Commun. Math. Phys. 315, 643–697 (2012)
Digital Object Identifier (DOI) 10.1007/s00220-012-1532-x

Communications in
**Mathematical
Physics**

An Example of a Nearly Integrable Hamiltonian System with a Trajectory Dense in a Set of Maximal Hausdorff Dimension

Vadim Kaloshin[1], **Maria Saprykina**[2]

[1] Department of Mathematics, University of Maryland, College Park, MD 20742-4015, USA.
E-mail: kaloshin@math.umd.edu
[2] Institutionen för Matematik, KTH, 10044 Stockholm, Sweden. E-mail: masha@kth.se

Received: 1 June 2011 / Accepted: 26 January 2012
Published online: 18 September 2012 –

Abstract: The famous ergodic hypothesis suggests that for a typical Hamiltonian on a typical energy surface nearly all trajectories are dense. KAM theory disproves it. Ehrenfest (The Conceptual Foundations of the Statistical Approach in Mechanics. Ithaca, NY: Cornell University Press, 1959) and Birkhoff (Collected Math Papers. Vol 2, New York: Dover, pp 462–465, 1968) stated the quasi-ergodic hypothesis claiming that a typical Hamiltonian on a typical energy surface has a dense orbit. This question is wide open. Herman (Proceedings of the International Congress of Mathematicians, Vol II (Berlin, 1998). Doc Math 1998, Extra Vol II, Berlin: Int Math Union, pp 797–808, 1998) proposed to look for an example of a Hamiltonian near $H_0(I) = \frac{\langle I,I \rangle}{2}$ with a dense orbit on the unit energy surface. In this paper we construct a Hamiltonian $H_0(I)+\varepsilon H_1(\theta, I, \varepsilon)$ which has an orbit dense in a set of maximal Hausdorff dimension equal to 5 on the unit energy surface.

Contents

1. Introduction

The famous question called the ergodic hypothesis, formulated by Maxwell and Boltzmann, suggests that for a typical Hamiltonian on a typical energy surface all, but a set of zero measure of initial conditions, have trajectories covering densely this energy surface itself. However, KAM theory showed that for an open set of nearly integrable systems there is a set of initial conditions of positive measure of almost periodic trajectories. This disproved the ergodic hypothesis and forced to reconsider the problem. A quasi-ergodic hypothesis, proposed by Ehrenfest [E] and Birkhoff [Bi], asks if a typical Hamiltonian on a typical energy surface has a dense orbit. A definite answer whether this statement is true or not is still far out of reach of modern dynamics. There was an attempt to prove this statement by E. Fermi [Fe], which failed (see [G] for a more detailed account). To simplify the quasi-ergodic hypothesis, M. Herman [H] formulated the following question:

Can one find an example of a C^∞–Hamiltonian H in a C^r–small neighborhood of $H_0(I) = \frac{\langle I,I\rangle}{2}$ such that on the unit energy surface $\{H^{-1}(1)\}$ there is a dense trajectory?

Many people believe that such examples do exist and are C^∞–generic (see [E, Bi, Ar1]). In this paper we make a step in the direction of answering Herman's question.

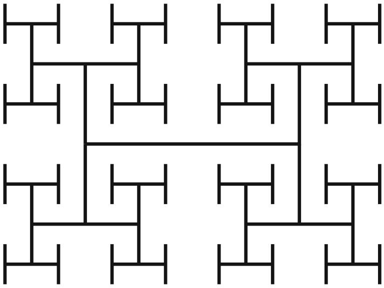

Fig. 1. Construction of an H-tree

For any r we construct a C^∞–Hamiltonian, which is C^r-close to $H_0(I) = \frac{\langle I,I\rangle}{2}$ and has a trajectory dense in a set of maximal Hausdorff dimension on the energy surface 1/2. Here is the exact statement.

Let $\theta = (x, y, z) \pmod 1 \in \mathbb{T}^3$, $I \in \mathbb{R}^3$ and $H_0(I) = \frac{\langle I,I\rangle}{2} = |I|^2/2$ be the unperturbed Hamiltonian, where $\langle I, I\rangle$ is the dot product in $\mathbb{R}^3$.

Theorem 1. *For any* $2 \le r \le \infty$ *there is a* C^∞*–smooth* C^r*–small*[1] *perturbation of* H_0*:*

$$H_\varepsilon(\theta, I) = H_0(I) + \varepsilon H_1(\theta, I, \varepsilon) \tag{1}$$

such that on the energy surface $\{H_\varepsilon = 1/2\}$ *there is an orbit* $(I, \theta)(t)$ *of* H_ε *whose closure has Hausdorff dimension 5.*

This paper is followed by [KZZ], where it is announced that the closure of an orbit can have a positive 5-dimensional Lebesgue measure. Naturally, there is a number of ideas in common.

The basic idea of the proof of this theorem is as follows. Consider the unperturbed system. The phase space of this system is $\mathbb{T}^3 \times \mathbb{R}^3$ and the energy surface has the form $\mathbb{T}^3 \times \mathbb{S}^2$, where $\mathbb{S}^2$ is the 2-dimensional unit sphere in $\mathbb{R}^3$. Denote by $\mathcal{K}$ the set of totally irrational vectors, i.e., $\omega \in \mathbb{R}^3$ such that for all $k \in \mathbb{Z}^3 \setminus \{0\}$ we have $\langle k, \omega\rangle \neq 0$. We shall construct the following objects:

- a Cantor set $F^\infty \subset \mathbb{S}^2 \cap \mathcal{K}$ of Hausdorff dimension 2 of the type of H-tree (see Fig. 1, Sect. 2 for a sketch of the construction, and Sect. 7 for the details). The product $\mathcal{F}^\infty = \mathbb{T}^3 \times F^\infty \subset \mathbb{T}^3 \times \mathbb{R}^3$ is a set of Hausdorff dimension 5 in the phase space.
- an open set $\mathcal{U} \subset \mathbb{R}^3$ such that its boundary $\partial\mathcal{U}$ contains F^∞, i.e., $F^\infty \subset \partial\mathcal{U}$.
- a perturbation $\varepsilon H_1(I, \theta, \varepsilon)$ supported on $\mathbb{T}^3 \times \mathcal{U}$. In other words, the perturbation vanishes on $\mathcal{F}^\infty$ and, therefore, all the invariant tori of the unperturbed system H_0 in $\mathcal{F}^\infty$ are also invariant for $H_\varepsilon = H_0 + \varepsilon H_1$, and form a set of Hausdorff dimension 5 in the phase space. Moreover, all of them belong to the energy surface $\{H_\varepsilon = 1/2\}$ which coincides with $\{H_0 = 1/2\}$ on $\mathcal{F}^\infty$.

The main feature of this construction is that there is a trajectory of H_ε which visits an arbitrarily small neighborhood of each torus in $\mathcal{F}^\infty$. In particular, one can arrange that $\mathcal{F}^\infty$ contains tori with any prescribed irrational frequency. The second main result of this paper is the following.

[1] Notice that r can be equal to infinity.

Theorem 2. *Let $\vec{\omega}' = (\omega_1', \omega_2', \omega_3')$ and $\vec{\omega}'' = (\omega_1'', \omega_2'', \omega_3'')$ be a pair of totally irrational unit vectors, i.e., $\vec{\omega}', \vec{\omega}'' \in \mathcal{K}$ and $|\vec{\omega}'| = |\vec{\omega}''| = 1$. For any $2 \le r \le \infty$ there is a C^r-smooth function $H_1(\theta, I, \varepsilon)$ that vanishes on the invariant tori $\mathbb{T}^3_{\vec{\omega}'}$ and $\mathbb{T}^3_{\vec{\omega}''}$ and an $\varepsilon_0 > 0$ such that the following holds. For any $\varepsilon < \varepsilon_0$ the Hamiltonian $H_\varepsilon(\theta, I) = H_0(I) + \varepsilon H_1(\theta, I, \varepsilon)$ has an orbit $(\theta, I)(t)$ on the energy surface $\{H_\varepsilon = 1/2\}$ whose ω-limit set contains $\mathbb{T}^3_{\vec{\omega}'}$ and α-limit set contains $\mathbb{T}^3_{\vec{\omega}''}$.*

2. Outline of the Proof

We give here a heuristic outline of the proof of the main theorem. Since the Hamiltonians we shall study are close to $|I|^2/2$, we have $\dot\theta = I + \varepsilon \partial_I H_1$, so in the outline below we shall make practically no distinction between velocity $\dot\theta$ and I (see (2-3)). A general method of constructing a diffusing trajectory is to make the velocity vector $\dot\theta$ follow a resonant plane of the form $\pi = \{k_1\dot x + k_2\dot y + k_3\dot z = 0\}$ with $k = (k_1, k_2, k_3) \in \mathbb{Z}^3 \setminus \{0\}$ (single resonance). Consider the energy surface $\{H_\varepsilon(\theta, I) = 1/2\}$. For $\varepsilon = 0$, its projection onto the I-space is the unit two-sphere $\mathbb{S}^2$. For small ε this projection, $\mathbb{S}^2_\varepsilon$, is a small smooth deformation of the sphere.

Consider the action component first. We shall choose a countable number of resonant planes $\{\pi_i\}_{i \in \mathbb{Z}_+}$ in such a way that the union of segments $l_i \subset \pi_i \cap \mathbb{S}^2$ over i form a fractal set F whose closure $\overline{F}$ has the maximal Hausdorff dimension on $\mathbb{S}^2_\varepsilon$. The set F is such that the stereographic projection from $\mathbb{S}^2$ onto $\mathbb{R}^2$ transforms F into a set close to the so-called H-tree, see Fig. 1. This model set is denoted by F^{mod}.

An H-tree on the plane is a fractal set, obtained as the closure of a countable union of self-similar H-letters. More exactly, fix $0 < \lambda < 1$ and define sets

— $F^{\mathrm{mod}}_{\lambda,1}$ consists of one horizontal segment of length one and two vertical segments of length λ centered at the end points of the horizontal segment. Thus, F_1 has the H-form.

— $F^{\mathrm{mod}}_{\lambda,n}$ consists of 2^{2n} translated copies of $\lambda^{2n} F_1^{mod}$ so that the center of each copy coincides with a vertex of a vertical segment inside $F^{\mathrm{mod}}_{\lambda,n-1}$,

$$F^{\mathrm{mod}}_\lambda = \overline{\cup_n F^{\mathrm{mod}}_{\lambda,n}}.$$

Notice that for $0 < \lambda \le 1/\sqrt{2}$ we have no loops in F^{mod}_λ. It is easy to compute that the Hausdorff dimension of the H-tree $F^{\mathrm{mod}}_{1/\sqrt{2}}$ is 2, and $F^{\mathrm{mod}}_{1/\sqrt{2}}$ fills an open set on the plane.

In our case the H-tree $\overline{F}$ belongs to the energy surface $\{H_\varepsilon = 1/2\}$. It is modeled on the above plain fractal and has Hausdorff dimension 2, but it differs from the model set in the following features. Each segment of F is the intersection of the energy surface $\{H_\varepsilon = 1/2\}$ with an appropriate resonant plane passing through the origin. Moreover, λ is not constant during the inductive construction and approaches $\frac{1}{\sqrt{2}}$ from below. Therefore, n^{th} order H-sets are only almost self-similar to those of order $n-1$. We define

$$F^\infty = \left(\bar F \setminus F\right) \cap \mathcal{K} \quad \text{and} \quad \mathcal{F}^\infty = F^\infty \times \mathbb{T}^3,$$

where $\mathcal{K}$ stands for the set of rationally independent vectors in $\mathbb{R}^3$. We shall prove that F^∞ has Hausdorff dimension 2.

The Hamiltonian H_ε will be constructed so that $H_\varepsilon(\theta, I) = H_0(I)$ for all $I \in F^\infty$. We shall construct a trajectory of H_ε whose velocity vector changes along the segments of F in a prescribed order. Moreover, it shadows the whole $\mathcal{F}^\infty$ in the limit. To make the diffusion process work, the resonant planes have to satisfy certain conditions which we specify in Sect. 7.

The construction of the Hamiltonian H_ε is done through a construction of the corresponding Lagrangian L_ε. Recall that for a strictly convex Hamiltonian $H(\theta, I)$ one can define a Legendre transform

$$L(\theta, \dot\theta) = \max_I \langle \dot\theta, I\rangle - H(\theta, I) =: \mathcal{L}(H) \tag{2}$$

and a diffeomorphism

$$\mathbb{L} : (\theta, I) \to (\theta, \dot\theta) = (\theta, \partial_I H). \tag{3}$$

The standard formalism (see e.g. [Ar2]) states that orbits of the Hamiltonian equation of H are mapped into orbits of the Euler-Lagrange equation of L:

$$\frac{d}{dt}\partial_{\dot\theta} L = \partial_\theta L.$$

Moreover, the Legendre transform is involutive, i.e., $\mathcal{L}(L) = \mathcal{L}^2(H) = H$. Thus, our system, governed by the Hamiltonian (1), corresponds to the Euler-Lagrange equation with the Lagrangian

$$L_\varepsilon(\theta, \dot\theta, \varepsilon) = \frac{\langle \dot\theta, \dot\theta\rangle}{2} + \varepsilon L_1(\theta, \dot\theta, \varepsilon). \tag{4}$$

The Lagrangian L_1 will have the form

$$U(\theta, \dot\theta, \varepsilon) - \varepsilon^{r+1}\beta(\theta, \dot\theta, \varepsilon), \quad \theta = (x, y, z) \in \mathbb{R}^3. \tag{5}$$

Remark. In the case $r = \infty$, replace the factor ε^{r+1} in front of the perturbation term by $e^{-\frac{1}{\varepsilon}}$. Then we use a standard metric in the space of C^∞-functions $f : \Omega \mapsto \mathbb{R}$. Recall the definition: consider the following family of semi-norms on Ω: $\|\cdot\|_{i,j} = \max_{|\alpha|\le i}\sup_{x\in K_j} |\partial^\alpha f(x)|$, where K_j is a countable family of compact sets that exhaust Ω. The standard metric in the space of $C^\infty(\Omega)$-functions is

$$d(f, g) = \sum_{i=0} \sup_{j\in\mathbb{N}} \frac{\|f - g\|_{i,j}}{1 + \|f - g\|_{i,j}} 2^{-i}.$$

We shall construct solutions of the Euler–Lagrange flow of (4) with the properties stated in the Main Theorems. The two important regimes of the diffusion process are *single resonant* (when the velocity vector changes close to a single resonant plane) and *double resonant* (when the velocity vector changes close to the intersection of two resonant planes).

We look for the solution to the Euler-Lagrange equation by minimizing the action $\int L$ and, using essentially the ideas of Mather and Fathi, carefully construct a variational problem with constraints which has an interior solution.

Recall that nearly integrable Hamiltonian systems with 3 degrees of freedom and 2.5 degrees of freedom are very similar. Locally the former one can be reduced to the later. It is well known that dynamics of typical perturbations of Hamiltonian systems with 3 degrees of freedom restricted to an energy surface and near a single resonance (and away from double resonances of small order) is a priori unstable. This means that the underlying Hamiltonian system has a 3-dimensional normally hyperbolic cylinder Λ and its 3-dimensional stable and unstable manifolds intersect transversally. Attempts of constructing diffusing orbits in this setting face serious difficulties. For example, if one uses the geometrical method proposed by Arnold [Ar] of shadowing a chain of *whiskered tori* for generic perturbations, then a problem of *large gaps between tori on* Λ arises (see e.g. [DLS]). Application of variational methods à la Mather [Be,CY,Ma1] can overcome this problem as well as modification of geometric methods [DLS,GL]. Analysis of the separatrix map in [T] does not see such a large gap problem. However, in all known proofs the details are extremely involved.

The study of perturbations of the product of a pendulum and a rotator was initiated in the famous paper by Arnold [Ar]. In this paper he discussed a particular example, which later was treated by Bessi [Bs1] for 2-degrees of freedom and recently by Zhang [Zha] for any number of degrees of freedom using variational methods. Arnold diffusion close to a double resonance for typical perturbations is a much more involved problem, and is studied by Mather [Ma1,Ma2]. An example of diffusion close to a double resonance was described by Bessi [Bs2]. Other examples of Arnold diffusion can be found in [Bo1,Bo2,LM,MS,BK]. Recently Zheng [Zhe] proved existence of Arnold diffusion for a special planar 5 body problem, proposed by Moeckel [Moe], heavily using its a priori unstable structure.

Theorem 2 is related to the work of Douady [Do], where he studies stability of totally elliptic fixed points (see also [KMV]). The present work can be viewed as an exposition of Mather's and Fathi's ideas and their application to the construction of examples of Arnold diffusion. In [KL1,KL2] we present an elementary example of Arnold diffusion of an expository nature. This example motivates further examples of Arnold diffusion for lattices [KLS].

The main idea of the construction in this paper is elementary. We perturb the integrable Hamiltonian system so that it is *maximally integrable* in the domain of diffusion. To make this idea more specific we discuss two model problems: for a double resonance and a single resonance.

2.1. Double resonant model. For a double resonance, i.e., for I in a neighborhood V_k of the intersection of two resonant planes $\pi_k \cap \pi_{k+1} \cap \mathbb{S}^2_\varepsilon$, we define H_ε in the following way: in some set of symplectic coordinates $(\theta_k, I_k),\ \theta_k = (x_k, y_k, z_k) \in \mathbb{T}^3$,

$$H_\varepsilon(\theta_k, I_k) = A_k H_0(I_k) - \varepsilon_{k-1} \cos^2 \frac{\pi y_k}{2} + \varepsilon_k \cos^2 \frac{a_k \pi z_k}{2} + \varepsilon_k^{r+1} \beta(\theta_k; \varepsilon_k)$$

for some positive integers $A_k, a_k \in \mathbb{Z}_+$, and β being a periodic C^∞ function whose support is localized at the vertices of a certain affine lattice $\Gamma_k \subset \mathbb{R}^3$ such that it contains the integer lattice $\mathbb{Z}^3 \subset \Gamma_k$. Notice that away from the support of β the Hamiltonian system is given as a product of

the pendulum $\times$ *the pendulum* $\times$ *the rotator*

and is completely integrable (i.e., have 3 first integrals in involution). This makes analysis of objects associated to this Hamiltonian system fairly transparent.

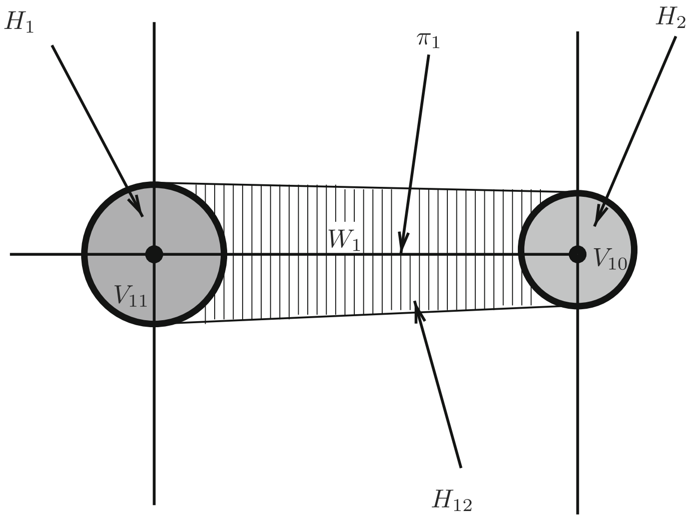

Fig. 2. Integrable deformations

2.2. Deformation problem. In order to construct a diffusing trajectory in a neighborhood W_k of $\pi_k \cap \mathbb{S}^2_\varepsilon$ (a single resonance) and avoid a long list of problems of nearly integrable a priori unstable systems we require that the system H_ε without the β-term is completely integrable in the whole of W_k. Hence, the following model problem arises. Given two completely integrable systems H_1 and H_2, where (Fig. 2)

$$H_1(\theta, I) = \left(\frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon \cos^2 \frac{\pi y}{2} \right) + \frac{I_z^2}{2} - \varepsilon' \cos^2 \frac{m\pi z}{2}$$

is defined for $(\dot{x}, \dot{y}) = (I_x, I_y)$ close to $(1, 0)$ and

$$H_2(\theta, I) = \left(\frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon \cos^2 \frac{\pi(ay - bx)}{2} \right) + \frac{I_z^2}{2} - \varepsilon' \cos^2 \frac{m\pi z}{2}$$

is defined for $(\dot{x}, \dot{y}) = (I_x, I_y)$ close to $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, a, b and m integers, find *a completely integrable system* H_{12}, defined in a neighborhood of the segment $\{(\dot{x}, \sqrt{1 - \dot{x}^2}) : \dot{x} \in [\frac{1}{\sqrt{2}}, 1]\}$, and coinciding with H_1 and H_2 in their respective domains of definition. We provide this construction in Sect. 6 for H_{12} having a certain special form and being close to H_0.

2.3. Single resonant model. The result of this construction is that near a single resonance the underlying Hamiltonian system is a product of

a completely integrable system of 2 *degrees of freedom* $\times$ *the pendulum.*

This system is completely integrable and can be analyzed without additional serious difficulties. After that we are ready to construct a variational problem with constraints à la Mather.

