepsilon, \mu) + \mu f(p' + S^{<l}_q, q, \psi) \right]_l \right)\,,$$

[cc]Notice that the pendulum and rotator terms cannot cancel each other, this is a consequence of the locality of our analysis.

the terms of order μ^l such that $\nu \neq \Lambda$ fix the value of $S^{(l)}_{\nu,k,h}$. We expand the Taylor series only in this expression. The symbol $\{k_i\}^r_k$ means the set of vectors in $\mathbb{N}^r$ such that $\sum_{i=1}^r k_i = k$, while $\{\nu_i\}^r_\nu$ is the set of r vectors in $\mathbb{Z}^n$ such that $\sum_{i=1}^r \nu_i = \nu$.

$$\begin{aligned} S^{(l)}_{\nu,k,h} = -\frac{1}{D(\nu,k,h)}\Bigg[&\sum_{\nu^{(1)}+\nu^{(2)}=\nu} \frac{1}{2} S^{(m)}_{\nu^{(1)},k_1,h_1} S^{(l-m)}_{\nu^{(2)},k-k_1,h-h_1}(\nu^{(1)}, A\nu^{(2)}) \\ &+ \sqrt{\varepsilon}\sum_{r\geq 2}^{l} \sum_{\substack{\{k_i\}^r_k,\{h_i\}^r_h,\\ \{l_i\}^r_l,\{\nu_i\}^r_\nu}} \left(\frac{1}{r!}\partial^r_J G(p'q)\Pi^r_{i=1} S^{(l_i)}_{\nu_i,k_i,h_i} h_i \right. \\ &+ \sum_{r\geq 1} \sum_{\substack{\{k_i\}^r_{k+r},\{h_i\}^r_h,\\ \{l_i\}^r_{l-1},\{\nu_i\}^r_{\nu_1}}} \frac{1}{r!}\partial^r_{p'} f(p',q,\Psi)\Pi^r_{i=1} k_i S^{(l_i)}_{\nu_i,k_i,h_i} \\ &- \sqrt{\varepsilon}\sum_{m=0}^{l-2}\sum_{r\geq 2}^{l-m} \sum_{\substack{\{k_i\}^r_k,\{h_i\}^r_h,\\ \{l_i\}^r_{l-m},\{\nu_i\}^r_\nu}} \left(\frac{1}{r!}\partial^r_J G_1(p'q;m)\Pi^r_{i=1} S^{(l_i)}_{\nu_i,k_i,h_i} k_i \right. \\ &- \sum_{m=1}^{l-1} \sum_{\substack{k_a+k_b=k,\\ h_a+h_b=h}} \sum_{\substack{l_a+l_b=l-m,\\ \nu_a+\nu_b=\nu}} \sum_{r\geq 0}^{l-m} \sum_{\substack{s\geq 0\\ r+s\geq 1}}^{l-m-r} \sum_{\substack{\{k_i\}^r_{k_a+r},\{h_i\}^r_{h_a}\\ \{l_i\}^r_{l_a},\{\nu_i\}^r_{\nu_a}}} \sum_{\substack{\{k_j\}^s_{k_b},\{h_j\}^s_{h_b},\\ \{l_i\}^s_{l_b},\{\nu_i\}^s_{\nu_b}}} \\ &\times \left(\frac{1}{r!s!}\partial^r_q \partial^s_{\psi_S} g_1(\psi_S, I', p', q; m)\Pi^r_{i=1} k_i (S^{(l_i)}_{\nu_i,k_i,h_i} \Pi^s_{j=1} \nabla_I S^{(l_i)}_{\nu_j,k_j,h_j}\right. \end{aligned} \tag{A.3}$$

To avoid proliferation of symbols we will set:
$\max(|f|_0, |G|_0) = E_0$ and choose $r_0 > 1$ so that $r_0\varepsilon^b \geq r_0\varepsilon \equiv \lambda_0 > k_0^2$. Finally we will call $b_j = b$ if $j = 1, \ldots, n-1$ and $b_n = 1$.

Proposition A.3. *Consider the nested domains*: $D_l \equiv D'(r_l) \times \mathbb{T}^n_{s_l} \times B^2_{k_l}$ *where* $r_l = \frac{1}{2}r_0 e^{-l\xi}$, $s_l = s_0(1 + l\xi)$ *and* $k_l = \frac{1}{2}k_0 e^{-l\xi}$; *the following bounds hold*[dd]:

$$|S^{(l)}_{\nu,k,h}|_l \leq C_1(l-1)!B^{l-1}, \ |G_1(J,l)|_l \leq C_2(l-1)!B^{l-1},$$

$$|g_1(\psi_S, I', p', q; l)|_l \leq C_3(l-1)!B^{l-1}$$

with $C_1 = \frac{E_0}{\beta}$, $C_2 = C_3 = E_0$ *and* $B = c\frac{E_0^2}{\beta^2 k_0^4 \xi^2}$ *for some small enough order one* c. *Moreover the so defined transformation is a biholomorphism*: $D_K \to D_0$ *provided that* $\xi = \frac{s_0}{4K}$, $\mu BK < 1$. *Thus the system can be written in normal form for*

$$\mu < \frac{\beta^2 k_0^4 \xi^2}{K^3} \tag{A.4}$$

in the domain $D(r) \times T^n_s \times B^2_k$, *with* $r = \frac{1}{2}r_0 e^{-s_0/4}$, $k = \frac{1}{2}k_0 e^{-s_0/4}$ *and* $s = s_0/4$.

[dd] By $|f|_l$ we mean $|f|_{D_l}$.

Remark A.4. Notice that for systems with one fast time scale, the domain P coincides with the whole $W(k, s_0/2)$ as all one-dimensional vectors of norm one are Diophantine with order one γ. Moreover in this case $\beta = O(1)$ as well, so if we choose $K = \frac{c}{\sqrt{\varepsilon}}$, the bound on μ is $\mu \leq \varepsilon^{\frac{5}{2}}$.

Remark A.5. Notice that if we choose $K = O_\varepsilon(1)$, we can perform some steps of the normal form theory for $\mu < \varepsilon$.

Proof. We proceed by induction, using the analyticity assumptions on G and f.

We will assume that the desired bounds hold for all $l < m$ and that $G_1(J, l)$ and $g_1(\psi_S, I', p', q, l)$ are analytic in D_{m-1}. This implies that the transformations

$$I = I' + \mu S_\psi^{<m}\,, \quad \psi' = \psi + \mu S_{I'}^{<m}\,,$$
$$p = p' + \mu S_q^{<m}\,, \quad q' = q + \mu S_{p'}^{<m}$$

are well defined and $D_m \to D_0$ if

$$\max(|\mu S_q^{<m}|_m, \mu S_{p'}^{<m}|_m) \leq \frac{1}{4} k_m\,, \quad |\mu S_{\psi_j}^{<m}|_m \leq \frac{1}{4} r_m \varepsilon^{b_j}\,,$$
$$|\mu S_{I'}^{<m}|_s \leq \frac{1}{4} s_0\,, \quad |\mu S_{\psi, I'}^{<m}|_m < 1\,.$$

Substituting the bounds in these inequalities (and using Cauchy estimates for the derivatives) we obtain the constraint $\mu \max(\frac{8C_1}{k_0^2\xi}, \frac{8C_1}{\lambda_0\xi^2}) < 1$ provided that $\mu K B \leq \frac{1}{2}$. Having verified the analyticity of the transformation up to order m, we use analytic bounds on G, G_1 and g_1 and the assumed bounds on the lower orders to bound $G_1(J; m)\ S^{(m)}$ and $g_1(\psi_S, I', p', q; m)$. We repeatedly use the inequality:

$$\sum_{\{k_i \geq 1\}_{i=1}^a : \sum_i k_i = k} \prod_{i=1}^a (k_i - 1)! \leq (k-1)!$$

Let us first consider $S^{(m)}$, it is composed of five sums. In each we substitute the Cauchy estimates and the bounds coming from the inductive hypothesis.