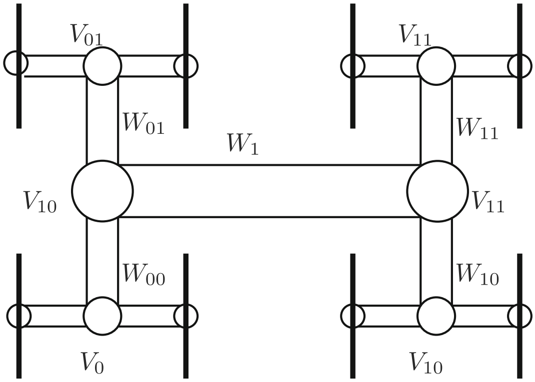

Fig. 3. Neighborhoods of single and double resonances

2.4. Deformation term. A system H_ε without "bumps" $\beta(\theta, I, \varepsilon)$ is completely integrable, therefore, no diffusion is possible. The support of β is contained in a countable collection of sets $\mathcal{L}_{\mathbf{n}}(I, \varepsilon)$ called *lenses*. A precise definition of the lenses is given in Sect. 7.3. To get an idea, suppose that the action variable I is inside V_{10}. The lenses $\mathcal{L}_{\mathbf{n}}$ are defined as convex open sets centered around points $\mathbf{n}$ of a certain lattice $(\mathbb{Z}^3 \subset\neq)\Gamma \subset \mathbb{R}^3$ in the configuration space (see Sect. 7.3) (Fig. 3):

$$\mathcal{L}_{\mathbf{n}}(I, \varepsilon) = \{(\theta, I) : \ |\theta - \mathbf{n}| \leq \varepsilon\, r(I), \ \mathbf{n} \in \Gamma\},$$

where the radius $r = r(I)$ is a function of I. The "bump"-function β is supported on the lenses. On a lens $\mathcal{L}_{\mathbf{n}}$ it has the form

$$\beta(\theta, I, \varepsilon) = \zeta\left(\frac{|\theta - \mathbf{n}|}{\varepsilon r(I)}\right)$$

where, to be specific, we take $\zeta([0, 1/2]) = 1$, $\zeta([1, \infty)) = 0$ with ζ being C^∞-smooth on $\mathbb{R}$, monotone decreasing on $[1/2, 1]$ and even. Actually, for our results to hold, ζ can be any smooth nonnegative function supported on $(-1, 1)$.

3. Heuristic Explanation of the Variational Method and Analytic Components of the Proof

We would like to construct an orbit $\gamma = \{(I, \theta)(t)\}_{t\in\mathbb{R}}$ from Theorem 1 as a limit of orbits $\gamma^n = \{(I^n, \theta^n)(t)\}_{t\in\mathbb{R}}$ solving certain variational problems with constraints. Loosely speaking an n^{th} orbit γ^n shadows the n^{th} generation of the underlying H-tree F^∞. The definitions below are valid for any convex Hamiltonian system on the cotangent bundle of a compact manifold M. However, we shall use them for 2 and 3 dimensional tori.

Associate to each Hamiltonian $H(I, \theta)$ on T^*M a convex in I Lagrangian using the Legendre transform (2) and the Legendre map (3). According to the general theory, the orbits of the flow of Hamiltonian H are mapped into the orbits of the Euler-Lagrange flow

$$\frac{d}{dt}\partial_{\dot\theta} L = \partial_\theta L.$$

Let $J = [a, b]$ be an interval of time. Solutions to the Euler-Lagrange equation are local extremals of the action $\int L(\dot\gamma(t), \gamma(t))dt$.

Suppose a Lagrangian L

- is *positive definite* in $\dot\theta$, i.e., $\partial^2_{\dot\theta\dot\theta}L$ is positive definite for each $(\theta, \dot\theta) \in TM$;
- has *super-linear growth*, i.e., $L(\dot\theta, \theta)/|\dot\theta| \to \infty$ as $\dot\theta \to \infty$;
- the Euler–Lagrange flow is *complete*, i.e., solutions are defined for all time.

Call a Lagrangian satisfying these conditions *a Tonelli Lagrangian*. It is known that the Legendre transform of a Tonelli Lagrangian is a diffeomorphism, i.e., smooth and has a smooth inverse. *In this paper we only consider Tonelli Lagrangians.*

Set

$$A^T(\theta_0, \theta_1) := \int_0^T L(\gamma(t), \dot\gamma(t))\, dt = \inf_{\tilde\gamma} \int_0^T L(\tilde\gamma(t), \dot{\tilde\gamma}(t))\, dt, \tag{6}$$

where the infimum is taken over all absolutely continuous curves $\tilde\gamma$ such that $\tilde\gamma(0) = \theta_0$, $\tilde\gamma(T) = \theta_1$. The curve $\gamma(t)$ is called a *minimizer* of $A^T(\theta_0, \theta_1)$. By Tonelli's Theorem and completeness assumption minimizers exist and are C^1-curves. Therefore, we get the same result minimizing in the space of C^1 curves. A minimizer is always a solution to the Euler-Lagrange equation above.

Before we sink into discussion of mathematical objects from Aubry-Mather theory we give a heuristic explanation of why we are interested in them. It turns out that if collections $i \in \mathbb{Z}$ of constraints (sections S_i of lenses $\mathcal{L}_i$, $\mathcal{L}_i$ being centered at certain points $\mathbf{n}_i$ of the lattices Γ_k), $\sigma_i > 0$ and time durations T_i^* are appropriately chosen, the following variational problem with constraints has an interior minimum:

$$\min_{\theta_i \in S_i,\ |T_i - T_i^*| < \sigma} \quad \sum_{i \in \mathbb{Z}} A^{T_i}(\theta_i, \theta_{i+1}). \tag{7}$$

We shall see that a solution is an orbit of the Euler-Lagrange flow (and, therefore, of the Hamiltonian flow) and satisfies the conclusion of Theorem 1.

To prove the existence of an interior minimum we analyze the action $A^{T_i}(\theta_i, \theta_{i+1})$ and determine that there is a collection of smooth periodic functions $u_i : \mathbb{T}^3 \to \mathbb{R}$, vectors $c_i^* \in \mathbb{R}^3$ such that $|c_i^*| \approx 1$, and constants $\alpha_i \approx 1$, slowly changing with i and having the property

$$\begin{aligned} A^{T_i}(\theta_i, \theta_{i+1}) = u_i(\theta_{i+1}) - u_i(\theta_i) + \alpha_i T_i - c_i^*(\theta_{i+1} - \theta_i) \\ - \operatorname{bump}^-(\theta_i) - \operatorname{bump}^+(\theta_{i+1}) + \frac{\text{“}K\text{”}}{T_i^*}, \end{aligned} \tag{8}$$

where $\operatorname{bump}^\pm(\theta_i)$ are smooth small non-negative functions supported in $S_i \cap \mathcal{L}_i$ having at least a certain ε-dependent value, and notation “K” stands for a term bounded in absolute value by K. Consider minimization w.r.t. θ_i. Notice that it is necessary to prove that for each $i \in \mathbb{Z}$ we have

$$\min_{\theta_i \in S_i} \quad A^{T_{i-1}}(\theta_{i-1}, \theta_i) + A^{T_i}(\theta_i, \theta_{i+1}) \tag{9}$$

has an interior minimum for any fixed pair T_{i-1}, T_i within the constraints. (We shall slightly modify this variational problem later.) Apply the formula above, concentrating only on θ_i-dependent terms

$$= u_{i-1}(\theta_i) - u_i(\theta_i) + (c_i^* - c_{i-1}^*)\theta_i - \text{bump}^+(\theta_i) - \text{bump}^-(\theta_i)$$
$$+ \text{terms independent of } \theta_i.$$

Since we assumed that u_i's and c_i^*'s change slowly with i, for T_i^*, T_{i+1}^* sufficiently large, the terms ($\text{bump}^\pm(\theta_i)$) can dominate. If "$-$ bumps" have minima of sufficient depth, then minimum w.r.t. θ_i is interior. To see interior minimum w.r.t. T_i we need a slightly more involved discussion of similar nature as above.

Unite this discussion with geometric discussion of single and double resonant models from the previous section. Based on an increment $\Delta\mathbf{n}_i = \mathbf{n}_{i+1} - \mathbf{n}_i$ and its direction $\omega_i = \Delta\mathbf{n}_i/|\Delta\mathbf{n}_i|$ we determine if the minimizer stays close to a single or a double resonance (see test (40) from Sect. 5) and determine $c_i = c(\omega_i)$.

— For a single resonance we construct u_c, compute $\alpha(c)$, in Sect. 4. Then we prove (8) in Lemma 4.
— Regularity of functions u_c and $\alpha(c)$ with respect to c and $c = c(\omega)$ with respect to ω, $|\omega| = 1$ is described in Lemma 1.
— For the double resonance case we construct u_c and compute $\alpha(c)$, in Sect. 5. Then we prove (8) in Lemma 8.
— Regularity of u_c and $\alpha(c)$ with respect to c and $c = c(\omega)$ with respect to ω, $|\omega| = 1$ follows from the explicit form of the Hamiltonian/Lagrangian (see Lemma 1 and Sect. 5.1 for details).

In order to prove that (7) has an interior minimum with respect to time T_i we analyze approximation formula (8) more closely and show that the minimum should be located close to a certain number $T_i^* = T^*(\Delta\mathbf{n}_i)$. See Lemma 4 for a single resonance and Lemma 8 for a double resonance.

Now we connect the aforementioned objects to those in Mather-Fathi theory. As we mentioned by the increment Δn_i we can determine a cohomology class $c_i \in H^1(\mathbb{T}^3, \mathbb{R}) \simeq \mathbb{R}^3$. Based on these cohomology classes and the underlying Lagrangian L we can define an α-function $\alpha : H^1(\mathbb{T}^3, \mathbb{R}) \to \mathbb{R}$ and a family of smooth periodic functions u_c on $\mathbb{T}^3$.

Due to a very special structure of Lagrangian L and a deep insight of Fathi [Fa1,Fa], these functions $u_c(\theta)$, $c \in \mathbb{R}^3$ define families of invariant sets as follows. Fix $c \in \mathbb{R}^3$ and consider the graph

$$G(u_c) = \{(\theta, \nabla u_c(\theta) + c) : \ \theta \in \mathbb{T}^3\} \subset T^*\mathbb{T}^3.$$

In our integrable situations the function u_c for all θ satisfies

$$H(\theta, \nabla u_c(\theta) + c) = \alpha(c),$$

and $G(u_c)$ gives rise to invariant sets. The set $G(u_c)$ contains one-sided and two-sided minimizers of L. More exactly, it contains so-called sets of Aubry, Mather, and Mañé (see Sect. 3.2). Using so called Fenchel-Legendre transform, based on c we can determine average rotation vector of these orbits. This is a family of functions which was defined in a more general convex situation by Fathi [Fa,Fa1]. In general situation, however, nice smooth dependence of u_c's on c and θ is lost.

To sum up, *if we have a good understanding of α-functions and the family of functions u_c, then we can solve the variational problem and with additional geometric arguments—prove Theorem 1.*

3.1. Plan of the paper. At this point the proof divides into four parts:

- First we consider model Hamiltonians, and prove that the action obeys estimate (8), and related objects are regular. We prove that the local minimum for model variational problems is interior, both with respect to θ and T. This is done in Sect. 4 for the single resonance, and is Sect. 5 for the double resonance case.
- In Sect. 6, the deformation problem from introductory Sect. 2.2 is solved.
- In Sect. 7, we select resonances and build the Hamiltonian $\varepsilon H_1(\theta, I, \varepsilon)$;
- Finally, in Sect. 8, we construct a variational problem by choosing $\mathbf{n}_i$'s.

Necessary properties of the pendulum are discussed in the Appendix.

In the next section we discuss some necessary aspects of the Aubry-Mather theory.

3.2. Description of Aubry–Mather–Mane Invariant sets for convex integrable Hamiltonian systems. Consider a Tonelli Lagrangian L as defined in Sect. 3. To set up an appropriate variational problem we need to modify the action functional in (6) as follows.

Let η_c be a closed one form on M with cohomology class $c = [\eta_c]_M \in H^1(M, \mathbb{R})$. In this case we are interested in $M = \mathbb{T}^d$, and we have $H^1(M, \mathbb{R}) \simeq \mathbb{R}^d$. A C^1-smooth curve $\gamma(t), \gamma : [0, T] \to M$ is called a *c−minimizer* or a *c-minimal* curve if it minimizes action of $(L - \eta_c)$ over all C^1 curves satisfying $\tilde{\gamma}(0) = \theta_0$ and $\tilde{\gamma}(T) = \theta_1$:

$$A^T_{\eta_c}(\theta_0, \theta_1) := \int_0^T (L - \eta_c)(\dot{\gamma}(t), \gamma(t))\, dt = \inf_{\tilde{\gamma}} \int_0^T (L - \eta_c)(\dot{\tilde{\gamma}}(t), \tilde{\gamma}(t))\, dt.$$

If the time interval J is not compact, a C^1 curve $\gamma : J \to M$ is called *c-minimal* if for every compact interval $\tilde{J} \subset J$ it is c-minimal. Denote by $\mathcal{G}_c \subset TM$ the set of c-minimal orbits. It turns out that c-minimality does not depend on a choice of η_c within its cohomology class [Ma].

To define another set of minimal orbits, called *a Mather set*, denoted $\mathcal{M}_c$, we need to extend the definition of the action along a C^1 curve to action along a probability measure. Let $\mathcal{M}_L$ be the set of Borel probability measures on TM, invariant for the Euler-Lagrange flow. For any $\nu \in \mathcal{M}_L$, the action $A_c(\nu)$ is defined as $\int (L - \eta_c) d\nu$. One shows that c-minimality of $\mu \in \mathcal{M}_L$ is independent of η_c with $[\eta_c] \in H^1(M, \mathbb{R})$ [Ma].

A probability measure μ is called *c-minimal invariant measure* if

$$A_c(\mu) = \min_{\nu \in \mathcal{M}_L} \int (L - \eta_c) d\nu.$$

Denote $\mathcal{M}_c$ the closure of the union of the supports of $c-$minimal invariant measures and call it the *Mather set*. One can show that $\mathcal{M}_c \subset \mathcal{G}_c$. A function

$$\alpha(c) := -A_c(\mu) : H^1(\mathbb{T}^3, \mathbb{R}) \to \mathbb{R}$$

is called *α−function*, where μ is a $c-$minimal invariant measure. Define also a rotation vector of a measure, denoted by $\omega(\mu) \in H_1(M, \mathbb{R})$ by the unique vector such that

$$(\omega(\mu) \cdot c) = \int \eta_c d\mu$$

for any closed one form on M with cohomology class $c = [\eta_c]_M \in H^1(M, \mathbb{R})$. Again the right-hand side is independent of a choice of η_c. Define

$$\beta(\omega) = \min_{\nu \in \mathcal{M}_L,\ \rho(\nu)=\omega} \int L d\nu.$$

It follows from the definition that $\beta(\omega)$ and $\alpha(c)$ are Legendre transforms of each other, i.e.

$$\beta(\omega) = \max_c \langle \omega, c\rangle - \alpha(c), \qquad \alpha(c) = \max_\omega \langle \omega, c\rangle - \beta(\omega). \tag{10}$$

Notice that in the case that β is differentiable at ω, then c, where the maximum for $\beta(\omega)$ is achieved satisfies $c = \beta'(\omega)$. In particular, the c-minimal measure has rotation vector ω.

Further properties of α and β functions of integrable systems.

- Let $L(\theta_1, \theta_2, \dot\theta_1, \dot\theta_2) = L_1(\theta_1, \dot\theta_1) + L_2(\theta_2, \dot\theta_2)$ and $\alpha_1(c_1)$ and $\alpha_2(c_2)$ be α-functions of L_1 and L_2, then the α-function of L is the sum: $\alpha(c_1, c_2) = \alpha_1(c_1) + \alpha_2(c_2)$. Similarly the β-function of L is the sum $\beta(\omega_1, \omega_2) = \beta_1(\omega_1) + \beta_2(\omega_2)$.
- Bernard [Be] proved that α and β functions are symplectic invariants. For any $(\phi, J) \in \Phi^{-1}(U)$ we have that $\beta(\omega) = \tilde L(\omega)$, and for any c such that $c = \partial_{\dot\phi} \tilde L(\omega)$ we have $\alpha(c) = \tilde H(c)$. Since in our case $\tilde H$ and $\tilde L$ are (direct sums of) strictly convex functions, in U there is one-to-one correspondence between c and ω such that both maps $c \to \omega$ or $\omega \to c$ are smooth.

Our main object in what follows will be properties of minimizers of

$$A_c^T(\theta_0, \theta_1) = \inf_{\tilde\gamma} \int_0^T (L(\dot{\tilde\gamma}(t), \tilde\gamma(t)) - \eta_c \dot{\tilde\gamma}(t) + \alpha(c))\, dt.$$

As $T \to \infty$, the limit is independent of the choice of η_c with $[\eta_c]_M = c$ and defines a c-minimizer.

For the Lagrangians that we study in this paper there is a natural choice of η_c and even finite time c-minimizers are parts of one-sided minimizers. In general, this is definitely not true.

3.3. Definitions of Aubry, Mather, and Mañé sets. In order to define Mañé and Aubry sets, denoted usually by $\mathcal{N}_c$ and $\mathcal{A}_c$ resp., we need to define Mañé potential. Fix a closed one-form η_c with $[\eta_c]_M = c$ and denote

$$L_c(\dot\theta(t), \theta(t)) = L(\dot\theta(t), \theta(t)) - \eta_c \dot\theta(t) + \alpha(c).$$

Define Mañé potential (see e.g. [CI])

$$A_c^\infty(\theta_0, \theta_1) = \inf_{T>0} A_c^T(\theta_0, \theta_1).$$

It is not difficult to see that $A_c^\infty(\theta_0, \theta_1)$ is Lipschitz in θ_0 and θ_1. Let $\gamma : \mathbb{R} \to M$ be a C^1 curve

- It is called *c-semi-static* (or *one-sided c-minimizer*) if for any $T > 0$ we have

$$A_c^{\infty}(\gamma(0), \gamma(T)) = \int_0^T L_c(\dot{\gamma}(t), \gamma(t))\, dt.$$

- It is called *c−static* (or *c-minimizer*) if for $T > 0$ we have

$$-A_c^{\infty}(\gamma(T), \gamma(0)) = \int_0^T L_c(\dot{\gamma}(t), \gamma(t))\, dt.$$

A static curve (or a c-minimizer) is always a semi-static curve (or a one-sided c-minimizer).

Denote the set of c–semi-static and c–static (resp. one-sided c-minimizer and c-minimizer) orbits by $\mathcal{N}_c$ and $\mathcal{A}_c$ respectively. Usually $\mathcal{N}_c$ is called *a Mañé set* and $\mathcal{A}_c$ is called *an Aubry set*.

$$\mathcal{M}_c \subseteq \mathcal{A}_c \subseteq \mathcal{N}_c \subseteq \mathcal{G}_c.$$

Lifts of these invariant sets to the tangent space TM are denoted by $\tilde{\mathcal{M}}_c$, $\tilde{\mathcal{A}}_c$, $\tilde{\mathcal{N}}_c$, and $\tilde{\mathcal{G}}_c$ respectively. Indeed c–static or c–semi-static (one-sided c-minimizer and c-minimizer) orbits can be considered in TM. *From now on we stick to names: one-sided c-minimizer and c-minimizer.*

In the autonomous setting $\tilde{\mathcal{N}}_c = \tilde{\mathcal{G}}_c$, while in the time-periodic case one can have $\tilde{\mathcal{M}}_c \subsetneq \tilde{\mathcal{G}}_c$ (see e.g., [CY,FM]). Mather [Ma] proved that $\tilde{\mathcal{M}}_c$ and $\tilde{\mathcal{A}}_c$ are Lipschitz graphs over M with respect to the natural projection $\pi : TM \to M$. It is not difficult to show that $\mathcal{A}_c = \{\theta \in M : A_c^{\infty}(\theta, \theta) = 0\}$.

Bernard [Be] proved that Aubry , Mañé and Mather sets (denoted $\mathcal{A}_c$, $\mathcal{N}_c$ and $\mathcal{M}_c$, resp.,) are symplectic invariants. Thus, Mañé sets of a completely integrable system are corresponding tori in TM of L. Their images under the Legendre map are invariant tori on H. These Hamiltonian tori are given as graphs of gradients of smooth functions $u_c(\theta)$, which depend smoothly on c. Below we consider several examples and present Aubry and Mañé sets in the corresponding cases.

Example 1. Let $H(\theta, I) = \frac{|I|^2}{2}$. Then its Legendre transform is the Lagrangian $L(\theta, \dot{\theta}) = \frac{|\dot{\theta}|^2}{2}$. Define a Lagrangian

$$L_c(\theta, \dot{\theta}) = L(\theta, \dot{\theta}) - c \cdot \dot{\theta} + \frac{|c|^2}{2} = \frac{|\dot{\theta} - c|^2}{2},$$

where the one form $\eta_c \equiv c$ and $\frac{|c|^2}{2} = \alpha(c)$.

Notice that the Euler-Lagrange flow of L_c and L are the same. Moreover, the minimization of action

$$\int_0^T L_c(\gamma(t), \dot{\gamma}(t))\, dt, \quad \text{where } \gamma(0) = \theta_0 \quad \text{and } \gamma : [0, T] \to \mathbb{T}^3 \quad \text{is } C^1$$

leads to the straight line with constant velocity c. The union of such trajectories is the Mañé set $\mathcal{N}_c = \{\dot{\theta} = c\}$, which coincides with the Aubry set $\mathcal{A}_c$ and the Mather set $\mathcal{M}_c$.

Example 2. c-minimizers of a pendulum. Let $H(y, I) = \frac{I^2}{2} - \varepsilon \cos^2 \frac{\pi y}{2}$, $\varepsilon > 0$ fixed, and let L be the corresponding Lagrangian:

$$L(y, \dot{y}) = \frac{\dot{y}^2}{2} + \varepsilon \cos^2 \frac{\pi y}{2}.$$

Define a function ν of two scalar variables: $y \in \mathbb{T}$—coordinate, and $0 \leq h$—energy by

$$\nu(y, h) = \sqrt{2\left(h + \varepsilon \cos^2 \frac{\pi y}{2}\right)},$$

and let

$$c = c(h) = \frac{1}{2}\int_0^2 \nu(y, h)dy = \int_0^2 \sqrt{2\left(h + \varepsilon \cos^2 \frac{\pi y}{2}\right)}\, dy > 0. \tag{11}$$

We omit dependence of c on ε not to overload notations. In fact, c can be interpreted as the mean (w.r.t. y) speed of a trajectory of the standard pendulum with energy h over one period. Denote $c(0)$ by c^+. It corresponds to the "upper" homoclinic solution of the saddle. For each $c \geq c^+$ the inverse function $h = h(c)$ is well defined. We define the c-Lagrangian as

$$L_c(y, \dot{y}) = L(y, \dot{y}) - \nu(y, h)\dot{y} + h = \frac{1}{2}(\dot{y} - \nu(y, h))^2.$$

The c-action is defined as

$$A_c^T(y_0, y_1) = \inf_\gamma \int_0^T L_c(y, \dot{y})\, dt$$

over all C^1 curves connecting y_0 to y_1 in time T. For $c = c(h)$ and any $h \geq 0$ a c-minimizer of the action $A_c^T(y_0, y_1)$ has energy h and clearly satisfies $\dot{y} = \nu(y, h)$ at every point. By definition such curves are trajectories of L.

Let $\nu(y, c) = \nu(y, h)$ for h such that $c = c(h)$. For $c \geq c^+$ this is uniquely defined. Set

$$u_c(y) = \int_0^y \nu(w, c)\, dw - c\, y \quad \text{for} \quad c \geq c^+. \tag{12}$$

By definition, for $c = c(h)$ we have: $H(y, u_c'(y) + c) = H(y, \nu(y, c)) = h$. So, $u_c(y)$ is a function whose gradient defines the graph of the c-minimizer.

Let us show that for each $|c| < c^+$, the corresponding c-minimizer with $\dot{y} > 0$ is located at the point $(y, \dot{y}) = (1, 0)$, which is the saddle point. To see this, use the fact that c-minimizers do not depend on a choice of η_c and let $\eta_c = \nu(y, c^+) + (c - c^+)$ and $L_c = L_{c_+} + (c - c^+)\dot{y}$. Then

$$\int_0^T L_c(y, \dot{y})\, dt = \int_0^T \frac{1}{2}(\dot{y} - \nu(y, c^+))^2 - (c - c^+)(y_T - y_0).$$

Thus, each loop $y_T = y_0 + 2$ has non-negative cost and A_c^T tends to infinity. If $\dot{y} < 0$, arguments are similar.

Thus, the saddle $(1, 0)$ is the Aubry set $\mathcal{A}_c$. It also coincides with the Mather set $\mathcal{M}_c$. This shows that one-sided c-minimizers (from the Mañé set $\mathcal{N}_c$) should approach the origin. The only orbits approaching the origin are separatrices of the pendulum. This implies that the set of one-sided minimizers $\mathcal{N}_c$ for the aforementioned c is the separatrix. To match these requirements we define $u_c(y)$ as follows:

$$u_c(0) = 0, \quad u_c'(y) = u_{c^+}'(y) \quad \text{for } 0 < c < c_+, \quad 0 < y < 2,$$

and extend by periodicity in y. Since the integral of $\int_0^2 (u_{c^+}'(y) - c)\, dy$ is not zero, this function has discontinuity at $y = 2k$, $k \in \mathbb{Z}$. Notice though that $u_c'(y)$ is a C^1-smooth periodic function away from the origin. Similarly using symmetry $L(y, \dot{y}) = L(y, -\dot{y})$ one can define $u_c(y)$ for $c < 0$ by $u_{-c}(y) = -u_c(y)$.

Example 3. c-minimizers of the direct product of two pendulums. Assume that

$$H(I, \theta; \varepsilon) = \frac{\langle I, I \rangle^2}{2} - \varepsilon' \cos^2 \frac{\pi y}{2} - \varepsilon \cos^2 \frac{\pi z}{2}.$$

In this setting we can describe the position of the sets $\tilde{\mathcal{A}}_c$ and $\tilde{\mathcal{N}}_c$ explicitly. There are three different cases:

1. The case $h' = h = 0$. Then $\tilde{\mathcal{N}}_c$ is a product of the circle, the separatrix of one pendulum, and the separatrix of the other pendulum. Most of trajectories on $\tilde{\mathcal{N}}_c$ are homoclinic to the periodic orbit $\tilde{\mathcal{A}}_c = \{(\theta, \dot{\theta}) : \dot{x} = c,\ \dot{y} = \dot{z} = y = z = 0\}$.
2. The case $h' > 0$ and $h = 0$. Then $\tilde{\mathcal{N}}_c$ is diffeomorphic to a product of a 2-dimensional (x, y)-torus and the separatrix of the z-pendulum. Most of trajectories on $\mathcal{N}_c$ are homoclinic to $\tilde{\mathcal{A}}_c = \{(\theta, \dot{\theta}) : \dot{\theta} = c,\ z = 0\}$.
3. The case $h' = 0$ and $h > 0$. Then $\tilde{\mathcal{N}}_c$ is diffeomorphic to a product of a 2-dimensional (x, z)-torus and the separatrix of the y-pendulum. Most of trajectories on $\mathcal{N}_c$ are homoclinic to $\tilde{\mathcal{A}}_c = \{(\theta, \dot{\theta}) : \dot{\theta} = c,\ y = 0\}$.

4. Diffusion in the Single-Resonance Case

We denote $\theta = (x, y, z) \in \mathbb{T}^3$, $I = (I_x, I_y, I_z) \in \mathbb{R}^3$. Let M be a 3×3 matrix with integer coefficients, and $M^* = (M^{tr})^{-1}$. Let $a \in \mathbb{Z} \setminus \{0\}$ and A be (large) integer constants. The model Hamiltonian in this section is

$$H(\theta, I, \varepsilon) = H^{\text{int}}(\theta, I, \varepsilon) - \varepsilon^{r+1} \beta(\theta, I, \varepsilon), \tag{13}$$

where $H^{\text{int}}(\theta, I, \varepsilon)$ is completely integrable, and $\beta(\theta, I, \varepsilon)$ is a C^r–smooth deformation. We assume that the integrable part has the form

$$H^{int}(\theta, I, \varepsilon) = A \left(H'((M\theta)_x, (M\theta)_y, (M^*I)_x, (M^*I)_y, \varepsilon) + \frac{|(M^*I)_z|^2}{2} \right) \\ - \varepsilon \cos^2 \frac{\pi}{2} a(M\theta)_z, \tag{14}$$

where $(v)_y$ denotes the y-component of a vector v, with the same notation for x and z. Let

$$(\tilde{\theta}, \tilde{I}) = (M\theta, M^* I).$$

The coordinate change $(\theta, I) \mapsto (\tilde{\theta}, \tilde{I})$ is symplectic. In terms of $(\tilde{\theta}, \tilde{I})$, the integrable Hamiltonian has the form

$$\tilde{H}^{\text{int}}(\tilde{\theta}, \tilde{I}, \varepsilon) = A\left(H'(\tilde{x}, \tilde{y}, \tilde{I}_x, \tilde{I}_y, \varepsilon) + \frac{\tilde{I}_z^2}{2}\right) - \varepsilon \cos^2 \frac{\pi}{2} a\tilde{z}. \tag{15}$$

Systems of this form are called *a priori unstable*. Proving existence of Arnold diffusion in this setting is a very difficult problem discussed in the Introduction. We choose our perturbation very carefully to avoid major difficulties of the generic case.

In this section we assume that the function $H'(\tilde{x}, \tilde{y}, \tilde{I}_x, \tilde{I}_y, \varepsilon)$ is ε-close to $\frac{\tilde{I}_x^2+\tilde{I}_y^2}{2}$. Moreover, for $|\tilde{I}_y| < \varepsilon^{1/10}$ we have

$$H'(\tilde{x}, \tilde{y}, \tilde{I}_x, \tilde{I}_y, \varepsilon) = \frac{\tilde{I}_x^2 + \tilde{I}_y^2}{2} - \varepsilon \cos^2 \frac{\pi}{2} \tilde{y},$$

compare with (58). The Hamiltonians we shall construct to solve our original problem do have this property, see Sect. 7.1–7.2 and Lemma 11.

Denote by h_y the "energy of the $\tilde{y}$-component":

$$h_y = \frac{\tilde{I}_y^2}{2} - \varepsilon \cos^2 \frac{\pi}{2} \tilde{y},$$

which is well defined for $|\tilde{I}_y| < \varepsilon^{1/10}$. Define h_z in the same way. In this section we assume that the energy of the $\tilde{z}$-component, h_z, is very close to zero (made precise by the construction),

$$\frac{\varepsilon}{100A} \le h_y \le \frac{1}{3A}, \tag{16}$$

and $\tilde{I}_x$ is such that $\tilde{H}^{int}(\tilde{\theta}, \tilde{I}, \varepsilon)$ is close to 1/2. For the above values of the actions we shall study our system as a "single resonance system", the single resonance being $\tilde{I}_z = 0$. The complementary values of the energies (i.e., when both h_z and h_y are close to zero), are considered in the next section. Here we do all the preliminary work needed to construct a diffusing trajectory whose actions change close to the "resonant plane" $\{\tilde{I}_z = 0\}$, and to the energy surface $E = \{H = 1/2\}$. A diffusion process for a system "close to (15)" with localized perturbation has been described in [KL1] for $H'(\theta_1, \theta_2, I_1, I_2) = \frac{I_1^2+I_2^2}{2}$. Here H' no longer has this special form and we need to modify calculations in [KL1].

The Lagrangian corresponding to $\tilde{H}(\tilde{\theta}, \tilde{I}, \varepsilon)$ by means of the Legendre transform has the form

$$\tilde{L}(\tilde{\theta}, \dot{\tilde{\theta}}) = \tilde{L}^{\text{int}}(\tilde{\theta}, \dot{\tilde{\theta}}) + \varepsilon^{r+1} \tilde{\beta}(\tilde{\theta}, \dot{\tilde{\theta}}, \varepsilon), \tag{17}$$

$$\tilde{L}^{\text{int}}(\tilde{\theta}, \dot{\tilde{\theta}}) = A\left(L'(\tilde{x}, \tilde{y}, \dot{\tilde{x}}, \dot{\tilde{y}}) + \frac{\dot{\tilde{z}}^2}{2}\right) + \varepsilon \cos^2 \frac{\pi}{2} a\tilde{z}, \tag{18}$$

where L' is completely integrable and C^{r-1}-close to $\frac{\dot{\tilde{x}}^2+\dot{\tilde{y}}^2}{2}$.

Remark 1. Consider the objects defined above: $L(\theta, \dot{\theta})$ is the Lagrangian corresponding to $H(\theta, I)$ by the Legendre transform, and $\tilde{L}(\tilde{\theta}, \dot{\tilde{\theta}})$ is the one corresponding to $\tilde{H}(\tilde{\theta}, \tilde{I})$. Then

$$\tilde{L}(\tilde{\theta}, \dot{\tilde{\theta}}) = L(\theta, \dot{\theta}).$$

Indeed, by definition of the Legendre transform, $L(\theta, \dot{\theta}) = I \cdot \dot{\theta} - H(\theta, I)$, and $\tilde{L}(\tilde{\theta}, \dot{\tilde{\theta}}) = \tilde{I} \cdot \dot{\tilde{\theta}} - \tilde{H}(\tilde{\theta}, \tilde{I})$. At the same time, we have: $H(\theta, I) = \tilde{H}(\tilde{\theta}, \tilde{I})$ and $\tilde{I} \cdot \dot{\tilde{\theta}} = (M^{tr})^{-1} I \cdot M\dot{\theta} = I \cdot \dot{\theta}$. The same relations hold for integrable parts of H and L.

By this remark,

$$L(\theta, \dot{\theta}) = L^{\text{int}}(\theta, \dot{\theta}) + \varepsilon^{r+1} \overline{\beta}(\theta, \dot{\theta}, \varepsilon), \tag{19}$$

$$L^{\text{int}} = A\left(L'((M\theta)_x, (M\theta)_y, (M\dot{\theta})_x, (M\dot{\theta})_y) + \frac{{(M\dot{\theta})_z}^2}{2} \right) + \varepsilon \cos^2 \frac{\pi}{2} a(M\theta)_z. \tag{20}$$

We shall look for solutions to the Euler-Lagrange equation of (20) as minimizers of the action functional (6).

4.1. Definition of $c(\omega)$, η_c and u_c. In order to study minimizers of the action functional $A^\tau(\theta_0, \theta_1)$, following Mather (see e.g. [Ma1]), it is useful to modify the Lagrangian by adding a closed one-form $\eta_c(\theta)\dot{\theta}$, where $\eta_c \in H^1(\mathbb{T}^3, \mathbb{R})$. The new Lagrangian has the same minimizers, but is easier to investigate. Here we define a suitable $\eta_c(\theta)$. It will be easier to define $\tilde{\eta}_{\tilde{c}}(\tilde{\theta})$ in terms of $\tilde{\theta}$, then pass to θ-coordinates. The form we choose only depends on the integrable parts of our Hamiltonians, namely $\tilde{H}^{int}$ and $\tilde{L}^{int}$. The form $\tilde{\eta}_{\tilde{c}}$ is a direct sum of two forms: $\tilde{x}\tilde{y}$-component and $\tilde{z}$-component. We define the $\tilde{z}$-components, $\tilde{\eta}_{\tilde{c}_z}(z)$ and $\tilde{u}_{\tilde{c}_z}(z)$, as in Example 2 of Sect. 3.2. In this case we only need to consider $h_z \geq 0$. To define the $\tilde{x}\tilde{y}$-component of $\tilde{\eta}_{\tilde{c}}$, consider the family of 2-dimensional tori

$$\mathcal{T}'_{\tilde{\omega}} = \{(\tilde{\theta}, \tilde{I}) : \ J_1(\tilde{x}, \tilde{y}, \tilde{I}_x, \tilde{I}_y) = f_1(\tilde{\omega}), \ J_2(\tilde{x}, \tilde{y}, \tilde{I}_x, \tilde{I}_y) = f_2(\tilde{\omega}), \ \tilde{z} = 1, \ \tilde{I}_z = 0\},$$

where tori are parameterized so that orbits on $\mathcal{T}'_{\tilde{\omega}}$ have rotation vector $\tilde{\omega} = (\tilde{\omega}_1, \tilde{\omega}_2)$. We can choose J_1 and J_2 to be close to $\tilde{I}_x$ and $\tilde{I}_y$. Then these tori are small deformations of $\{\tilde{I}_x = \tilde{\omega}_1, \tilde{I}_y = \tilde{\omega}_2, \tilde{z} = 1, \tilde{I}_z = 0\}$. Since these tori are Lagrangian submanifolds, there is a well defined function $\tilde{u}_{\tilde{c}_{xy}}(\tilde{x}, \tilde{y})$ such that

$$J_1(\tilde{x}, \tilde{y}, \nabla \tilde{u}_{\tilde{c}_{xy}}(x, y)) \equiv f_1(\tilde{\omega}), \qquad J_2(x, y, \nabla \tilde{u}_{\tilde{c}_{xy}}(x, y)) \equiv f_2(\tilde{\omega}),$$

(see, e.g., [MDS], Prop. 3.25). In other words, the gradient of $\tilde{u}_{\tilde{c}_{xy}}$ defines the invariant 2-torus $\mathcal{T}'_\omega$ and, therefore, near $\{\tilde{z} = 1, \ \tilde{I}_z = 0\}$ satisfies $\tilde{H}'(\tilde{x}, \tilde{y}, \nabla \tilde{u}_{\tilde{c}_{xy}}(\tilde{x}, \tilde{y}) + \tilde{c}_{xy}) =$ const. Let $\tilde{u}_{\tilde{c}_z}(\tilde{z})$ be a function defined for $\tilde{c}_z \geq \tilde{c}_z^+$ in (12). Set

$$\tilde{c}^+ = (\tilde{c}_{xy}, \tilde{c}_z^+), \quad u_{\tilde{c}^+}(\tilde{x}, \tilde{y}, \tilde{z}) = \tilde{u}_{\tilde{c}_{xy}}(\tilde{x}, \tilde{y}) + \tilde{u}_{\tilde{c}_z^+}(\tilde{z}).$$

Then $\tilde{H}(\tilde{\theta}, \nabla \tilde{u}_{\tilde{c}^+}(\tilde{\theta}) + \tilde{c}^+) =$const defines the set of trajectories homoclinic to $\mathcal{T}'_{\tilde{\omega}}$. In terms of the initial coordinates (θ, I), the invariant tori above are also obtained as the gradient of a function $u(\theta)$ which is related to $u_{\tilde{c}^+}$ by a linear transformation.

This invariant set can also be viewed in the configuration space. The Legendre transform relating $\tilde{H}$ and $\tilde{L}$ also defines a diffeomorphism $\mathcal{L} : (\tilde{\theta}, \tilde{I}) \to (\tilde{\theta}, \dot{\tilde{\theta}}) = (\tilde{\theta}, \partial_{\tilde{I}} \tilde{H}(\tilde{\theta}, \tilde{I}))$. This diffeomorphism maps the family $\{\mathcal{T}'_{\tilde{\omega}}\}_{\tilde{\omega}}$ into a family $\{\mathcal{T}_{\tilde{\omega}} = \mathcal{L}(\mathcal{T}'_{\tilde{\omega}})\}_{\tilde{\omega}}$. This map is close to the identity, thus, the family $\{\mathcal{T}_{\tilde{\omega}}\}_{\tilde{\omega}}$ consists of tori such that each $\mathcal{T}_{\tilde{\omega}}$ is close to $\{\dot{\tilde{x}} = \tilde{\omega}_1, \dot{\tilde{y}} = \tilde{\omega}_2, \tilde{z} = 1, \dot{\tilde{z}} = 0\}$. By the implicit function theorem there is a well defined one form $\tilde{\eta}_{\tilde{\omega}}(\tilde{\theta})$, whose graph coincides with the torus $\mathcal{T}_{\tilde{\omega}}$. This form is closed, since its graph is a Lagrangian submanifold (see, e.g., [MDS], Prop. 3.25). Let $\tilde{c} = \tilde{c}(\tilde{\omega})$ be the cohomology class of $\tilde{\eta}_{\tilde{\omega}}$. Note that, since $\mathcal{L}$ no longer satisfies $\tilde{I} = \dot{\tilde{\theta}}$, it is no longer true that $\nabla \tilde{u}_{\tilde{c}(\tilde{\omega})}(\tilde{\theta}) = \tilde{\eta}_{\tilde{\omega}}(\tilde{\theta})$. Our "main" parameter will be $\tilde{c}$, so we shall use notation $\tilde{\eta}_{\tilde{c}} = \tilde{\eta}_{\tilde{c}(\tilde{\omega})}$ instead of $\tilde{\eta}_{\tilde{\omega}}$.

For the Lagrangian $L(\theta, \dot{\theta})$ in the original coordinates we define

$$\eta_\omega(\theta) = M^{tr} \tilde{\eta}_{\tilde{\omega}}(M^{-1}\tilde{\theta}), \quad c = M^{tr}\tilde{c}, \quad \omega = M^{-1}\tilde{\omega}. \tag{21}$$

4.2. Definition of c-Lagrangian and c-action. For a given $c \in \mathbb{R}^3$, let $\alpha(c)$ and η_c be as defined above. Define the *c-Lagrangian* as

$$\begin{aligned} L_c(\theta, \dot{\theta}; \varepsilon) &= L(\theta, \dot{\theta}; \varepsilon) - \eta_c(\theta) \cdot \dot{\theta} + \alpha(c), \\ \tilde{L}_c(\tilde{\theta}, \dot{\tilde{\theta}}; \varepsilon) &= \tilde{L}(\tilde{\theta}, \dot{\tilde{\theta}}; \varepsilon) - \tilde{\eta}_c(\tilde{\theta}) \cdot \dot{\tilde{\theta}} + \alpha(\tilde{c}), \end{aligned} \tag{22}$$

and for $\theta_0, \theta_1 \in \mathbb{R}^3$, define the *c-action* as

$$A_c^\tau(\theta_0, \theta_1) = \inf_\gamma \int_0^\tau L_c(\dot{\gamma}(s), \gamma(s); \varepsilon) ds, \ \tilde{A}_c^\tau(\tilde{\theta}_0, \tilde{\theta}_1) = \inf_\gamma \int_0^\tau \tilde{L}_c(\dot{\gamma}(s), \gamma(s); \varepsilon) ds, \tag{23}$$

where the infimum is taken over all absolutely continuous curves $\gamma : [0, t] \to \mathbb{R}^3$ with $\gamma(0) = \theta_0$ and $\gamma(\tau) = \theta_1$. By Remark 1 and definitions (21), $\tilde{L}_c(\tilde{\theta}, \dot{\tilde{\theta}}; \varepsilon) = L_c(\theta, \dot{\theta}; \varepsilon)$ and $\tilde{A}_c^\tau(\tilde{\theta}_0, \tilde{\theta}_1) = A_c^\tau(\theta_0, \theta_1)$.

Notice that in coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$ the Euler-Lagrange dynamics of the integrable Lagrangian, as well as its modified counterpart $\tilde{L}_c^{int}$, splits into the direct product of $\tilde{x}\tilde{y}$-dynamics and $\tilde{z}$-dynamics. This makes it more convenient to work in coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$.

From now, till the end of this section we work in coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$, for simplicity omitting the tilde in the notations.