(1) The sum of quadratic terms is bounded by $(k-1)! B^{k-1} \frac{C_1^2}{s_0^2 \xi^2 \beta B}$.

(2) The terms due to G are bounded by

$$\frac{\sqrt{\varepsilon} E_0}{\beta} (m-1)! B^m \sum_{r \geq 2} \left(\frac{4C_1}{k_0^2 \xi B}\right)^r \leq \frac{8\sqrt{\varepsilon} E_0 C_1^2}{k_0^4 \xi^2 \beta B} (m-1)! B^{m-1}$$

provided that $\frac{4C_1}{k_0^2 \xi B} < \frac{1}{2}$.

(3) The terms due to f are bounded by

$$\frac{E_0}{\beta} (m-1)! B^{m-1} \sum_{r \geq 1} \left(\frac{2C_1}{k_0^2 \xi B}\right)^r \leq \frac{4 E_0 C_1}{k_0^2 \xi \beta B} (m-1)! B^{m-1}$$

provided that $\frac{2C_1}{k_0^2 \xi B} < \frac{1}{2}$.

(4) The terms due to G_1 has the same bound as (2) if we fix $C_2 = E_0$.

(5) If we fix $C_3 = E_0$ as well, the terms due to g_1 are bounded by

$$\frac{E_0}{\beta}(m-1)!B^{m-1}\sum_{r\geq 0}\sum_{s\geq 0, r+s\geq 1}\left(\frac{2C_1}{k_0^2\xi B}\right)^r\left(\frac{2C_1}{\lambda_0\xi B}\right)^s \leq \frac{4C_1E_0}{\beta k_0^2\xi B}(m-1)!B^{m-1}$$

provided that $\frac{2C_1}{\lambda_0\xi B} \leq \frac{2C_1}{k_0^0\xi B} < \frac{1}{2}$.

These five bounds must be all set $< \frac{1}{5}C_1$. It is easily seen that, as $b \leq 1$ and $\lambda_0 \geq k_0^2$, all the desired bounds are implied by $\max(\frac{8C_1}{\lambda_0\xi^2}, \frac{8\sqrt{\varepsilon}E_0C_1}{k_0^4\xi^2\beta B}) \leq \frac{1}{5}$. Now we discuss the bounds on G_1 and g_1. There are always the same five terms times a factor $\frac{\beta}{\sqrt{\varepsilon}}$ for G_1 and β for g_1. So all the bounds are verified if, $\frac{E_0C_1}{k_0^4\xi^2\beta B} \leq c \ll 1$. We fix $C_1 = \frac{E_0}{\beta}$ as this comes from the first order and $B = c\frac{E_0^2}{k_0^4\xi^2\beta^2}$. □

A.3. *Proof of Lemma 4.23*

Proof. We define an operator

$$O_j[g] := Q_j(G) + \frac{1}{2}\sum_{i=0,1} x_j^i\Im(x_j^{[i]}G)$$

so that $\psi_j^k = O_j(F_j^k)$. We consider the vector $V = \binom{O_j(G)}{\partial_t(O_j(G))}$. By the definition of O_j, it is a solution of $\dot{V} = L_jV + G$ where L_j is the 2×2 matrix:

$$L_j = \begin{vmatrix} 0 & 1 \\ \delta_{j0}g(q^0(t)) & 0 \end{vmatrix}, \quad g(\psi_0) = -\partial_{\psi_0}^2 f^0(\psi_0)\,.$$

We derive with respect to t:

$$\ddot{V} = L_j\dot{V} + (\dot{L}_j)V + \dot{G})\,.$$

The first line of the solution $\dot{V}$ is

$$\partial_t(O_j(G)) = O_j(-\delta_{j0}\dot{q}^0(t)\partial_0^3 f^0(t)O_0(G) + \dot{G})$$

plus the first component of a solution of the homogeneous equation $t \to W(t)X$ that we determine via the initial data. $O_j^t(F)$ is zero for $t = 0$, and the initial datum is determined by the boundedness condition $\partial_t(O_j(G))|_{t=0} = \Im^0(x_j^0G)$, so

$$\partial_t(O_j(G)) = O_j\left(\frac{2}{c}\delta_{j0}x_0^0(t)\partial_0^3 f^0(t)O_0(G) + \dot{G}\right) + x_j^0(t)\Im^0(x_j^0G)$$

and as G is odd we can substitute $Q_j(G) = O_j(G)$.