4.3. Regularity of $\omega(c)$, u_c and η_c in c.

Lemma 1. *Let $H(\theta, I)$ be a C^∞-smooth Hamiltonian as in (15), such that the corresponding H' is convex with respect to the action variables, completely integrable and C^{r+1}-close to $\frac{I_x^2+I_y^2}{2}$. Given a rotation vector $\omega \in \mathbb{R}^3$, define $c = c(\omega)$, u_c and η_c as in Sect. 4.1. Then*

1. *c is smooth in ω;*
2. *$u_c(\theta)$ and $\eta_c(\theta)$ are smooth in c.*

Proof. Since H' (resp., L') is completely integrable, every orbit lies on one of the invariant tori $\mathcal{T}'_\omega$ (resp. $\mathcal{T}_\omega$). Thus, every orbit is either periodic or quasi-periodic and fills one of smooth 2-tori $\mathcal{T}'_\omega$ (resp. $\mathcal{T}_\omega$). Notice that for each irrational ω, i.e. ω viewed as a vector in $\mathbb{R}^2$ with incommensurable components, there is a unique invariant probability measure, denoted ν_ω, supported on $\mathcal{T}_\omega$. This is the only ergodic invariant measure of rotation number ω. Indeed, rotation vector is independent of coordinate system and in action-angle coordinates of a convex Hamiltonian clearly there is only one 2-torus of rotation vector ω.

The family of measures $\{\nu_\omega\}_\omega$ with irrational ω can be extended to all ω. By the implicit function theorem, the tori $\mathcal{T}_\omega$ depend on ω smoothly, therefore, $\int L\, d\nu_\omega$ depends on ω smoothly. Due to convexity of the β-function we have

$$\beta(\omega) = \int L\, d\nu_\omega.$$

Moreover, $\beta(\omega)$ is differentiable in ω, because of smooth dependence of ν_ω on ω. It follows from the definition of the α-function that $\alpha(c_{xy}) = \langle c_{xy}, \omega\rangle - \beta(\omega)$ for ω such that $\beta'(\omega) = c$. This implies that the dependence $c \mapsto c(\omega)$ is smooth and invertible with smooth inverse.

Invariant tori $\{\mathcal{T}_\omega\}$ are smooth in ω and, therefore, in c. The function u_c and the one-form η_c describe $\mathcal{T}_\omega$ and $\mathcal{T}'_\omega$ with $c = c(\omega)$. Thus, u_c and η_c are smooth in c too. □

4.4. Useful facts about the lenses. System (13) without the deformation term $\beta(\theta, I, \varepsilon)$ is completely integrable, hence no diffusion would be possible. The support of $\beta(\theta, I, \varepsilon)$ is contained in the union of sets $\mathcal{L}_{\mathbf{n}}$ called *lenses*. Their detailed definition appears in Sect. 7.3. A lens $\mathcal{L}_{\mathbf{n}}$ is a round ball in the phase variables θ of a certain (small) radius r, centered a point $\mathbf{n}$. Both radii and positions of the lenses depend on I. Fix a pair of integers a and a_1. For each I we have two families of lenses (Fig. 4).

- Family $\mathcal{F}1$ consists of lenses of radius $r \le \sqrt{\varepsilon}$ centered at the points of the lattice $\Gamma = (2, 2, \frac{2}{a}) \cdot \mathbb{Z}^3$.
- Family $\mathcal{F}2$ consists of lenses of radius $r \le \frac{\sqrt{\varepsilon}}{a_1}$ centered at the points of the lattice $\Gamma_1 = (\frac{2}{a_1}, \frac{2}{a_1}, \frac{2}{a}) \cdot \mathbb{Z}^3$ where $a_1 > a$.

We shall often use the following notation. Given a vector v and a lens $\mathcal{L}_{\mathbf{n}}$, consider a *section* $S(\mathbf{n}, v)$:

$$S(\mathbf{n}, v) = \{\theta \in \mathcal{L}_j : (\theta \mod 2) \cdot v = 0\}, \tag{24}$$

which is a 2-dimensional disk concentric with $\mathcal{L}_j$ of the same radius as $\mathcal{L}_j$.

Lemma 2. *Let $H^{\text{int}}(\theta, I, \varepsilon)$ be as in (15).*

1. *Suppose that $\mathcal{L}_0$ and $\mathcal{L}_1$ are two lenses from the family $\mathcal{F}1$ above, centered at the points $\mathbf{n}_0 = (0, 0, 0)$ and $\mathbf{n}_1 = 2(m, n, \frac{1}{a})$, respectively, with $m, n \in \mathbb{Z}$ (it is important that the third coordinates differ by $\frac{2}{a}$, which is one step of the corresponding lattice). By Tonelli's Theorem, for any $\theta_0 \in \mathcal{L}_0$, $\theta_1 \in \mathcal{L}_1$ there is a unique minimizer of $A^\tau(\theta_0, \theta_1)$. It satisfies the following.*
   a) *If τ is sufficiently large, the minimizer of $A^\tau(\theta_0, \theta_1)$ does not intersect any lenses except for $\mathcal{L}_0$ and $\mathcal{L}_1$.*

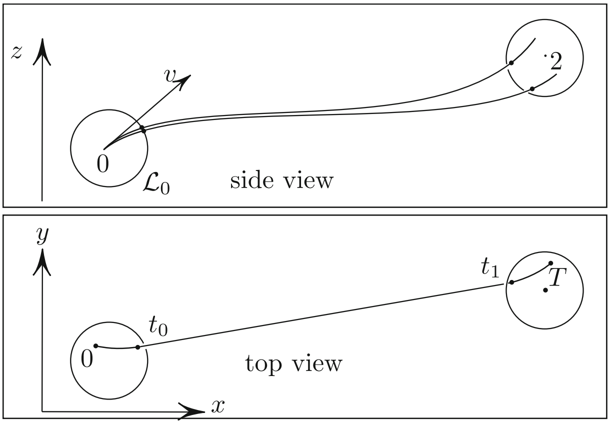

Fig. 4. Projection of minimizing orbits

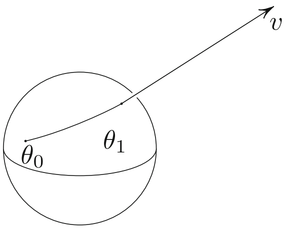

Fig. 5. Initial point θ_0 and exit velocity v define the exit point θ_1

b) Set $\Delta \mathbf{n} = (\mathbf{n}_1 - \mathbf{n}_0)$. Fix any vector v in the ε-cone around $\Delta \mathbf{n}$. If $\theta_0 \in \partial S(\mathbf{n}_0, v)$ (see (24)), then the minimizer of $A^\tau(\theta_0, \theta_1)$ does not intersect $\mathcal{L}_0$.

2. Analogous statement holds for two lenses from $\mathcal{F}1$ centered at points $\mathbf{n}_0 = (0, 0, 0)$ and $\mathbf{n}_1 = (\frac{2m}{a_1}, \frac{2n}{a_1}, \frac{2}{a})$, respectively, with $m, n \in \mathbb{Z}$.

The proof is close to that of Lemma 2 in [KL1].

On the picture t_0 and t_1 denote times of crossing boundaries of the lenses $\mathcal{L}_0$ and $\mathcal{L}_1$ respectively. It is based on the following lemma (Fig. 5):

Lemma 3. *Outside the lenses $\mathcal{L}_\mathbf{n}$, the flow is completely integrable. Inside any lens $\mathcal{L}_0$ we have the following:*

A) For any $\theta_0 \in Int\, \mathcal{L}_0$, and $v \in \mathbb{R}^3 \setminus \{0\}$ there exists a unique trajectory starting at θ_0 and exiting $\mathcal{L}_0$ with velocity v. The exit point θ_1 depends smoothly on θ_0 and v, with $\frac{\partial \theta_1}{\partial \theta_0}$, $\frac{\partial \theta_1}{\partial v}$ bounded by an ε-independent constant;

B) For a trajectory in $\mathcal{L}_0$ we have: $\ddot{z} = O(\varepsilon)$ and deviation of (x, y)-component from the integrable system is $O(\varepsilon^r)$.

The proof is as in [KL1].

4.5. Evaluation of action for the integrable system. We formulate and prove the following lemma for the integrable Lagrangian (18), corresponding to the coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$.

In view of Remark 1 and relations (21), the same statement holds for the original Lagrangian (20). We omit the tilde in the notations.

Let $\mathbf{n}_1$ and $\mathbf{n}_0$ be centers of two lenses, $\mathcal{L}_0$ and $\mathcal{L}_1$, from one of the families $\mathcal{F}1$ or $\mathcal{F}2$ above. Suppose that $|\mathbf{n}_1 - \mathbf{n}_0|_z = \frac{2}{a}$. There is an orbit connecting $\mathbf{n}_0$ with $\mathbf{n}_1$ and belonging to the energy surface $\{H = 1/2\}$. Denote by γ^* this orbit. We know that it is c-static for some $c = c^*$. Let T^* be the time a minimizer on the energy surface $\{H = 1/2\}$ takes to connect $\mathbf{n}_0$ and $\mathbf{n}_1$. Let ω^* be the rotation vector of this minimizer, and let $u_{c^*}(\theta)$ be the smooth function on $\mathbb{T}^3$ graph of whose gradient gives an invariant torus consisting of c^*-minimizers (see Sect. 4.1).

Lemma 4. *There exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ the following holds. Let $\mathbf{n}_0$, $\mathbf{n}_1$, γ^*, c^* T^*, ω^* and $u_{c^*}(\theta)$ be as above, and let $K > 0$ be a fixed constant. Suppose that T^* is sufficiently large.*

Pick two points $\theta_i \in \mathcal{L}_i$, $i = 0, 1$, and let τ be such that $|\tau - T^*| < K$. Let $\gamma : [0, \tau] \to \mathbb{R}^3$ be the minimal orbit connecting θ_0 with θ_1 in time τ. This orbit is c-static for some c. Denote by ω its rotation vector. Then we have:

$$\left|\omega - \omega^*\right| \leq \frac{2K}{T^*}, \tag{25}$$

$$|c - c^*| < \frac{k_1}{T^*}, \tag{26}$$

$$\left|\int_0^\tau L(\gamma(t), \dot{\gamma}(t))dt - u_{c^*}(\theta_1) + u_{c^*}(\theta_0) + \alpha(c^*)\tau - c^*(\theta_0 - \theta_1)\right| \leq \frac{k}{T^*} \tag{27}$$

for some positive constants k_1, k depending on K and ε.

Proof. Let γ be the minimizer connecting θ_0 with θ_1 in time τ. Since the minimizer of the direct product system is the direct product of minimizers of components, one component is the minimizer of the pendulum, the other (2-dimensional) is the minimizer of the completely integrable xy-component. Now we have two trajectories of our system, $\gamma(t)$ and $\gamma^*(t)$, such that $\gamma(0) = \theta_0$, $\gamma(\tau) = \theta_1$, $\gamma^*(0) = \mathbf{n}_0$, $\gamma^*(T^*) = \mathbf{n}_1$. Let us compare the corresponding rotation vectors ω and $\omega^* \in \mathbb{R}^3$. Recall that our system is completely integrable. More exactly, the xy-part of our system can be brought to the form $\bar{H}(I_x, I_y)$ by a smooth symplectic coordinate change. This Hamiltonian is close to $\bar{I}_x^2/2 + \bar{I}_y^2/2$ and the coordinate change is close to identity and preserves integer points. By Bernard's theorem [Be2], a symplectic coordinate change leaves invariant the sets of Aubry, Mañé, and Mather ($\mathcal{A}_c$, $\mathcal{N}_c$ and $\mathcal{M}_c$). This means that the images of the above two curves, $\bar{\gamma}(t)$ and $\bar{\gamma}^*(t)$, are c- and c^*-minimizers, respectively. The curves $\bar{\gamma}(t)$ and $\bar{\gamma}^*(t)$ are trajectories of the system in the new coordinates, and satisfy $\bar{\gamma}(0) = \bar{\theta}_0$, $\bar{\gamma}(\tau) = \bar{\theta}_1$, $\bar{\gamma}^*(0) = \mathbf{n}_0$, $\bar{\gamma}^*(T^*) = \mathbf{n}_1$. In the new variables all the trajectories of the system satisfy $\dot{\bar{x}} = const.$, $\dot{\bar{y}} = const.$ outside the lenses. In particular, outside the lenses $\bar{\gamma}_{xy}(t)$ gets the form $(\dot{\bar{x}}, \dot{\bar{y}}) = \omega_{xy}$, and $\bar{\gamma}^*(t)$ gets the form $(\dot{\bar{x}}, \dot{\bar{y}}) = \omega^*_{xy}$. We have:

$$\omega_{xy} = \frac{(\mathbf{n}_1 - \mathbf{n}_0)_{xy} + 3\text{``}\sqrt{\varepsilon}\text{''}\vec{e}}{\tau}, \quad \omega^*_{xy} = \frac{(\mathbf{n}_1 - \mathbf{n}_0)_{xy}}{T^*},$$

where $\vec{e} = (1, 1)$. This implies

$$\left|\omega_{xy} - \omega^*_{xy}\right| \leq \frac{2K}{T^*}.$$

Now, (26) follows from Lemma 1.

By Fathi's formula [Fa1], we have

$$\int_0^\tau L(\gamma(t))dt = u_c(\theta_1) - u_c(\theta_0) - \alpha(c)\tau + c(\theta_1 - \theta_0).$$

Recall that the α-function of the direct product is the sum of α-functions of components. For the xy-part we have:

$$\alpha_{xy}(c_{xy}) - \alpha_{xy}(c^*_{xy}) = \alpha'_{xy}(c^*_{xy})(c_{xy} - c^*_{xy}) + R(c_{xy})\,(c_{xy} - c^*_{xy})^2. \tag{28}$$

By the properties of α-functions we have $\alpha'_{xy}(c^*_{xy}) = \omega^*_{xy}$. Condition (16) implies that in the case of single resonance, both ω^*_x and ω^*_y are bounded away from zero together with their derivatives. This implies that

$$|R(c_{xy})| \le C_0, \tag{29}$$

where C_0 is a constant independent of c.

For the z-component, we have: $|c_z|,\ |c^*_z| \le C_1 e^{-C_2 T^*}$ for some positive constants C_1, C_2. Indeed, let $z(t)$ be the z-component of the minimizer $\gamma^*(t)$. The energy h_z is conserved, and $z(t)$ is confined to a fixed neighborhood of the saddle for the duration T^*. By the asymptotic formula (68),

$$0 < h_z = O\left(\exp\left(-\frac{\sqrt{\pi\varepsilon}\,T^*}{2}\right)\right).$$

The same argument holds for the z-component of γ. Further notice that $\alpha'(c^*)T^* = \omega^* T^* = \mathbf{n}_1 - \mathbf{n}_0$. Now we can rewrite the last two terms in the form:

$$\begin{aligned}
\alpha(c)\tau - c(\theta_1 - \theta_0) - \alpha(c^*)\tau + c^*(\theta_1 - \theta_0) &= (\alpha(c) - \alpha(c^*))\tau - (c - c^*)(\theta_1 - \theta_0)\\
&= \alpha'(c^*)T^*(c-c^*) + \alpha'(c^*)(\tau - T^*)(c-c^*) - (\theta_1-\theta_0)(c - c^*) + \text{“}C_0\text{”}\,(c - c^*)^2\tau\\
&= [(\mathbf{n}_1 - \mathbf{n}_0) - (\theta_1 - \theta_0)](c - c^*) + \alpha'(c^*)K(c - c^*) + \text{“}C_0\text{”}\,(c - c^*)^2\tau\\
&= (K\alpha'(c^*) + \text{“}4\sqrt{\varepsilon}\text{”})(c - c^*) + \text{“}C_0\text{”}\,(c - c^*)^2\tau.
\end{aligned}$$

Recall that, by Lemma 1, for all θ we have: $|u_c(\theta) - u_{c^*}(\theta)| \le k_0|c - c^*|$ for some constant k_0. Combined with (26), this completes the proof. □

4.6. Minimum in θ is interior. The next lemma is proved for the Lagrangian (17), (18) (omitting the tilde). Due to Remark 1 the same statement holds for the Lagrangian (19), (20) (in original coordinates). Fix three lenses $\mathcal{L}_-$, $\mathcal{L}$ and $\mathcal{L}_+$ centered at points $\mathbf{n}_-$, $\mathbf{n}$ and $\mathbf{n}_+$, respectively. Consider the trajectory $\gamma^-(t)$ that connects $\mathbf{n}_-$ to $\mathbf{n}$ on the energy surface $\{H = 1/2\}$, and let T_- be the corresponding time. This trajectory is c-minimal for some $c = c_-$. Let $\gamma^+(t)$ be the trajectory connecting $\mathbf{n}$ to $\mathbf{n}_+$ on the energy surface $\{H = 1/2\}$, and define c_+ and T_+ similarly. Let $\Delta\mathbf{n}_+ = \|\mathbf{n}_+ - \mathbf{n}\|$. Define the section $S(\mathbf{n}, \Delta\mathbf{n}_+)$ as in (24). Define the corresponding c-Lagrangians by (22). Fix any $\theta_\pm \in \mathcal{L}_\pm$, τ_- and τ_+, and define $A^{\tau_\pm}_{c_\pm}$ by the formula (23). Consider the following function of $\theta \in S(\mathbf{n}, \Delta\mathbf{n}_+)$:

$$s(\theta) = A^{\tau_-}_{c_-}(\theta_-, \theta) + A^{\tau_+}_{c_+}(\theta, \theta_+).$$

Lemma 5. *Assume the notations above. Given an ε-independent constant K, let $|\tau_\pm - T_\pm| < K$. There exist constants $\varkappa$, e.g. $\varkappa = \varepsilon^{r+3}$ and μ, e.g. $\mu \geq \varepsilon^{-r-3}$, such that if*

$$|c_+ - c_-| < \varkappa, \quad T_- > \mu, \ T_+ > \mu,$$

then $s(\theta)$ attains its minimum with respect to θ in the interior of $\mathcal{L}$, and moreover,

$$\min_{\theta \in S(\mathbf{n}, \Delta\mathbf{n}_+)} s(\theta) + \varepsilon^{r+2}/2 < \min_{\theta \in \partial S(\mathbf{n}, \Delta\mathbf{n}_+)} s(\theta). \tag{30}$$

Proof. For simplicity, remove the deformation term β supported on $\mathcal{L}_\pm$; at the end of the proof we shall show that this does not change the result.

The idea of the proof is the following. Consider the system with the β-term supported on $\mathcal{L}$ removed. The resulting system is integrable. We shall show that the action of the integrable system changes very little (of order $O(\frac{1}{T})$) when θ moves from the center of $\mathcal{L}$ to its boundary. Then we plug in the lens at $\mathcal{L}$. If θ lies on the boundary of S_c, then, by Lemma 2b, the minimizer passing through θ does not intersect $\mathcal{L}$, so its action is the same as the one of the integrable system. The minimal action of the original system between θ_- and $\mathbf{n}$ does not exceed its action over the minimizer of the integrable system, connecting the same points. This proves that the minimum of $s(\theta)$ with respect to θ is attained in the interior of $\mathcal{L}$, since β is of order $-\varepsilon^{r+2}$.

Here are the details. Fix some τ_- and τ_+ satisfying $|\tau_\pm - T_\pm| < K$. Recall that $L_{c_\pm}(\theta, \dot\theta) = L(\theta, \dot\theta) - \eta_{c_\pm}(\theta) \cdot \dot\theta + \alpha(c_\pm)$. By formula (27) we have:

$$\begin{aligned}
\min_{\theta \in S(\mathbf{n}, \Delta\mathbf{n}_+)} s(\theta) &= \min_{\theta \in S(\mathbf{n}, \Delta\mathbf{n}_+)} \int_0^{\tau_-} L_{c_-}(\gamma_-(t), \dot\gamma_-(t))dt + \int_0^{\tau_+} L_{c_+}(\gamma_+(t), \dot\gamma_+(t))dt \\
&= u_{c_-}(\theta) - u_{c_-}(\theta_-) - \alpha(c_-)\tau_- - c_-(\theta - \theta_-) + \frac{\text{“}k\text{”}}{T_-} \\
&\quad + u_{c_+}(\theta_+) - u_{c_+}(\theta) - \alpha(c_+)\tau_+ - c_+(\theta_+ - \theta) + \frac{\text{“}k\text{”}}{T_+} - \text{bump}(\theta) \\
&= \big(u_{c_-}(\theta) - u_{c_+}(\theta)\big) + \theta(c_- - c_+) - \text{bump}(\theta) + \frac{\text{“}k\text{”}}{T_+} + \frac{\text{“}k\text{”}}{T_-} \\
&\quad - u_{c_-}(\theta_-) + u_{c_+}(\theta_+) + (\alpha(c_+)\tau_+ - \alpha(c_-)\tau_-) + c_-\theta_- - c_+\theta_+.
\end{aligned}$$

The expression in the last line is independent of θ. Thus, we need to estimate θ-dependence only in the line before. Recall that $u_c(\theta)$ and $\eta_c(\theta)$ are smooth in c. Therefore, if $c_+ - c_-$ is sufficiently small, then each of $(u_{c_-}(\theta) - u_{c_+}(\theta))$, and $|\theta(c_- - c_+)|$ is smaller than $C\varepsilon^{r+3}$. If, at the same time, T_-, T_+ are sufficiently large, i.e. $\geq \varepsilon^{-r-3}$, then and the bump(θ) has depth $\geq \varepsilon^{r+2}$ inside of the lens, then the minimum w.r.t. θ is interior and satisfies (30). □

4.7. Minimum in τ is interior. Consider the Lagrangian defined by (17), (18). Define $A_c^\tau(\theta_0, \theta_1)$ by (23).

Lemma 6. *There exists $\varepsilon_0 > 0$ and a constant $\sigma > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ the following holds. Let $\mathbf{n}_0$ and $\mathbf{n}_1$ be centers of two lenses with $|(\mathbf{n}_1 - \mathbf{n}_0)_z| = \frac{2}{a}$ and $|\mathbf{n}_0 - \mathbf{n}_1|$ large. Consider a trajectory of the integrable system connecting $\mathbf{n}_0$ to $\mathbf{n}_1$ on the energy surface $\{H^{int} = 1/2\}$, and let $T^* \in \mathbb{R}$ be the corresponding time. Suppose that $h_y > 0.01\varepsilon$ (as assumed by (16)).*

Then for any $\theta_i \in \mathcal{L}_i$ ($\mathcal{L}_i$ centered at $\mathbf{n}_i$) the minimum of $A_c^\tau(\theta_0, \theta_1)$ with respect to τ satisfies

$$\min_{|T^*-\tau|\le 5\sqrt{\varepsilon}} A_c^\tau(\theta_0, \theta_1) < \min_{\sigma-1\le|T^*-\tau|\le\sigma} A_c^\tau(\theta_0, \theta_1) - \frac{1}{T^*}, \tag{31}$$

provided that T^* is sufficiently large.

Proof. For each τ there exists a unique minimizer $\gamma(t)$ of A_c^τ, and Lemma 2 implies that γ exits $\mathcal{L}_0$ at some time t_0 and enters $\mathcal{L}_1$ at some time t_1 without meeting other lenses for $t \in (t_0, t_1)$. Outside the lenses the system is completely integrable. Outside the lenses, $A_c^\tau(\theta_0, \theta_1)$ is the direct sum of the xy-part and the z-part. Let us study the xy-component first. To simplify the notation, in this part of the proof let us call (x, y) by θ, and the Lagrangian $L'(x, y, \dot{x}, \dot{y})$ (see formula (18)) by L.

Introduce a symplectic change of coordinates $\Phi : (\theta, I) \mapsto (\phi, J)$ bringing our system to the action-angle variables. Since our system is assumed to be integrable such Φ exists. The motion in the new variables is $J = const.$, $\dot{\phi} = \omega(J)$, and it is governed by the Hamiltonian $\bar{H}(J) = H(\theta(\phi, J), I(\phi, J))$. Moreover, after this change of variable the one-form $\eta_c(\theta) \cdot \dot{\theta}$ in the definition of the c-Lagrangian becomes the constant form, i.e., $c \cdot \dot{\phi}$. It will be more convenient to prove formula (31) in the new variables. To justify this, we show that the action over the minimizer, computed in the new variables, differs from the one computed in the old variables by an error of order $O(\frac{1}{T})$. Indeed, if $L(\theta, \dot{\theta})$ is the Legendre transform of $H(\theta, I)$, then for $\theta = \gamma(t)$, $\dot{\theta} = \dot{\gamma}(t)$ we have:

$$L(\theta, \dot{\theta}) = \dot{\theta} \cdot I - H(\theta, I).$$

Let $\bar{L}(\phi, \dot{\phi})$ be the Legendre transform of $\bar{H}(J)$. Then we have:

$$\bar{L}(\phi, \dot{\phi}) = \bar{L}(\dot{\phi}) = \dot{\phi} \cdot J - \bar{H}(J).$$

By assumption, the change of coordinates was symplectic, which means that

$$I d\theta = J d\phi + dS(\phi, J),$$

where S is the generating function of the coordinate change. By construction in Deformation Lemma 11 for $0 < h_y < \varepsilon^{1/10}$ the Hamiltonian is the sum of a rotator and two pendulums (see (48)). We change to action-angle coordinates of $(x, y, \dot{x}, \dot{y})$. This corresponds to straightening level sets $h_y = \dot{y}^2/2 - \varepsilon \cos^2 \frac{\pi y}{2}$. Such a change of coordinates can be made explicit. It is defined $(y, \dot{y}) \to (\phi_y, J_y)$ with $J_y = \sqrt{\dot{y}^2/2 - \varepsilon \cos^2 \frac{\pi y}{2}}$ and ϕ_y is given by $\phi_y(0, \dot{y}) = 0$ and the differential relation $d\phi_y(y, \dot{y}) = \frac{J_y}{\dot{y}} dy$. In particular,

$$dJ_y = \frac{\dot{y}}{J_y} d\dot{y} + \varepsilon \sin \pi y \, dy, \qquad d\phi_y(y, \dot{y}) = \frac{J_y}{\dot{y}} dy.$$

It follows from the exact form of J_y that for $h_y > 0.01\varepsilon$ the ratio $\frac{J_y}{\dot{y}} < 20$. In other words, $S(\phi, J)$ has a uniformly bounded derivative. For $h_y > \varepsilon^{\frac{1}{10}}$, the change of coordinates is ε-close to identity by the Deformation Lemma 11. Since this lemma will be used for Hamiltonians of type (14) with (large) constants A and a, we choose not to compute this constant explicitly.

For the Lagrangian in terms of the generating function we have

$$L(\theta, \dot\theta) = \bar L(\dot\phi) + \frac{d}{dt} S(\phi, J),$$

where $(\phi, J) = \Phi(\theta, I)$. Then

$$\int_0^\tau L(\theta, \dot\theta) + \eta_c(\theta) \cdot \dot\theta \, dt = \int_0^\tau \bar L(\dot\phi) + c \cdot \dot\phi \, dt + S_0(\phi_0, \phi_1, J),$$

where $S_0(\phi_0, \phi_1, J) = S(\phi_1, J) - S(\phi_0, J)$, and ϕ_0, ϕ_1 are the images of the points θ_0, θ_1 under the coordinate change. Notice that the values of the first integrals are preserved along the minimizer of $\bar A_c^\tau(\phi_0, \phi_1)$, so that J is constant along the flow. Moreover, the dependence between J and the rotation vector ω is smooth. Hence, if $\gamma_\tau(t)$ is the minimizer of $\bar A_c^\tau(\phi_0, \phi_1)$ and $\gamma_T(t)$ is the minimizer of $\bar A_c^T(\phi_0, \phi_1)$ for some $|T - \tau| < \sigma$, if J_τ, J_T are the corresponding values of the action variable, and ω_T, ω_τ — the corresponding rotation vectors, then

$$|S_0(\phi_0, \phi_1, J_\tau) - S_0(\phi_0, \phi_1, J_T)| \le \left| \frac{\partial S_0}{\partial J}(\tilde J) \right| |J_\tau - J_T| \le C_0 |\omega_\tau - \omega_T| \le \frac{C_0 \sigma}{T}, \tag{32}$$

see (25). We use the fact that $|\partial S_0 / \partial J| < C_0$ for some constant C_0 depending on the Hamiltonian.

Now let us prove estimate (31) in the new coordinates. Here the c-Lagrangian gets the form:

$$\bar L(\dot\phi) - c \cdot \dot\phi + \alpha(c) = \bar L(\omega) + \frac{\partial \bar L}{\partial \dot\phi}(\omega)(\dot\phi - \omega) + (\dot\phi - \omega)^{tr} \frac{\partial^2 \bar L}{\partial \dot\phi^2}(\omega)(\dot\phi - \omega)$$
$$+ O((\dot\phi - \omega)^3) - c \cdot \dot\phi + \alpha(c) = \frac{1}{2}(\dot\phi - \omega)^{tr} \frac{\partial^2 \bar L}{\partial \dot\phi^2}(\omega)(\dot\phi - \omega) + O((\dot\phi - \omega)^3).$$

The latter equality holds true due to the following cancelations. As discussed above we have: $\frac{\partial \bar L}{\partial \dot\phi}(\omega) = c$, and (since $\bar L$ is the Legendre transform of $\bar H$), we have

$$\left(\frac{\partial \bar L}{\partial \dot\phi} \cdot \dot\phi - \bar L(\dot\phi) \right) |_{\dot\phi = \omega} = \frac{\partial \bar L}{\partial \dot\phi}(\omega)\omega - \bar L(\omega) = \bar H(c) = \alpha(c).$$

Hence, in the new variables the action takes the form

$$\bar A_c^\tau(\theta_0, \theta_1) = \frac{1}{2} \int_0^\tau (\dot\phi - \omega)^{tr} \frac{\partial^2 \bar L}{\partial \dot\phi^2}(\omega)(\dot\phi - \omega) + O((\dot\phi - \omega)^3) \, dt. \tag{33}$$

The rest of the proof is close to that of Lemma 2 in [KL1]. Since the minimizer does not meet any lenses between $\mathcal{L}_0$ and $\mathcal{L}_1$, we have

$$\dot\phi = \text{const.} \quad \text{for} \ \ t \in (t_0, t_1),$$

and, therefore,

$$\phi \, |_{t_0}^{t_1} = \dot\phi(\tau + 5\text{``}\sqrt{\varepsilon}\text{''} \vec e) = \mathbf{n}_1 - \mathbf{n}_0 + 3\text{``}\sqrt{\varepsilon}\text{''},$$

where $\mathbf{n}_i$ denotes the ϕ-component of $\mathbf{n}_i$, and $\vec{e} = (1, 1)$. Thus we obtain a horizontal velocity estimate:

$$\dot{\phi} - \omega = \frac{\mathbf{n}_1 - \mathbf{n}_0 + 3``\sqrt{\varepsilon}"\,\vec{e}}{\tau + 5``\sqrt{\varepsilon}"} - \omega = \frac{(T-\tau)\omega + 9``\sqrt{\varepsilon}"\,\vec{e}}{\tau + 5``\sqrt{\varepsilon}"}. \tag{34}$$

Recall that $(D^2\bar{L} - I)$ by construction is uniformly bounded (here I stands for the identity matrix). Therefore, there is a constant C_1 such that for any vector v with $1 \le |v| \le 100$ we have:

$$\frac{1}{C_1}|v|^2 \le |v^{tr}(D^2\bar{L})v| \le C_1|v|^2. \tag{35}$$

To verify estimate (31), we shall take τ and τ' such that $|T-\tau| < 5\sqrt{\varepsilon}$ (Case 1 below) and $|T-\tau'| > \sigma_-$ (Case 2 below), and compare the actions for τ and τ' component-wise.

1. First let $|T-\tau| = \kappa\sqrt{\varepsilon}$ for some $\kappa < 5$. By (34), then

$$|\dot{\phi} - \omega| = \frac{\sqrt{\varepsilon}}{T}(``9.5"\,\vec{e} + \kappa\omega).$$

   Substituting the latter estimate into the ϕ-part of the action, we get

$$\begin{aligned} &1/2\int_{t_0}^{t_1} (\dot{\phi}-\omega)^{tr}(D^2\bar{L}(\omega))(\dot{\phi}-\omega)dt \\ &\le T\frac{\varepsilon}{2T^2}\left|(``9.5"\,\vec{e}+\kappa\omega)^{tr}(D^2\bar{L}(\xi))(``9.5"\,\vec{e}+\kappa\omega)\right|. \end{aligned}$$

   If $|``9.5"\,\vec{e}+\kappa\omega| < 1$, then the integral above is less than $C_1\varepsilon/T$. Otherwise, we can use estimate (35) and compute this integral to be less than $120\frac{C_1\varepsilon}{T}$.
2. Now take $\sigma_- < |T-\tau| \le \sigma_+$. Then the integral

$$(33) \ge \frac{1}{2C_1}\frac{(\sigma-1)^2}{2T} = \frac{(\sigma-1)^2}{4C_1T}.$$

   Choosing σ so that

$$\frac{(\sigma-1)^2}{4C_1} > 2(C_0\sigma + 120\varepsilon C_1) + 1$$

   we obtain that the action in Case 2 dominates the sum of the action in Case 1 above and the error term (32). This gives the desired relation (31), modulo the following estimates.

The error term in (33) for both cases is small for large T:

$$\int_{t_0}^{t_1} O((\dot{\phi}-\omega)^3)\,dt = O(\frac{1}{T^2}).$$

The z-components of actions for both τ and τ' are exponentially (in T) close to a constant. Indeed, let $z(t)$ be the z-component of the minimizer. For $t \in (t_0, t_1)$ the energy h_z

is conserved, and $z(t)$ is confined to a fixed neighborhood of the saddle for the duration $t_1 - t_0 = O(T)$. By the asymptotic formula (68),

$$0 < h_z = O\left(\exp\left(-\frac{\sqrt{\pi\varepsilon}T}{2}\right)\right).$$

Finally, we estimate the impact of the lenses. Let $\gamma_\tau(t)$ be the minimizer of $A_c^\tau(\theta_0, \theta_1)$, and $\gamma_{\tau'}(t)$ be that of $A_c^{\tau'}(\theta_0, \theta_1)$. Let v be the exit velocity from $\mathcal{L}_0$ of $\gamma_\tau(t)$, and v' be the exit velocity from $\mathcal{L}_0$ for $\gamma_{\tau'}(t)$. Then we have

$$|v - v'| \leq \text{ const.}\frac{\sqrt{\varepsilon}}{T},$$

as we have shown above. Then, by Lemma 3 A), we have:

$$|\gamma_\tau - \gamma_{\tau'}|_{C^1[0,t_0]} \leq \text{ const.}\frac{\sqrt{\varepsilon}}{T},$$

hence the difference between A_c^T and A_c^τ inside $\mathcal{L}_0$ is bounded by const.$t_0\frac{\varepsilon}{T^2} \leq$ const.$\frac{\varepsilon^{3/2}}{T^2}$. So, if we choose T (and $\mathbf{n}_1 - \mathbf{n}_0$ accordingly) sufficiently large, then the minimum of the action is attained for some τ strictly inside the interval $|T - \tau| \leq \sigma$ and, moreover, (31) holds. □

5. Diffusion in the Double Resonance Case

In order to define a class of Hamiltonians describing dynamics near double resonances we fix an integer matrix $M \in GL_3(\mathbb{Z})$ and a pair of small positive numbers $\varepsilon, \varepsilon' > 0$. Let a and A be positive integers. Denote by

$$H^{\text{int}}(I, \theta; \varepsilon, \varepsilon') = A\frac{\langle (M^*I), (M^*I)\rangle}{2} - \varepsilon\cos^2\frac{\pi(M\theta)_y}{2} - \varepsilon'\cos^2\frac{\pi a}{2}(M\theta)_z, \qquad (36)$$

where $(v)_y$ denotes the y-component of a vector v, with the same notation for z. Consider the model Hamiltonian:

$$H(\theta, I, \varepsilon, \varepsilon') = H^{\text{int}}(\theta, I, \varepsilon, \varepsilon') - \varepsilon^{r+1}\beta(\theta, I, \varepsilon, \varepsilon').$$

where $\beta(\theta, I, \varepsilon, \varepsilon')$ is a smooth deformation which is described later in this section. In this case the coordinate change $(\theta, I) \mapsto (\tilde{\theta}, \tilde{I}) = (M\theta, M^*I)$ is symplectic. In terms of $(\tilde{\theta}, \tilde{I})$, we have:

$$H^{\text{int}}(I, \theta; \varepsilon, \varepsilon') = \tilde{H}^{\text{int}}(\tilde{I}, \tilde{\theta}; \varepsilon, \varepsilon') = A\frac{\langle \tilde{I}, \tilde{I}\rangle}{2} - \varepsilon\cos^2\frac{\pi\tilde{y}}{2} - \varepsilon'\cos^2\frac{\pi a}{2}\tilde{z}, \qquad (37)$$

and the corresponding Lagrangian is

$$\tilde{L}^{\text{int}}(\tilde{\theta}, \dot{\tilde{\theta}}; \varepsilon, \varepsilon') = A\frac{\langle \dot{\tilde{\theta}}, \dot{\tilde{\theta}}\rangle}{2} + \varepsilon\cos^2\frac{\pi\tilde{y}}{2} + \varepsilon'\cos^2\frac{\pi a}{2}\tilde{z}. \qquad (38)$$

By Remark 1, the Lagrangian corresponding to $H^{\text{int}}(I, \theta; \varepsilon, \varepsilon')$ is $L^{\text{int}}(\theta, \dot{\theta}; \varepsilon, \varepsilon')$ equal to

$$\tilde{L}^{\text{int}}(\tilde{\theta}, \dot{\tilde{\theta}}; \varepsilon, \varepsilon') = A\frac{\langle M\dot{\theta}, M\dot{\theta}\rangle}{2} + \varepsilon\cos^2\frac{\pi(M\theta)_y}{2} + \varepsilon'\cos^2\frac{\pi a}{2}(M\theta)_z. \qquad (39)$$

The Lagrangian $\tilde{L}^{\text{int}}$ is the direct sum of two pendula and a rotation. Therefore, it is convenient to work in coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$. In this section we assume that "the energy of both the $\tilde{z}$ and $\tilde{y}$-component is small": namely,

$$|h_{\tilde{y}}| \leq \frac{\varepsilon}{10A}, \tag{40}$$

the energy $h_{\tilde{z}}$ being very small (prescribed by the construction), and $\tilde{I}_x$ such that the total energy is close to $1/2$. (Compare this condition with its "complement", condition (16)). Let us assume $A = 1$ here.

5.1. Definition of u_c, η_c, c-Lagrangian and c-action. As in Sect. 5, we modify the Lagrangian (39) by a closed one form $\eta_c(\theta)$. We shall define $\tilde{\eta}_{\tilde{c}}(\tilde{\theta})$ for the Lagrangian $\tilde{L}^{int}$ as follows

$$c = M^{tr}\tilde{c}, \quad \eta_c(\theta) = M^{tr}\tilde{\eta}_{\tilde{c}}(M\theta).$$

With these definitions we shall have, as in the single resonance case, $\tilde{L}_c = L_c$ and $\tilde{A}_c = A_c$. Therefore,

in this section, we work in coordinates $(\tilde{\theta}, \dot{\tilde{\theta}})$, for simplicity, omitting the tilde.

For this Lagrangian one can write explicit formulas for η_c and u_c. We define these functions component-wise. Let $\eta_{c_x} = c_x$, and $u_{c_x} = xc_x$. Let $\eta_{c_y}(y)$, $u_{c_y}(y)$ be defined as in Example 2, Sect. 3.3. The same definition for z-component, though for z we only need to consider the case $h_z \geq 0$.

Given a vector $c = (c_x, c_y, c_z)$ with $c_x \sim 1$ and $0 < c_y \leq \sqrt{2\varepsilon(1+\alpha)}$ and $c_z \geq 0$ define a closed one-form

$$\eta_c(\theta) = \begin{cases} (c_x, u'_{c_y}(y) + c_y, u'_{c_z}(z) + c_z) & \text{for} \quad c_y \geq c_y^+ \\ (c_x, u'_{c_y^+}(y) + c_y, u'_{c_z}(z) + c_z) & \text{for} \quad 0 \leq c_y \leq c_y^+. \end{cases} \tag{41}$$

Its cohomology class is c. For $c_y < 0$ we use the symmetry, and define $u_{c_y}(y) = -u_{-c_y}(y)$, and then η_c is again defined by (41). For $c_y = 0$ we have a freedom of picking either the top or the bottom separatrix. This freedom can be used to find a closed one-form of cohomology class (c_x, c_y, c_z) for $|c_y| < c_y^+$. We, however, design η_c's to approximate the action and satisfy formula (8).

Denote branches of the separatrices by $u_0^{\pm}(y) = \pm u_{c_y^+}(y)$ and let

$$\eta^{\pm}_{(c_x,0,c_z)}(\theta) = (c_x, \nabla u_0^{\pm}(y), c_z).$$

We have

$$\alpha(c) = \alpha(c_x, c_y, c_z) = \frac{c_x^2}{2} + h_y(c_y) + h_z(c_z),$$

where $h_y(c_y)$ and $h_z(c_z)$ are functions given by Example 2 (see (11) relating c's and h's). We also set $h_y(0) = h_z(0) = 0$.

5.2. Definition of c-Lagrangian and c-action. For $|c_y| \ge c_y^+$ define the c-Lagrangian

$$L_c(\theta, \dot\theta) = L(\theta, \dot\theta) - \eta_c(\theta) \cdot \dot\theta + \alpha(c) \tag{42}$$

and the c-action

$$A_c^\tau(\theta_0, \theta_1) = \inf_\gamma \int_0^\tau L_c(\theta, \dot\theta; \varepsilon) dt = A_0^\tau(\theta_0, \theta_1) + \int_0^\tau \eta_c(\theta)\dot\theta \, dt + \alpha(c)\tau,$$

where the infimum is taken over all C^1-curves $\gamma : [0, \tau] \to \mathbb{R}^3$ with $\gamma(0) = \theta_0$ and $\gamma(\tau) = \theta_1$. Notice that the form $\eta(\theta) \cdot \dot\theta$ is closed, and therefore, $\int_0^\tau \eta(\theta) \cdot \dot\theta \, dt$ is independent of γ. Hence, the action $A_c^\tau(\theta_0, \theta_1)$ has the same minimizers as $A_0^\tau(\theta_0, \theta_1)$ (i.e., the Euler-Lagrange flows of L_c and L are the same). We can rewrite the c-Lagrangian in the form

$$L_c(\theta, \dot\theta) = \frac{\langle \dot\theta - \eta_c(\theta), \dot\theta - \eta_c(\theta)\rangle}{2}.$$

5.3. Useful facts about the flow inside the lenses. System (37) without the β-term is completely integrable, hence no diffusion would be possible. The support of β is contained in the union of sets $\mathcal{L}_\mathbf{n}$ called lenses (see the detailed definition in Sect. 7.3). For this section it is enough to see them as round balls $\mathcal{L}_\mathbf{n}$ in the phase variables θ of radius $r = \sqrt{\varepsilon'}$, centered at points $\mathbf{n}$ of the lattice $(2, 2, \frac{2}{a}) \cdot \mathbb{Z}^3$. Define $S(\mathbf{n}, v)$ as in (24).

Lemma 7. *Analog of Lemma 2 holds.*

As discussed above, the same statement holds for the Lagrangian (39).