Next we notice that the vectors $W^i = \binom{x_0^0(t)}{\dot{x}_0^0(t)}, \binom{\sigma(t)x_0^1(t)}{\sigma(t)\dot{x}_0^1(t)}$ are solutions of the system $\dot{W} = L_0W$. So we apply the time derivative and obtain[ee]

$$\dot{x}_0^i = \frac{2}{c}Q_0(x_0^ix_0^0\partial_0^3 f^0(t)) + \delta_{i1}\sigma(t)x_0^0\,. \tag{A.5}$$

[ee]We are using the fact that $O_0(\sigma(\tau)F) = \sigma(t)O_0(F)$.

The last term is added to have the right behavior in $t = 0$ ($d_t\sigma(t)x_0^1|_0 = 1$).

$$O_j\left(\frac{2}{c}\delta_{j0}x_0^0(t)\partial_0^3 f^0(t)O_0(G) + \dot{G}\right) + x_j^0(t)\Im^0(x_j^0 G)$$

$$= Q_j\left(\frac{2}{c}\delta_{j0}x_0^0(t)\partial_0^3 f^0(t)Q_0(G) + \dot{G}\right)$$

$$+ \frac{1}{2}\sum_i \left(x_j^i \Im x_j^{[i]}\left(\frac{2}{c}\delta_{j0}x_0^0(t)\partial_0^3 f^0(t)Q_0(G) + \dot{G}\right) + x_j^0(t)\Im^0(x_j^0 G)\,. \qquad \text{(A.6)}$$

The last two sums cancel each other via relation A.5 and Proposition 2.6(i) and (iv), for $j = 0$, and using the fact that if $j \neq 0$, then $\dot{x}_j^0 = 0$ and $\dot{x}_j^1 = \sigma(t)$:

$$\frac{1}{2}\sum_i \left(x_j^i \Im x_j^{[i]}\left(\frac{2}{c}\delta_{j0}x_0^0(t)\partial_0^3 f^0(t)Q_0(G) + \dot{G}\right) + x_j^0(t)\Im^0(x_j^0 G)$$

$$= \frac{1}{2}\sum_i \left(x_j^i \Im G Q_0\left(\frac{2}{c}\delta_{j0}x_j^{[i]}(t)x_0^0(t)\partial_0^3 f^0(t)\right)\right.$$

$$- \frac{1}{2}\Im(\dot{x}_j^{[i]}G) + x_j^0(t)\Im^0(x_j^0 G)\,. \qquad \text{(A.7)}$$

$\square$

References

[1] V. I. Arnold, Instability of dynamical systems with several degrees of freedom, *Sov. Math. Dokl.* **5** (1964), pp. 581–5.

[2] L. Chierchia and G. Gallavotti, Drift and diffusion in phase space, *Annales de l'IHP*, *Section Physique Théorique* **60** (1994), pp. 1–144; see also the *Erratum* **68** (1998), 135.

[3] G. Gallavotti, Twistless KAM tori, quasi flat homoclinic intersections and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A Review, *Rev. Math. Phys.* **6** (1994), 343–411.

[4] G. Gallavotti, G. Gentile and V. Mastropietro, Separatrix splitting for systems with three time scales, *Comm. Math. Phys.* **202** (1999), 197–236.

[5] L. Chierchia, Arnold instability for nearly-integrable analytic Hamiltonian systems, *Proc. Workshop "Variational and Local Methods in the Study of Hamiltonian Systems"*, A. Ambrosetti, G. F. Dell'Antonio eds. Trieste, 1994.