5.4. Evaluation of the action for integrable system. Recall that $a \in \mathbb{Z} \setminus \{0\}$ is a non-zero integer. Let $\mathbf{n}_1, \ \mathbf{n}_0 \in 2\mathbb{Z} \times 2\mathbb{Z} \times \frac{2}{a}\mathbb{Z}$ be centers of two lenses, $\mathcal{L}_0$ and $\mathcal{L}_1$, and denote $\Delta\mathbf{n} = \mathbf{n}_1 - \mathbf{n}_0 = (\mathbf{n}_x, \mathbf{n}_y, 2/a)$. There is an orbit of the integrable system connecting $\mathbf{n}_0$ with $\mathbf{n}_1$ and belonging to the energy surface $\{H = 1/2\}$. Denote this orbit by γ^*. Let T^* be the time it takes γ^* to connect $\mathbf{n}_0$ and $\mathbf{n}_1$, and let ω^* be the rotation vector of γ^*. We know that γ^* is c-static for some $c^* = c(\gamma^*)$. There are explicit formulas to define $c^* = (c_x^*, c_y^*, c_z^*)$. Namely: $c_x^* = \frac{(\mathbf{n}_1 - \mathbf{n}_0)_x}{T^*}, \quad c_z^* = \int_0^2 \sqrt{2(h_z + \varepsilon' \cos^2 \frac{\pi y}{2})}$, where h_z is the energy of the z-component (cf. (11)). To define c_y^*, consider three cases.

If $\mathbf{n}_y \ge 2$ (in which case $h_y > 0$), define c_y^* by the same formula as c_z^* with h_y instead of h_z.

If $-\mathbf{n}_y \ge 2$, define $c_y^*(\mathbf{n}_y) = -c_y^*(-\mathbf{n}_y)$.

If $\mathbf{n}_y = 0$, define $c_y^* = 0$. Notice that we have a discontinuity in the dependence of c_y^* on $\Delta\mathbf{n}$. This fact produces additional difficulties in having c as a parameter in our constructions. In Sect. 5.6 we explain a way of overcoming this inconvenience. The following lemma is an analog of Lemma 4.

Lemma 8. *There exist positive $\varepsilon_0, \ \varepsilon_0'$ such that for all $0 < \varepsilon < \varepsilon_0, \ 0 < \varepsilon' < \varepsilon_0'$ the following holds. Let $\mathbf{n}_0$, $\mathbf{n}_1$, γ^*, c^* T^*, ω^* and $u_{c^*}(\theta)$ be as above, and let $K > 0$ be a fixed constant. Suppose that T^* is sufficiently large. Pick two points $\theta_i \in \mathcal{L}_i, \ i = 0, 1$, and take any τ such that $|\tau - T^*| < K$. Let $\gamma : [0, \tau] \to \mathbb{R}^3$ be the minimal orbit connecting θ_0 with θ_1 in time τ. This orbit is c-static for some c. Denote by ω its rotation vector.*

1. *In the case* $\mathbf{n}_y = (\mathbf{n}_1 - \mathbf{n}_0)_y \neq 0$ *we have:*

$$\left|\omega - \omega^*\right| \leq \frac{2K}{T^*}, \quad |c - c^*| < \frac{k_1}{T^*},$$
$$\left|\int_0^\tau L(\gamma(t), \dot\gamma(t), \varepsilon, \varepsilon')dt - u_{c_y^*}(y_1) + u_{c_y^*}(y_0) + \alpha(c^*)\tau - c^*(\theta_0 - \theta_1)\right| \leq \frac{k}{T^*} \tag{43}$$

for some positive constants k_1, k *depending on* K *and* ε.

2. *In the case* $\mathbf{n}_y = 0$ *and* $\pm\dot\gamma_y(0) > 0$ *formula (43) holds with the* y*-component of* u_c, $u_{c_y^*}(y_0)$, *replaced by* $u_0^{\pm}(y_0)$, *and* $u_{c_y^*}(y_1)$ *replaced by* $u_0^{\mp}(y_1)$. *Also* $c^* = (c_x^*, c_y^*, c_z^*)$ *is replaced by* $c^* = (c_x^*, 0, c_z^*)$.

Proof. Consider the case $\mathbf{n}_y \neq 0$. First, we show that $|\omega^* - \omega| < \frac{k}{T^*}$ for some constant k. The proof repeats that of Lemma 4. The x-component of the system is a rotator, so the calculation as in Lemma 4 can be made for it.

It is left to show that the difference $|\omega_y - \omega_y^*|$ is of order $\frac{k_0}{T^*}$ for some constant k_0. The idea is the same as in the case of single resonance: if rotation numbers ω_y and ω_y^* of two trajectories differ by more than a certain constant times $\frac{1}{T^*}$, then the endpoints of the two trajectories above would be more than 1 unit apart. Smoothness of $c(\omega)$ with respect to ω implies the second formula.

The proof of estimate (43) follows the same lines as the proof of (27), but requires more care. Since the energy of the y-component is not bounded away from zero in the present case ($0 < h_y < 0.1\varepsilon$), we cannot guarantee that $\alpha(c)$ be smooth in c. We shall prove that the first derivative of this function is bounded, but the second derivative does not have to be bounded. Nevertheless, we shall prove that the product $R_\alpha = |\alpha_y''(\bar c_y)(c_y - c_y^*)^2|$ (cf. formulas (28) and (29)) is bounded by $\frac{C_1}{T^2}$ for some constant C_1 only depending on ε.

Let γ be a minimizer connecting θ_0 with θ_1 in time τ. By Fathi's formula [Fa1], we have

$$\int_0^\tau L(\gamma(t))dt = u_c(\theta_1) - u_c(\theta_0) - \alpha(c)\tau + c(\theta_1 - \theta_0).$$

Recall that an α-function of a direct sum of Lagrangians is the sum of α-functions of components. Thus, we have

$$\alpha_{xy}(c_{xy}) - \alpha_{xy}(c_{xy}^*) = \alpha_{xy}'(c_{xy}^*)(c_{xy} - c_{xy}^*) + \alpha_{xy}''(\bar c)(c_{xy} - c_{xy}^*)^2$$

for some $\bar c_{xy}$ between c_{xy} and c_{xy}^*. For the x-component of this relation, the argument is the same as that the in the single resonance case. For the y-component it is slightly different. Consider the Lagrangian $L(y, \dot y) = \frac{\dot y^2}{2} + \varepsilon\cos^2\frac{\pi y}{2}$. Its dual Hamiltonian is $H(y, I) = \frac{I^2}{2} - \varepsilon\cos^2\frac{\pi y}{2}$. In Example 2 we computed c as a function of energy h of the Hamiltonian H: $c(h) = \int_0^2 \sqrt{2(h + \varepsilon\cos^2\frac{\pi y}{2})}\,dy$. Since $\alpha(c) = H(\theta, \nabla u_c(\theta) + c)$ (this relation is independent of θ on a fixed trajectory), we obtain an implicit function $\alpha(c(h)) = h$. Differentiating this expression with respect to h, we get $\alpha'(c(h))\,c'(h) = 1$. Differentiating again, we get

$$\alpha''(c(h))\,(c'(h))^2 + \alpha'(c(h))\,c''(h) = 0.$$

This gives $\alpha''(c(h)) = -\frac{c''(h)}{(c'(h))^3}$. Calculations from Appendix 9 show that

$$c'(h) \sim \frac{\ln \varepsilon/h}{\sqrt{\varepsilon}}, \quad c''(h) \sim -\frac{c}{h\sqrt{\varepsilon}}$$

(here "$\sim$" means equality modulo some multiplicative constants). Substituting this into the previous relation, we get:

$$\alpha'(c) \sim \frac{\sqrt{\varepsilon}}{\ln \varepsilon/h} \le \sqrt{\varepsilon}, \quad \alpha''(c(h)) \sim \frac{\varepsilon}{h(\ln \varepsilon/h)^3}.$$

Let us use this asymptotic formula to compute the error in determining h_y. Denote by h_y^* the energy of the y-component of γ^*, and by h_y—that of γ. Let $\Delta h_y := h_y - h_y^*$. If the lenses have radius $\sqrt{\varepsilon}$ and the y-component of the speed near the lenses is $\sqrt{\varepsilon}$; we get that $T = \mathbf{n}_y \mathcal{T}(h_y^*) = \mathbf{n}_y \mathcal{T}(h_y) + C_0$ for some C_0 of order of one. We know that $\mathbf{n}_y \sim T/\mathcal{T}(h_y)$. Therefore, $\mathcal{T}(h_y) - \mathcal{T}(h_y^*) \sim \frac{K\mathcal{T}(h_y)}{T}$. Apply the mean value theorem: $\mathcal{T}(h_y) - \mathcal{T}(h_y^*) = \mathcal{T}'(h_y')\Delta h_y$ with $h' \in [h_y, h_y^*]$. Plug in the asymptotic values of $\mathcal{T}(h_y)$ and $\mathcal{T}'(h_y)$ given by (68) and (69) respectively. Then relation between Δh_y and $\mathcal{T}(h_y)$'s quantities becomes

$$\frac{c}{h_y\sqrt{\varepsilon}}\Delta h_y \sim \frac{\ln(\varepsilon/h_y)}{T\sqrt{\varepsilon}} \quad \text{or} \quad \Delta h_y \sim \frac{h_y \ln(\varepsilon/h_y)}{T}.$$

Now compute Δc_y. By the mean value theorem, $\Delta c_y \sim c_y'(h_y)\Delta h_y \sim \dfrac{\Delta h_y \ \ln \varepsilon/h_y}{\sqrt{\varepsilon}}$. Finally,

$$\alpha''(c_y)(\Delta c_y)^2 \sim \frac{\varepsilon}{h_y(\ln \varepsilon/h_y)^3} \frac{h_y^2\,(\ln \varepsilon/h_y)^4}{\varepsilon T^2} = \frac{h_y \ \ln \varepsilon/h_y}{T^2} \le \frac{C_1}{T^2}$$

for some constant C_1.

One can prove that for fixed θ_0 and θ_1, $u_c(\theta_0)$ and $u_c(\theta_1)$ are smooth in c. Indeed, $\frac{\partial u_c(\theta_0)}{\partial c} = \frac{\partial u_c(\theta_0)}{\partial h_y}\frac{\partial h_y}{\partial c}$. By the asymptotic expression (68), $\frac{\partial h_y}{\partial c} \to 0$ when $h_y \to 0$. By the formula for u_c, $\frac{\partial u_c(\theta_0)}{\partial h_y}$ is bounded. With the calculations above, in the same way as in Lemma 4, we get the estimate

$$\begin{aligned} &\alpha(c)\tau - c(\theta_1 - \theta_0) - \alpha(c^*)K\tau + c^*(\theta_1 - \theta_0) \\ &\quad = (\alpha(c) - \alpha(c^*))\tau - (c - c^*)(\theta_1 - \theta_0) \\ &\quad = (K\alpha'(c^*) + \text{“}4\sqrt{\varepsilon}\text{”})(c - c^*) + \frac{\text{“}C_1\text{”}}{T^2}\tau. \end{aligned}$$

This implies the desired estimate in the case $\mathbf{n}_y \neq 0$. The case when $\mathbf{n}_y = 0$ is similar. □

5.5. Minimum in θ is interior. Consider three lenses $\mathcal{L}_-$, $\mathcal{L}$ and $\mathcal{L}_+$ centered at points $\mathbf{n}_-$, $\mathbf{n}$ and $\mathbf{n}_+$, respectively. Consider the trajectory $\gamma^-(t)$ that connects $\mathbf{n}_-$ to $\mathbf{n}$ on the energy surface $\{H = 1/2\}$, and let T_- be the corresponding time. This trajectory is c-minimal for some $c = c_-$. Define η by (41) and the c-Lagrangian by (42). Let $\gamma^+(t)$ be the trajectory connecting $\mathbf{n}$ to $\mathbf{n}_+$ on the energy surface $\{H = 1/2\}$, and define c_+ and T_+ similarly. Given a vector v and a lens $\mathcal{L}_\mathbf{n}$, consider a 2-dimensional disk $S(\mathbf{n}, v)$, defined in (24).

Fix any $\theta_\pm \in \mathcal{L}_\pm$,$\tau_-$ and τ_+, and consider the following function of θ:

$$s(\theta) = A_{c_-}^{\tau_-}(\theta_-, \theta) + A_{c_-}^{\tau_+}(\theta, \theta_+).$$

Lemma 9. *In the above notations, let $\tau_\pm$ satisfy $|\tau_\pm - T_\pm| < 10$. Then there exist constants $\varkappa$, e.g. $\varkappa = \varepsilon^{r+3}$ and μ, e.g. $\mu = \varepsilon^{-r-3}$ such that if*

$$|c_+ - c_-| < \varkappa, \quad T_- > \mu, \; T_+ > \mu,$$

then $s(\theta)$ attains its minimum with respect to θ in the interior of $\mathcal{L}$, and moreover,

$$\min_{\theta \in S(\mathbf{n}, c_+)} s(\theta) + \varepsilon^{r+2}/2 < \min_{\theta \in \partial S(\mathbf{n}, c_+)} s(\theta). \tag{44}$$

Proof. The proof repeats that of Lemma 5. □

5.6. Passing through a double resonance. In the case we need to diffuse across a double resonance in notations of Sect. 5 we have two possibilities: diffuse to annihilate the energy h_y and increase the energy h_z of the z-component (i.e., we turn the corner from the y-resonance to the z-resonance), or change the sign of the y-component of the velocity (i.e., we go through the double resonance along the y-resonant line). Consider the first situation, with the second being similar.

We consider a collection of lenses $\mathbf{n}_0 = 0$ and $\Delta\mathbf{n}_i = (\mathbf{n}_x^i, \mathbf{n}_y^i, 2), i = 1, 2, \dots$ such that $|\Delta\mathbf{n}_i|$ are large enough to satisfy Lemma 8, and the ratios $\mathbf{n}_y^i/\mathbf{n}_x^i$ monotonically decrease to zero until $\mathbf{n}_y^j = 2$ for some j. At this moment we need to deal with the discontinuity of c_y (see (43)). Indeed, as long as $\mathbf{n}_y^j \geq 2$, then $c_y^* \geq c_y^+ = 2\sqrt{2\varepsilon}$. However, to switch $\mathbf{n}_y$ from positive to negative we need to change the sign of non-zero c_y^* to the opposite one. To overcome this difficulty we consider a somewhat artificial procedure. We insert a long collection of steps with the same $\Delta\mathbf{n}_i = (\mathbf{n}_x^j, 2, 2)$ with $j \leq i \leq j'$. Along this collection we select c_i to be decreasing slowly in the y component: $c_i = (1, c_y^i, c_z^+), c_j = c_y^+, \dots, c_y^{j'} = 0$. In order to be able to apply Lemma 8 for such c's we add a constant boundary term:

$$\int_0^{T_i} L(\gamma(t), \dot\gamma(t))dt + (c_y^i - c_y^+)(y_{i+1} - y_i), \tag{45}$$

where $\gamma(0) = \theta_i = (x_i, y_i, z_i)$, $\gamma(T_i) = \theta_{i+1} = (x_{i+1}, y_{i+1}, z_{i+1})$. Notice that the terms we added do not depend on a minimizer γ and do not affect the minimization process. Absorbing this term we can change the cohomology class of the closed one-form η_c defined in (41).[2]

[2] Certainly, the minimization of the sum of two terms like (45) leads to a corner for the minimizing solution due to $c_y^i \neq c_y^{i+1}$. To have a smooth minimizing solution we make a smooth deformation/gluing of closed one forms $\eta_{c_{i-1}}$ and η_{c_i} to match the forms in the lens $\mathcal{L}_i$. Then the minimization provides smooth solutions. See Sect. 8 for details.

5.7. Minimum in τ is interior. The next lemma is formulated for $h_y(c_y) > 0$, i.e., for $c_y \geq c_y^*$. In the case when we change the direction of the trajectory in y, the same argument as above leads to the conclusion that the minimum with respect to T is interior. Recall that $\mathcal{T}(h_y)$ denotes the period of a trajectory of the standard pendulum with energy h_y. The necessary estimates are given in Appendix 9.

Lemma 10. *There exists $\varepsilon_0 > 0$ and $\sigma > 0$ (one can take $\sigma = 10$) such that for all $0 < \varepsilon < \varepsilon_0$ the following holds. Let $\mathbf{n}_0$ and $\mathbf{n}_1$ be centers of two lenses with $\pi_z(\mathbf{n}_1 - \mathbf{n}_0) = 2/a$, $|\mathbf{n}_0 - \mathbf{n}_1|$ large. Consider a trajectory γ^* of the integrable system, connecting $\mathbf{n}_0$ to $\mathbf{n}_1$ on the energy surface $\{H = 1/2\}$, and let $T^* \in \mathbb{R}$ be the corresponding time. This trajectory is a c-minimizer for some c^*. Suppose that $h_y < 0.1\varepsilon$.*

Then for any $\theta_i \in \mathcal{L}_i$ ($\mathcal{L}_i$ centered at $\mathbf{n}_i$) the minimum of $A^\tau(\theta_0, \theta_1)$ with respect to τ satisfies

$$\min_{|T^*-\tau|\leq 5\sqrt{\varepsilon}} A_c^\tau(\theta_0,\theta_1) < \min_{\sigma-1\leq|T^*-\tau|\leq\sigma} A_c^\tau(\theta_0,\theta_1) - \frac{1}{T}. \tag{46}$$

Proof. By Tonelli's Theorem, there exists a minimizer $\gamma(t) = (x(t), y(t), z(t))$, $t \in [0, \tau]$, with $\gamma(0) = \theta_0$ and $\gamma(\tau) = \theta_1$. By Lemma 7, γ exits $\mathcal{L}_0$ at some time $t_0 > 0$, enters $\mathcal{L}_\mathbf{n}$ at time $t_1 < \tau$ without meeting any other lenses for $t \in (t_0, t_1)$. Outside the lenses the system is a direct sum of three 1-d systems. Decompose the action A_c^τ into the following components:

$$\begin{aligned} A_c^\tau(\theta_0,\theta_1) &= \frac{1}{2}\int_{t_0}^{t_1}(\dot{x}-c_x)^2dt + \frac{1}{2}\int_{t_0}^{t_1}(\dot{y}-\nu(y,h_y))^2dt \\ &\quad + \frac{1}{2}\int_{t_0}^{t_1}(\dot{z}-\nu(z,h_z))^2dt + \Delta A \\ &:= A_{c,x}^\tau + A_{c,y}^\tau + A_{c,z}^\tau + \Delta A, \end{aligned}$$

where ΔA is the impact of the lenses.

For the x- and the z-components of the action we have exactly the same estimates as in Lemma 4. We repeat the proof for the x-component here for the sake of completeness. First we estimate the x-part of the action over (t_0, t_1). Since $|\dot{x}| \geq \frac{4}{5}$ on $(0, \tau)$, and the diameter of the lenses is $2\sqrt{\varepsilon}$, we have: $t_0, \tau - t_1 \leq 2.5\sqrt{\varepsilon}$, and hence we have:

$$t_1 - t_0 = \tau - \text{“}5\sqrt{\varepsilon}\text{”},$$

and thus

$$x(t_1) - x(t_0) = \dot{x}(\tau - \text{“}5\sqrt{\varepsilon}\text{”}) = n_x + \text{“}2\sqrt{\varepsilon}\text{”}.$$

Hence we obtain an x-velocity estimate:

$$\dot{x} - c_x = \frac{n_x + \text{“}2\sqrt{\varepsilon}\text{”}}{\tau - \text{“}5\sqrt{\varepsilon}\text{”}} - c_x = \frac{c_x|T-\tau| + \text{“}8\sqrt{\varepsilon}\text{”}}{\tau - \text{“}5\sqrt{\varepsilon}\text{”}}.$$

Let $|T - \tau| \leq 5\sqrt{\varepsilon}$. Substituting the last estimate into the x-part of the action over (t_0, t_1), we get

$$\frac{1}{2}\int_{t_0}^{t_1}(\dot{x}-c)^2dt \leq \frac{(10+5)^2}{2}\frac{\varepsilon}{T} \leq 125\frac{\varepsilon}{T^2}T < 125\frac{\varepsilon}{T}.$$

If $|T-\tau|\geq 9$, then

$$\frac{1}{2}\int_{t_0}^{t_1}(\dot{x}-c)^2dt\geq\frac{10}{T}.$$

Let us estimate the y-component of the action, $A^{\tau}_{c,y}$ outside the lenses. Here the y-component of the solution curve satisfies the system with the Hamiltonian $H(y,I_y)=\frac{1}{2}I_y^2-\varepsilon U_y(y;\varepsilon)$. We show that for $h_y<0.1\varepsilon$ the y-component of the action is smaller than $\frac{\varepsilon}{T}$, which does not affect estimate (46).

By definition of $\nu(y,h_y)$, on the curve $y(t)$ with energy h_y such that $\mathcal{T}(h_y)=\frac{T}{n_y}$ we have: $\dot{y}-\nu(y,h_y)=0$ for all t. To study the $A^{\tau}_{c,y}$, we relate the time τ to the energy of the minimizer $\tilde{y}(t)$ such that $\tilde{y}(0)=\pi_y\theta_0:=y_0$, $\tilde{\theta}(\tau)=\pi_y\theta_1:=y_1$. Outside the lenses it satisfies $\dot{\tilde{y}}^2/2-\varepsilon U_y(\tilde{y};\varepsilon)=\tilde{h}_y$ for some constant $\tilde{h}_y$.

By assumption, for the y-component we have: $\mathcal{T}(h_y)=\frac{T}{n_y+\text{“}\sqrt{\varepsilon}\text{”}}$. Let us estimate the difference $|h_y-\tilde{h}_y|$. Close to even integer points, the velocity is $\geq\sqrt{\varepsilon}/2$; the radius of the lenses is at least $\sqrt{\varepsilon}/2$. Therefore, $n_y\mathcal{T}(\tilde{h}_y)=\tau+\text{“4”}$. Hence,

$$\mathcal{T}(h_y)-\mathcal{T}(\tilde{h}_y)=\frac{T-\tau+\text{“4”}}{n_y}=\frac{\text{“14”}}{n_y}.$$

Then, using the asymptotic formula (69) for the period, for some $\hat{h}_y$ between h_y and $\tilde{h}_y$, we get:

$$|\tilde{h}_y-h_y|=|\mathcal{T}(h_y)-\mathcal{T}(\tilde{h}_y)|\frac{1}{|\mathcal{T}'(\hat{h}_y)|}\leq|\mathcal{T}(h_y)-\mathcal{T}(\tilde{h}_y)|c_2\hat{h}_y\sqrt{\varepsilon}\leq 10c_2\frac{\sqrt{\varepsilon}\hat{h}_y}{n_y}.$$

Now we are ready to estimate the action:

$$\begin{aligned}
&\frac{1}{2}\int_0^{\tau}\left(\dot{\tilde{\theta}}(t)-\nu(\tilde{\theta}(t),h_y)\right)^2dt\\
&=\frac{1}{2}\int_0^{\tau}\left(\sqrt{2(\tilde{h}_y+\varepsilon U_y(\tilde{\theta}(t);\varepsilon))}-\sqrt{2(h_y+\varepsilon U_y(\tilde{\theta}(t);\varepsilon))}\right)^2dt\\
&=\int_0^{\tau}\frac{(\tilde{h}_y-h_y)^2}{\left(\sqrt{2(\tilde{h}_y+\varepsilon U_y(\tilde{\theta}(t);\varepsilon))}+\sqrt{2(h_y+\varepsilon U_y(\tilde{\theta}(t);\varepsilon))}\right)^2}\leq\tau\frac{|\tilde{h}_y-h_y|^2}{2h_y}\\
&\leq\tau(14c_2)^2\frac{\varepsilon h_y}{2n_y^2}\leq 20\pi\frac{\varepsilon(\mathcal{T}(h_y))^2h_y}{T}\leq 20\pi\frac{\varepsilon h_y\ln^2(\frac{4}{\pi^2}\frac{h_y}{\varepsilon})}{2^3\pi\varepsilon}\frac{1}{T}\leq\frac{12\varepsilon}{T}.
\end{aligned}$$

The impact of the lenses, as well as $A^{\tau}_{c,z}$, can be shown to be negligible exactly as in Lemma 6. □

6. Deformation Within Completely Integrable Hamiltonians

Consider a configuration of three resonant planes:

$$\pi_1 = \{I_y = 0\},\ \ \pi_2 = \{I_z = 0\},\ \ \pi_3 = \{I_x = I_y\}.$$

Our aim is to construct a Hamiltonian $H(I, \theta, \varepsilon)$ close to $H_0(I)$ whose actions change along the contour $(\pi_1 \cup \pi_2 \cup \pi_3)$ close to the energy surface $\{H(\cdot, \varepsilon) = 1/2\}$. Recall that $\mathbb{S}^2_\varepsilon$ denotes the projection of this energy surface into the space of actions. In order to use the arguments of Sect. 5, we have to make H have a certain special form in the neighborhoods of the "double resonances", i.e., of $\pi_1 \cap \pi_2 \cap \mathbb{S}^2_\varepsilon$ and $\pi_2 \cap \pi_3 \cap \mathbb{S}^2_\varepsilon$. Moreover, in order to use the arguments of Sect. 4, we have to make $H(\cdot, \varepsilon)$ completely integrable in a neighborhood of $\pi_2 \cap \mathbb{S}^2_\varepsilon$ and convex with respect to the actions. In fact, H will have the special form (15) here. The main result of this section proves that a Hamiltonian H with these properties does exist (Fig. 6).

Lemma 11 (Deformation lemma). *Consider a configuration of three resonant planes*

$$\pi_1 = \{I_y = 0\}, \quad \pi_2 = \{I_z = 0\}, \quad \pi_3 = \{aI_y - bI_x = 0\}, \tag{47}$$

where a and b are integers. Fix $0 < \varepsilon', \varepsilon'', \bar{\varepsilon}$ small and $\varepsilon = \max\{\varepsilon', \varepsilon'', \bar{\varepsilon}\}$. Let V_- be an $\varepsilon^{\frac{1}{10}}$-neighborhood of $p_1 \in \pi_1 \cap \pi_2 \cap \mathbb{S}^2_\varepsilon$, and V_+ be an $\varepsilon^{\frac{1}{10}}$-neighborhood of $p_2 \in \pi_2 \cap \pi_3 \cap \mathbb{S}^2_\varepsilon$. Let W be the convex hull of $V_- \cup V_+$. For non-zero integers a, b, m, consider two Hamiltonians, H_+ and H_-, defined in V_+ and V_-, respectively:

$$H_-(\theta, I) = \left(\frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon' \cos^2 \frac{\pi y}{2}\right) + \frac{I_z^2}{2} - \bar{\varepsilon} \cos^2 \frac{m\pi z}{2} \tag{48}$$

and

$$H_+(\theta, I) = \left(\frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon'' \cos^2 \frac{\pi(ay - bx)}{2}\right) + \frac{I_z^2}{2} - \bar{\varepsilon} \cos^2 \frac{m\pi z}{2}. \tag{49}$$

There exists a completely integrable Hamiltonian H on W that coincides with H_- on V_-, and with H_+ on V_+ and has the form

$$H(\theta, I) = H'(x, y, I_x, I_y, \varepsilon) + \frac{I_z^2}{2} - \bar{\varepsilon} \cos^2 \frac{m\pi z}{2},$$

where H' is completely integrable and C^1 ε-close to $\frac{I_x^2}{2} + \frac{I_y^2}{2}$ outside of $V_- \cap V_+$[3].

Proof. For the proof we shall set $a = b = m = 1$. The general case can be treated in the same way. Since the (z, I_z)-parts of H_- and H_+ are the same, we shall omit them in the following considerations. Let $\tilde{V}_\pm$, $\tilde{W}$ and $\tilde{H}_\pm$ denote the projection of $V_\pm$, W and $H_\pm$, respectively, onto the space (x, y, I_x, I_y). We shall prove that, given $\tilde{H}_\pm$ in $\tilde{V}_\pm$, where

$$\tilde{H}_- = \frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon' \cos^2 \frac{\pi y}{2}, \quad \tilde{H}_+ = \frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon'' \cos^2 \frac{\pi(x - y)}{2},$$

there exists a completely integrable Hamiltonian $\tilde{H}$ on $\tilde{W}$ that coincides with $\tilde{H}_-$ in $\tilde{V}_-$, and with $\tilde{H}_+$ on $\tilde{V}_+$. Let us omit the tilde in the notations. The proof is divided into six steps.

[3] Inside these neighborhoods V_- and V_+ we have explicit formulas.

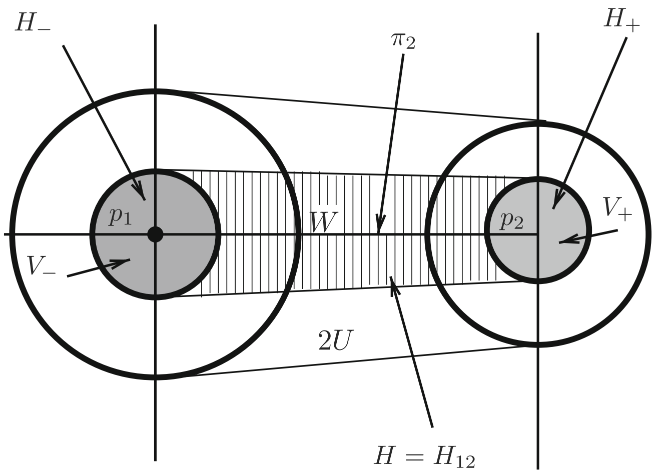

Fig. 6. Deformation neighborhoods

1) *Local reduction to a* 1.5-*degrees of freedom system.* For h near $1/2$, consider the energy surface $\{H_\pm = h\}$. Denote by U_- and U_+ the intersection of neighborhoods V_- and V_+ with this energy surface.
By the Implicit Function Theorem we can express I_x on the energy surface h. Since in our case $H_\pm = \frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon P_\pm(x, y)$, we have an explicit formula

$$I_x = \sqrt{2(h + \varepsilon P_\pm(x, y)) - I_y^2} := K_\pm(I_y, x, y; h).$$

Note that the phase trajectories of the system with the Hamiltonian $H_\pm$ on the energy surface $H_\pm = h$ satisfy the Hamiltonian equations

$$\frac{dy}{dx} = \frac{\partial K_\pm}{\partial I_y}, \quad \frac{dI_y}{dx} = -\frac{\partial K_\pm}{\partial y},$$

where x plays the role of time, and $K_\pm$ is the time-periodic Hamiltonian function defined above, see [Ar2]. We shall call these Hamiltonian systems "reduced systems".
2) *Poincaré map.* For a fixed x, consider a cylinder

$$\mathbb{A}_x = \{x\} \times \mathbb{T} \times \mathbb{R} \ni (x, y, I_y).$$

For the reduced time-periodic systems $K_\pm$, consider the Poincaré map of the cylinder $\mathbb{A}_x$ into itself, denoted by

$$\mathcal{F}_x^\pm : (y, I_y) \to (y', I_y').$$

Since $I_y \in [0, 1/\sqrt{2}]$, we have that $I_x \geq 1/(\sqrt{2}) > 0$. This implies that $\mathcal{F}_x^\pm$ are well defined. Since each $\mathcal{F}_x^\pm$ is a time-one map of a Hamiltonian vector field, it is an exact area-preserving (EAPT) map of the cylinder $\mathbb{A}_x$, see [MDS], Sect. 9.3.

3) *Analytic foliations of* $\mathbb{A}_0$. Due to complete integrability of $H_\pm$, additional integrals define two analytic foliations of $\mathbb{A}_x$ for each x, that are invariant for the Poincaré maps $\mathcal{F}_x^+$ and $\mathcal{F}_x^-$, respectively. Here we compute these two foliations:
H_- has an additional first integral

$$\frac{I_y^2}{2} - \varepsilon' \cos^2 \frac{\pi y}{2},$$

which defines a foliation on $\mathbb{A}_x$ for any x.
To see an additional first integral of H_+ we rewrite

$$\frac{I_x^2}{2} + \frac{I_y^2}{2} - \varepsilon'' \cos^2 \frac{\pi(x-y)}{2} = \frac{(I_x+I_y)^2}{4} + \frac{(I_x-I_y)^2}{4} - \varepsilon'' \cos^2 \frac{\pi(x-y)}{2}.$$

Since the change of coordinates $(x, y, I_x, I_y) \mapsto (u, v, I_u, I_v)$ with $u = (x+y)/\sqrt{2}$ and $v = (x-y)/\sqrt{2}$ is symplectic, we see that our system has two first integrals in involution:

$$(I_x + I_y) \quad \text{and} \quad \frac{(I_x-I_y)^2}{4} - \varepsilon'' \cos^2 \frac{\pi(x-y)}{2}.$$

Since we study the energy surface $H_+ = h \sim 1/2$, for each initial condition we have

$$I_x + I_y = c \quad \text{with some } c \text{ and}$$
$$\frac{(I_x-I_y)^2}{4} - \varepsilon'' \cos^2 \frac{\pi(x-y)}{2} = h - \frac{c^2}{4}.$$

Now we can express I_x and get the following first integral

$$\frac{(c-2I_y)^2}{4} - \varepsilon'' \cos^2 \frac{\pi(x-y)}{2} = h - \frac{c^2}{4}.$$

Define the following foliations of $\mathbb{A}_0$ by 2-tori:
— the map $\mathcal{F}_0^+$ preserves level sets of

$$\Gamma_c^- = \{(y, I_y) : \frac{I_y^2}{2} - \varepsilon' \cos^2 \frac{\pi y}{2} = c\};$$

— the map $\mathcal{F}_0^-$ preserves level sets of

$$\Gamma_c^+ = \{(y, I_y) : \frac{(c-2I_y)^2}{4} - \varepsilon'' \cos^2 \frac{\pi y}{2} - h + \frac{c^2}{4} = 0\}$$

with varying c.

4) *Action-angle like variables.* Denote by $\hat{\Gamma}_c^\pm$ the set of (x, y, I_y) such that a trajectory starting at this initial condition on the energy surface $H = h$ crosses the Poincaré section $\{x = 0\} \cap \Gamma_c^\pm$ respectively. These sets are 2-tori, invariant under the flow of the reduced time-periodic systems $K_\pm$, respectively. Let $\hat{\Gamma}^\pm$ be the foliations of $\mathbb{T} \times \mathbb{T} \times \mathbb{R}$ into the above tori.
We shall construct the reduced time-periodic Hamiltonian K in such a way that in U_+ it preserves the leaves of $\hat{\Gamma}^+$, in U_- it preserves the leaves of $\hat{\Gamma}^-$, and on the

whole $\mathbb{T}\times\mathbb{T}\times\mathbb{R}$ it has a smooth invariant foliation by 2-dimensional tori. We shall start by defining this invariant foliation "between" U_- and U_+. It is more convenient to do it in a new symplectic system of coordinates, (ϕ, J), in which both foliations $\Gamma^-\cap U_-$ and $\Gamma^+\cap U_+$ have the form $J = const$. For each fixed x we introduce such a symplectic change of coordinates,

$$\Phi_x : (y, I_y) \mapsto (\phi, J),$$

in $\mathbb{A}_x$ that the resulting global change of coordinates is smooth and 2-periodic in x. To do this, fix an x, and consider the leaves of the foliation $\hat\Gamma^-(x) := \mathbb{A}_x \cap \hat\Gamma^-$. Define a change of coordinates $\Phi_x^- : (y, I_y) \mapsto (\phi, J) \in \mathbb{T}\times\mathbb{R}$ so that J is the area under the invariant curve passing through (y, I_y), (evidently, J is constant on leaves of $\Gamma^-(x)$), and let ϕ be such that $(y, I_y) \to (\phi, J)$ is area-preserving. This transformation can be given by a generating function $S_x^-(y, J)$. Moreover, if $I_y > \varepsilon^{1/4}$, then the leaves of Γ^- are "almost horizontal", so in this domain

$$S_x^- = yJ + \sqrt{\varepsilon}g^-(y, J),$$

where g^- is a smooth function. In a similar way, we define $\Gamma^+(x)$ and a symplectic transformation $\Phi_x^+ : (y, I_y) \mapsto (\phi, J) \in \mathbb{T}\times\mathbb{R}$ so that J is constant on the leaves of $\Gamma^+(x)$. This transformation can be given by a generating function $S_x^+(y, J)$. If $(I_x - I_y) > \varepsilon^{1/4}$, then $S_x^+ = yJ + \sqrt{\varepsilon}g^+(y, J)$, where g^+ is a smooth function. Let $\alpha(J)$ be a smooth monotone function on $[0, \sqrt{2}/2]$ such that

$$\alpha(J) = \begin{cases} 0, & J < \frac{1}{10} \\ 1, & J > \sqrt{2}/2 - \frac{1}{10}. \end{cases}$$

Define a new function:

$$S_x(y, J) = \alpha(J)S_x^+(y, J) + (1 - \alpha(J))S_x^-(y, J).$$

If $\frac{1}{10} \le J \le \sqrt{2}/2 - \frac{1}{10}$, then the corresponding $I_y > \varepsilon^{1/4}$ and $(I_x - I_y) > \varepsilon^{1/4}$, so the corresponding generating functions $S_x^{\pm}$ are $\sqrt{\varepsilon}$-close to Jy. This implies

$$\frac{\partial^2 S_x(y, J)}{\partial y\, \partial J} \ne 0.$$

If $J < \frac{1}{10}$, then $S_x = S_x^-$, and for $J > \sqrt{2}/2 - \frac{1}{10}$, $S_x = S_x^+$, so the above inequality holds also. Therefore, $S_x(y, J)$ is a generating function of a symplectic transformation. This is the desired change of coordinates Φ_x.
Since both $S_x^{\pm}$ are smooth and periodic in x, the same is true for Φ_x. By construction, Φ_x coincides with Φ_x^- for $I_y < \frac{1}{20}$, and with Φ_x^- for $I_y > \sqrt{2}/2 - \frac{1}{20}$.

5) *Construction of the reduced time-periodic Hamiltonian K.* In the new coordinates both $\hat\Gamma^-\cap U_-$ and $\hat\Gamma^+\cap U_+$ have the form $J = const$, because the flow of $K_\pm$ preserves the area form $dy\wedge dI_y$. Now we shall construct the reduced time-periodic system K so that it preserves the foliation $J = const.$ for all J. In order to do this, we define a smooth family of symplectic maps $F_x : (J_0, \phi_0, 0) \mapsto (J_x, \phi_x, x)$ with $F_0 = id$, and then, by a standard fact in symplectic topology, produce an x-periodic Hamiltonian function whose time-x map coincides with F_x for each x. Note that

for $(J_0, \phi_0) \in U_\pm$, this time-x map is already defined as the flow map of $K_\pm$, and has the form

$$\begin{cases} J_x = J_0, \\ \phi_x = \phi_0 + \omega_x^\pm(J_0, \varepsilon), \end{cases} \tag{50}$$

where $\omega_x^\pm$ are smooth in (x, J, ε), monotone increasing in J, and for all x we have:

$$\max_{W_x^-} \omega_x^-(J, \varepsilon) < \min_{W_x^+} \omega_x^+(J, \varepsilon) - c \quad \text{for some} \quad c > 0.$$

Moreover, $\omega_x^\pm(J, \varepsilon)$ converges smoothly to the function Jx when $\varepsilon \to 0$. Let $\omega_x(J)$ be a function, increasing in J for each x and smooth in (x, I), such that

$$\omega_x(J, \varepsilon) = \begin{cases} \omega_x^-(J, \varepsilon), & J < 1/10 \\ \omega_x^+(J, \varepsilon), & J > \sqrt{2}/2 - 1/10. \end{cases} \tag{51}$$

Such a function exists due to the properties of $\omega_x^\pm$. For each x define $F_x : \mathbb{A}_0 \mapsto \mathbb{A}_x$ as

$$\begin{cases} J_x = J_0, \\ \phi_x = \phi_0 + \omega_x(J_0, \varepsilon). \end{cases}$$

This map is area-preserving for each x by construction, and $F_0 = id$. Moreover, for each x,

$$F_x^* \lambda - \lambda = J_1 d\phi_1 - J_0 d\phi_0 = J_0 \frac{\partial \omega(J_0, x)}{\partial J_0} dJ_0 = dG_x(J_0),$$

and G_x is smooth in x. Therefore, the flow F_x, $x \geq 0$, is generated by a Hamiltonian function K, see [MDS], Prop 9.19. This is the desired reduced time-periodic Hamiltonian K.

6) *Construction of* $\tilde{H}(I, \theta; \varepsilon)$. By a standard procedure, we construct the corresponding autonomous Hamiltonian $\tilde{H}$ in the variables (x, I_x, J, ϕ). Returning to the variables (x, y, I_x, I_y), we obtain the desired $\tilde{H}$.

Note that J is the first integral of $\tilde{H}$ by construction, and J is in involution with $\tilde{H}$. The latter follows from the fact that $\partial \tilde{H}/\partial \phi = 0$.

We note also that if $\varepsilon = 0$ then both H_+ and H_- equal $\frac{I_x^2}{2} + \frac{I_y^2}{2}$. Naturally we define $\tilde{H}_0$ by the same formula. This can also be achieved by the construction above choosing $\omega_x(J, 0)$ in formula (51) to be $\omega_x(J, 0) = Jx$.

Since for $\varepsilon = 0$ we have $\tilde{H}(I, \theta; \varepsilon) = \frac{I_1^2}{2} + \frac{I_2^2}{2}$, the difference $\tilde{H}(I, \theta; \varepsilon) - (\frac{I_1^2}{2} + \frac{I_2^2}{2})$ goes to zero with ε. This shows that, for small ε, $\tilde{H}$ is convex. For the corresponding Lagrangian $\tilde{L}(\phi, \dot{\phi}, \varepsilon)$, this implies that $\tilde{L}(\phi, \dot{\phi}; \varepsilon) - (\frac{\dot{\phi}_1^2}{2} + \frac{\dot{\phi}_2^2}{2})$ goes to zero with ε. Outside of $\varepsilon^{1/10}$-neighborhood influence of $\varepsilon \cos(\cdot)$ becomes negligible. This completes the proof of the Lemma. □

7. Construction of the Fractal Set F and the Hamiltonian $H{=}H_\varepsilon$

Recall that $\mathbb{S}^2_\varepsilon$ denotes the natural projection of the energy surface $\{H = 1/2\}$ onto the space of actions. Then $\mathbb{S}^2_\varepsilon$ is a smooth ε-small deformation of a unit sphere in $\mathbb{R}^3$. Fix a fast decaying sequence $\{\varepsilon_n\}_n$ of positive real numbers. Below we define a fractal set F with its closure homeomorphic to the "H-tree" F^{mod}, see Fig. 1. Namely, $F = \cup_{n\in\mathbb{Z}_+} F_n$, where $\cup_{k\le n} F_k$ is a tree of generation n (consisting of $\sum_{j=0}^n 2^j$ segments up to generation n).

Let $\bar n$ stand for an n-tuple of 0's and 1's. The set F_0 consists of one segment passing through zero. Suppose that $l_{\bar n}$ is one of the segments of the n^{th} generation. Two segments of the $(n + 1)^{\text{st}}$ generation, called $l_{\bar n,0}$ and $l_{\bar n,1}$, are are attached by their midpoints to the ends of $l_{\bar n}$. Segments $l_{\bar n,0}$ and $l_{\bar n,1}$ are almost perpendicular to $l_{\bar n}$. More details are in Sect. 7.1. In this section we prove the following theorem.