[6] G. Gallavotti, G. Gentile and V. Mastropietro, Melnikov approximation dominance. Some examples, *Rev. Math. Phys.* **11** (1999), 451–461.

[7] G. Gallavotti, *Reminiscences on Science at the I.H.E.S. A Problem on Homoclinic Theory and a Brief Review.*

[8] M. Berti and P. Bolle, A functional analysis approach to Arnold diffusion, *Annales de l'IHP*, *Section Analyse Non-Lineaire* **19**, 4 (2002), pp. 795–811.

[9] G. Gallavotti, G. Gentile and V. Mastropietro, Hamilton–Jacobi equation and existence of heteroclinic chains in three time scales systems, *Nonlinearity* **13** (2000), 323–340.

[10] C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer-Verlag (Graduate texts in mathematics, 207), Berlin, Heidelberg, New York, 2001.

[11] S. Lang, *Algebra*, Addison Wesley Publishing Co. 1993.
[12] G. Benettin and G. Gallavotti, *Stability of Motions Near Resonances in Quasi-Integrable Systems*.
[13] Pöschel, Nekhoroshev estimates for quasi-convex Hamiltonian systems, *Math. Z.* **213** (1993), 187–216.
[14] V. I. Arnold, Small denominators and problems of stability of motion in classical and celestial mechanic, *Russ. Math. Surveys* **18:6** (1963), pp. 85–191.
[15] M. Berti and P. Bolle, Fast Arnold diffusion in systems with three time scales, *Discrete and Continuous Dynamical Systems*, *Series A* **8**, 3 (2002), pp. 795–811.
[16] U. Bessi, An approach to Arnold diffusion through the calculus of variations, *Nonlinear Analysis T. M. A.* **26** (1996), pp. 1115–35.
[17] U. Bessi, L. Chierchia and E. Valdinoci, Upper bounds on Arnold diffusion time via Mather theory, *J. Math. Pures Appl.* **80-1** (2001), pp. 105–129.
[18] B. Bollobas, *Graph Theory*, Springer-verlag (Graduate texts in mathematics, 63), Berlin, Heidelberg, New York, 1979.
[19] L. Chierchia and E. Valdinoci, A note on the construction of Hamiltonian trajectories along heteroclinic chains, *Forum Math. 12*, **2** (2000), 247–254.
[20] V. I. Arnold, Dynamical Systems 3, *Encyclopedia of Math. Sci.* **3** (1963).
[21] A. Delshams, V. G. Gelfreich, V. G. Jorba and T. M. Seara, Exponentially small splitting of separatrices under fast quasi-periodic forcing, *Comm. Math. Phys.* **150** (1997), 35–71.
[22] G. Gallavotti, G. Gentile and V. Mastropietro, *Lindstedt Series and Hamilton–Jacobi Equation for Hyperbolic Tori in Three Time Scales Problems*.
[23] G. Gentile, Whiskered tori with prefixed frequencies and Lyapunov spectrum, *Dynamic Stability Systems* **10**, 3 (1995), 269–308.
[24] V. Gelfreich, Melnikov method and exponentially small splitting of separatrices, *Physica D* **101** (1997), pp. 227–248.
[25] P. Holmes and J. Marsden, Melnikov method and Arnold diffusion for perturbations of integrable Hamiltonian systems, *J. Math. Phys.* **23** (1982), pp. 669–675.
[26] P. Lochak, J. P. Marco and D. Sauzin, *On the Splitting of Invariant Manifolds in Multidimensional Hamiltonian Systems*, preprint Université Jussieu.
[27] H. Poincaré, *Les Methodes Nouvelles de la Mécanique Céleste*, Vol. I-III Paris, 1892, 93, 99.
[28] J. P. Serre, *Trees*, Springer Verlag, 1980.
[29] E. Valdinoci, Families of whiskered tori for *a priori* stable/unstable Hamiltonian systems and construction of unstable orbits, *Math. Phys. Electr.* **6** (2000).