Theorem 3. *1. There exist a fractal set F of type "H-tree", see Fig. 1, on $\mathbb{S}^2_\varepsilon$ with Hausdorff dimension of $\overline{F}$ equal to 2, any of its segments being the intersection of $\mathbb{S}^2_\varepsilon$ with an appropriate "resonant plane" of the form $\{I \mid (k\cdot I) = 0\}$ for $k \in \mathbb{Z}^3$. Moreover, for any of its segments, say, $l_{\bar n}$, of generation n, there is an open set $U_{\bar n}$ containing $l_{\bar n}$ that does not intersect any segment of generation higher than $\ge (n + 2)$ or lower than $\le n - 2$.*

2. There is a C^r-smooth Hamiltonian $H(\theta, I) = H^{int}(\theta, I) + \tilde\beta(\theta, I, \varepsilon)$, depending on $\{\varepsilon_n\}_n$, defined in a neighborhood of F with the following properties.

- *$H(\theta, I) - \frac{|I|^2}{2}$ vanishes on $\overline{F} \setminus F$.*
- *For each vertex, $p_{\bar n}$, of F there is a neighborhood $V_{\bar n}$ and an integer matrix $M_{\bar n} \in GL_3(\mathbb{Z})$ such that for $I \in V_{\bar n}$ the Hamiltonian H^{int} has the form (37), where $\varepsilon = \varepsilon_{n-1}$, $\varepsilon' = \varepsilon_n$, matrix $M = M_{\bar n}$ is defined in (56), and the integer constant $a = c_{\bar n}$ is defined below in (52) and (54).*
- *For each segment, $l_{\bar n}$, of F there is an open set $U_{\bar n} \supset l_{\bar n}$ in which H^{int} has the form (14) with $\varepsilon = \varepsilon_n$, the matrix $M = M_{\bar n}$ defined in (56), and the constant $a = c_{\bar n}$ is defined in (52) and (54).*
- *One can define a C^r-smooth "bump" function $\tilde\beta(I, \theta, \varepsilon)$, 2-periodic in θ, whose support is contained in the union of convex sets called lenses. For $I \in V_{\bar n}$, the corresponding lens has the form $M^{-1}_{\bar n} B(\mathbf{n}, \sqrt{\varepsilon_n})$ where $B(\mathbf{n}, \sqrt{\varepsilon_n})$ is a ball centered at $\mathbf{n} \in 2\mathbb{Z}^3$ of radius $\sqrt{\varepsilon_n}$, and*

$$\mathbf{n} \in \Gamma_{\bar n} = M^{-1}_{\bar n}(2, 2, \frac{2}{c_{\bar n}}) \cdot \mathbb{Z}^3.$$

For $I \in W_{\bar n,i}$, $i = 0$ or 1, the definition of a "bump" function $\beta(I, \theta, \varepsilon)$ is more involved, see Sect. 7.3. Roughly speaking, it is a smooth deformation between the aforementioned bump function supported in $M^{-1}_{\bar n} B(\mathbf{n}, \sqrt{\varepsilon_n})$ and the bump function from the next stage, supported in $M^{-1}_{\bar n,i} B(\mathbf{n}', \sqrt{\varepsilon_{n+1}})$ with $\mathbf{n}' \in \Gamma_{\bar n,i} = M^{-1}_{\bar n,i}(2, 2, \frac{2}{c_{\bar n,i}}) \cdot \mathbb{Z}^3$.

In Subsect. 7.1 we construct the fractal set F; Subsect. 7.2 is devoted to the construction of the integrable Hamiltonian H^{int}. In Subsect. 7.3 we construct the perturbation term $\beta(I, \theta, \varepsilon)$ supported on the lenses. The desired Hamiltonian is $H(I, \theta) = H^{int}(I, \theta, \varepsilon) + \varepsilon^{r+1}\beta(I, \theta, \varepsilon)$.

7.1. Construction of the fractal set F. Our model for F is the "H-tree" F^{mod}, see Fig. 1. It is obtained as a limit of the following iterative procedure. Given $\alpha > 0$ and a monotone increasing sequence λ_n, $0 < \lambda_n < \frac{1}{\sqrt{2}}$, $\lambda_n \to \frac{1}{\sqrt{2}}$, start with a straight line segment l_0 of length α (l_0 is the segment "of level 0"). The segments of the H-tree, denoted above by $l_{\bar{n}}$, are indexed by dyadic numbers as follows. Inductively, two new segments, $l_{\bar{n},0}$ and $l_{\bar{n},1}$ are attached by the midpoints to the endpoints of $l_{\bar{n}}$. They are perpendicular to $l_{\bar{n}}$ and have equal lengths $|l_{\bar{n},0}| = |l_{\bar{n},1}| = \lambda_n |l_{\bar{n}}|$. The set F^{mod} is the closure of the union of these segments. One way to verify that $\dim_H(F^{mod}) = 2$ is to define F^{mod} using an iterated function scheme (see, e.g., Falkoner's book [Fa]).

The construction of $F \subset \mathbb{S}^2_\varepsilon$ is a little more involved: here the consecutive segments are no longer parallel, and $\lambda_{\bar{n}}$ depends on the whole sequence $\bar{n}$. Namely, we define $F = \cup_{n\in\mathbb{Z}_+} F_n$, where F_n is a "tree of generation n". Consider the planes: $\pi_0 = \{I_y = 0\}$, $\pi_{01} = \{I_z = 0\}$. Let $\pi_{011} = \{aI_y - bI_x = 0\}$, $\pi_{010} = \{aI_y + bI_x = 0\}$ for some $a, b \in \mathbb{Z}$. Let $p_{01} = \pi_0 \cap \pi_{01} \cap \mathbb{S}^2_\varepsilon$, $p_{011} = \pi_{01} \cap \pi_{011} \cap \mathbb{S}^2_\varepsilon$, and $p_{010} = \pi_{01} \cap \pi_{010} \cap \mathbb{S}^2_\varepsilon$. Denote by l_0 the segment of $\pi_0 \cap \mathbb{S}^2_\varepsilon$ with one end at p_{01} and the length $|l_0| = \alpha$ sufficiently small. Denote by l_{01} the segment of $\pi_{01} \cap \mathbb{S}^2_\varepsilon$ between p_{010} and p_{011}. We assume that a and b are chosen so that $|l_{01}| = |l_0|\lambda_0$ for some λ_0 "close to" $\frac{1}{\sqrt{2}}$.

To continue, suppose that $(\theta_{\bar{n}}, I_{\bar{n}})$ is a system of symplectic coordinates in a neighborhood of $\mathbb{S}^2_\varepsilon$, and suppose we have defined two consecutive segments of F: $l_{\bar{n}} \subset \pi_{\bar{n}} \cap \mathbb{S}^2_\varepsilon$ and $l_{\bar{n}1} = \pi_{\bar{n}1} \cap \mathbb{S}^2_\varepsilon$, where

$$\pi_{\bar{n}} = \{I_{y_{\bar{n}}} = 0\}, \quad \pi_{\bar{n}1} = \{I_{z_{\bar{n}}} = 0\}.$$

(The case of $l_{\bar{n}}$ and $l_{\bar{n}0}$ is analogous). Choose integers $a_{\bar{n}}, b_{\bar{n}}$, and define

$$\pi_{\bar{n}10} = \{a_{\bar{n}} I_{y_{\bar{n}}} + b_{\bar{n}} I_{x_{\bar{n}}} = 0\}, \quad \pi_{\bar{n}11} = \{a_{\bar{n}} I_{y_{\bar{n}}} - b_{\bar{n}} I_{x_{\bar{n}}} = 0\} \tag{52}$$

in such a way that the following two requirements hold:

1. Let $l_{\bar{n}1}$ denote the interval connecting the points $p_{\bar{n}10} := \pi_{\bar{n}1} \cap \pi_{\bar{n}10} \cap \mathbb{S}^2_\varepsilon$ and $p_{\bar{n}11} := \pi_{\bar{n}1} \cap \pi_{\bar{n}11} \cap \mathbb{S}^2_\varepsilon$. The length of $l_{\bar{n}1}$ satisfies
$$|l_{\bar{n}1}| = |l_{\bar{n}}|\lambda_{\bar{n}},$$
where $\lambda_{\bar{n}}$ satisfies
$$\lambda_{\bar{n}} = 1/\sqrt{2} - \delta_n, \quad \text{where } \alpha 2^{-(n+1)/4} < \delta_n < \alpha 2^{-n/4}. \tag{53}$$
2. Integers $a_{\bar{n}}$ and $b_{\bar{n}}$ in the definition of $\pi_{\bar{n}11}$ satisfy:
$$c_{\bar{n}} := \sqrt{a_{\bar{n}}^2 + b_{\bar{n}}^2} \in \mathbb{N}. \tag{54}$$
In this case the triple $(a_{\bar{n}}, b_{\bar{n}}, c_{\bar{n}})$ is called "Pythagorean triple". This assumption is used, in particular, in Sect. 7.2.

Recall that a Pythagorean triple is a triple (a, b, c) of integers such that $a^2 + b^2 = c^2$. Such triples can be generated as follows:

$$a = m^2 - n^2, \quad b = 2mn, \quad c = m^2 + n^2, \quad \text{where } m > n > 0, \quad m, n \in \mathbb{Z}.$$

The latter implies that for any constants β and δ there is a Pythagorean triple $(a, b, \sqrt{a^2 + b^2})$ such that $|\beta - \frac{a}{b}| < \delta$. This implies, in its turn, that requirements 1 and 2 above can be met simultaneously.

7.1.1. Coordinate change and related notations. To make the recursion work, consider the following symplectic **change of coordinates**. For a fixed $\bar{n}$, it acts on θ_n by the integer matrix $T_{\bar{n}}$:

$$\theta_{\bar{n}1} = T_{\bar{n}1}\theta_{\bar{n}}, \quad T_{\bar{n}1} = \begin{pmatrix} a_{\bar{n}} & b_{\bar{n}} & 0 \\ 0 & 0 & -c_{\bar{n}} \\ -b_{\bar{n}} & a_{\bar{n}} & 0 \end{pmatrix}. \tag{55}$$

The change to $\theta_{\bar{n}0}$ is analogous, the corresponding matrix $T_{\bar{n}0}$ is as $T_{\bar{n}1}$ with $b_{\bar{n}}$ replaced by $-b_{\bar{n}}$.

For the fixed $\bar{n}$, let $\bar{n}_j$ stand for the sequence of first j digits of $\bar{n}$. We denote

$$M_{\bar{n}} = T_{\bar{n}} T_{\bar{n}_{n-1}} \dots T_{\bar{n}_2}; \quad \text{then } \theta_{\bar{n}} = M_{\bar{n}}\theta_0. \tag{56}$$

Define

$$I_{\bar{n}1} = T^*_{\bar{n}1} I_{\bar{n}}, \quad \text{where } T^*_{\bar{n}1} = (T^{-1}_{\bar{n}1})^{tr} = \begin{pmatrix} \frac{a_{\bar{n}}}{c^2_{\bar{n}}} & \frac{b_{\bar{n}}}{c^2_{\bar{n}}} & 0 \\ 0 & 0 & -\frac{1}{c_{\bar{n}}} \\ -\frac{b_{\bar{n}}}{c^2_{\bar{n}}} & \frac{a_{\bar{n}}}{c^2_{\bar{n}}} & 0 \end{pmatrix}.$$

The resulting coordinate change is symplectic. Notice that it brings the planes $\pi_{\bar{n}1}$ and $\pi_{\bar{n}11}$ to the form

$$\pi_{\bar{n}1} = \{I_{y_{\bar{n}1}} = 0\}, \quad \pi_{\bar{n}11} = \{I_{z_{\bar{n}1}} = 0\},$$

and the segment $l_{\bar{n}1} \subset \pi_{\bar{n}1} \cap \mathbb{S}^2_\varepsilon$ is defined above by its endpoints $p_{\bar{n}1}$ and $p_{\bar{n}11}$. This ends the recursive construction.

About the coordinate change, notice that

$$H_0(I) = \left(\prod_{j=1}^{n-1} c^2_{\bar{n}_j}\right) H_0(I_{\bar{n}}).$$

Here we study the structure of F and prove that $\dim_H(F) = 2$.

Lemma 12. *The set F constructed above has no self-intersections. Moreover, each segment $l_{\bar{n}}$ of generation n has an open neighborhood of width $d_n \sim 2^{-n}$, which does not intersect any segment of generation $\geq (n+2)$ and $\leq (n-2)$.*

Proof. Fix two consecutive segments: $l_{\overline{n-1}}$ and $l_{\bar{n}}$. Let $F(l_{\bar{n}})$ denote the connected component of $F \setminus l_{\overline{n-1}}$ containing $l_{\bar{n}}$. Informally, it is the branch of the tree growing on $l_{\bar{n}}$. We shall show that the distance $d_{\overline{n-1}}$ between $l_{\overline{n-1}}$ and $F(l_{\bar{n}})$ is larger than some positive number depending only on n. For this lemma denote $s_k = \frac{1}{2}\max|l_{\bar{k}}|$, where the maximum is taken over all segments of generation k in $F(l_{\bar{n}})$. In particular, $s_n = |l_{\bar{n}}|$. If α is sufficiently small, we can estimate:

$$\begin{aligned} d_{\overline{n-1}} &\geq s_n - \sum_{j=1}^{\infty} s_{n+2j} - 2\sum_{j=0}^{\infty} s_{n+2j+1} \sin\left(\sum_{j=0}^{\infty} s_{n+2j}\right) \\ &\geq s_n - 2\left(\frac{1}{\sqrt{2}} - \delta_n\right)\left(\frac{1}{\sqrt{2}} - \delta_{n+1}\right) s_n - 4 s_{n+1} \sin s_n \\ &\geq s_n \left(\frac{1}{\sqrt{2}}(\delta_n + \delta_{n+1}) - \delta_n\delta_{n+1} - 4s_{n+1}\right). \end{aligned}$$

Since $\alpha 2^{-n/4} \leq \delta_n \leq \alpha 2^{1-n/4}$ by assumption, and s_n is of order $2^{-n/2}$, the latter expression is positive, and moreover, for small α we have:

$$d_{\overline{n-1}} \geq \frac{1}{2} s_n^{3/2}.$$

Value d_n can be chosen to be as large as the minimum of $d_{\bar{n}}$ over all $|\bar{n}| = n$. Value $d_n \sim 2^{-n}$ also works. □

To estimate the Hausdorff dimension, we use the following lemma from [Fa].

Lemma 13. *If $f : G \to \mathbb{R}^m$ is a bi-Lipschitz transformation, then* $\dim_H f(G) = \dim_H G$.

This lemma can indeed be used due to the following:

Lemma 14. *Let $\overline{F} \subset \mathbb{S}^2_\varepsilon$ be the set constructed above, and $F^{mod} \subset \mathbb{R}^2$ be the model set constructed in the beginning of the section with the same $\lambda_{\bar{n}}$. Then there is a bi-Lipschitz transformation $f : \overline{F} \to F^{mod}$.*

Proof. By construction, any segment $l_{\bar{n}}$ of F has an open neighborhood free from segments of order $\geq (n+2)$. Moreover, any two intersecting segments, $l_{\bar{n}}$ and $l_{\overline{n+1}}$, intersect at right angle. This implies the existence of a transformation claimed by the lemma. □

7.2. Construction of H^{int}. Let $F \subset \mathbb{S}^2_\varepsilon$ be the fractal set constructed in the previous section. For any $\bar{n}$, let $V_{\bar{n}}$ be a ball centered at $p_{\bar{n}}$ with radius $d_n/2$ (where d_n is defined in Lemma 12, e.g. $d_n = 2^{-n}$), and $2V_{\bar{n}}$ be a ball centered at $p_{\bar{n}}$ with twice the radius of $V_{\bar{n}}$, i.e., d_n. Let $U_{\bar{n},1}$ be the convex hull of $V_{\bar{n}}$ and $V_{\bar{n},1}$, and $2U_{\bar{n},1}$ denote the convex hull of $2V_{\bar{n}}$ and $2V_{\bar{n},1}$. See Fig. 6 with $V_- = V_{\bar{n}}$, $V_+ = V_{\bar{n},1}$, $U = U_{\bar{n}}$ similarly $2V_- = 2V_{\bar{n}}$, $2V_+ = 2V_{\bar{n},1}$ are balls of double radius with the same center. Since the width of $2U_{\bar{n},1}$ is smaller than d_n, $2U_{\bar{n},1}$ does not intersect $l_{\bar{j}}$ with $j \neq \bar{n}$, $(\bar{n}, 1)$, see Lemma 12. Introduce analogous notation for $\bar{n}, 0$.

Let $\frac{1}{2}V_{\bar{n}}$ be a ball centered at $p_{\bar{n}}$ of half the radius of $V_{\bar{n}}$. Define $W_{\bar{n},1}$ to be the convex hull of $V_{\bar{n}} \cup V_{\bar{n},1}$ minus $V_{\bar{n}} \cup V_{\bar{n},1}$. The sets $W_{\bar{n}}$ are the neighborhoods in which we study the single-resonant case. Let (ε_n), $n \geq 0$, be a sufficiently fast decaying sequence of positive constants. Let $2U_{\overline{n}}$ and $2U_{\bar{n},i}$ be two consecutive neighborhoods with $i = 0$ or 1. For $I \in 2U_{\overline{n}} \cup 2U_{\bar{n},i}$, multiply ε_n by a cut-off function τ_n,

$$\varepsilon_{\bar{n}}(I) = \begin{cases} 1, & I \in U_{\bar{n}} \cup U_{\bar{n},i} \\ 0, & I \notin 2U_{\bar{n}} \cup 2U_{\bar{n},i}. \end{cases} \tag{57}$$

Since ε_n decays fast with n and each neighborhood $U_{\bar{n}}$ is the convex hull of $V_{\bar{n}}$ and $V_{\bar{n},i}$, it contains an $\varepsilon_n^{\frac{1}{10+r}}$-neighborhood of $l_{\bar{n}}$. Thus, $\varepsilon_{\bar{n}}(I)$ can be chosen C^r-small. Define $(\theta_0, I_0) \equiv (\theta, I)$. Let $a_{\bar{n}}$, $b_{\bar{n}}$ and $c_{\bar{n}}$ be the integers that appeared in the definition (52) of the resonant plane $\pi_{\bar{n}11}$. Define H^{int} by the following recursive procedure:

– For $I \in V_{01}$, define

$$H^{int}(\theta, I) = H_0(I) - \varepsilon_0 \cos^2 \frac{\pi}{2} y - \varepsilon_1 \cos^2 \frac{\pi}{2} c_0 z;$$

Fix an $\bar n \in \mathbb{Z}_2^n$. As before, $\bar n_j$ denotes the sequence of first j digits of $\bar n$, $j = 1, \dots, n$. Suppose that H^{int} is defined for $I \in V_{\bar n}$ and, written in coordinates $(\theta_{\bar n}, I_{\bar n})$, has the form

$$H^{int}(\theta_{\bar n}, I_{\bar n}) = \left(\prod_{j=1}^{n-1} c_{\bar n_j}^2\right) H_0(I_{\bar n}) - \varepsilon_{n-1}\cos^2\frac{\pi}{2} y_{\bar n} - \varepsilon_n \cos^2\frac{\pi}{2} c_{\bar n} z_{\bar n}. \tag{58}$$

For $I \in V_{\bar n 1}$ define $H^{int}(\theta_n, I_n)$ as follows:

$$H^{int} = \left(\prod_{j=1}^{n-1} c_{\bar n_j}^2\right) H_0(I_{\bar n}) - \varepsilon_{n+1}\cos^2\frac{\pi}{2} c_{\bar n,1}(a_{\bar n} y_{\bar n} - b_{\bar n} x_{\bar n}) - \varepsilon_n \cos^2\frac{\pi}{2} c_{\bar n} z_{\bar n}. \tag{59}$$

The above pair of Hamiltonian functions has the form (48), (49). For $I \in U_{\bar n,1}$ by Lemma 11 there is an integrable Hamiltonian denoted $H^{int}(\theta_{\bar n}, I_{\bar n})$.

To continue the recursive process, notice that (55) gives $H_0(I_{\bar n,1}) = c_{\bar n}^{-2} H_0(I_{\bar n})$, and that for $I \in V_{\bar n,,1}$, H^{int} written in the new symplectic coordinates $(\theta_{\bar n,1}, I_{\bar n,1})$ (given by (55)), has the form

$$H^{int}(\theta_{\bar n 1}, I_{\bar n,1}) = \left(\prod_{j=1}^{n} c_{\bar n_j}^2\right) H_0(I_{\bar n,1}) - \varepsilon_n \cos^2\frac{\pi}{2} y_{\bar n,1} - \varepsilon_{n+1}\cos^2\frac{\pi}{2} c_{\bar n,1} z_{\bar n,1}. \tag{60}$$

We define H^{int} for $I \in V_{\bar n,0}$ in an analogous way. This completes the recursive step of the construction.

Remark 2. Notice that for the above Hamiltonian we have:

$$H^{int}(\theta, I) = H_0(\theta, I) \quad \text{for} \;\; I \in \overline{F} \setminus F.$$

Indeed, I cannot lie in any of U_n (U_n are defined in the beginning of this section) since U_n intersects at most two different segments of F.

Denote

$$\mathcal{U} = \bigcup_{\bar n} U_{\bar n}, \quad \text{and} \quad F^\infty = (\overline{F} \setminus F) \cap \mathcal{K}, \tag{61}$$

where $\mathcal{K}$ stands for the set of rationally independent vectors in $\mathbb{R}^3$. Then we have: $\mathrm{supp}(H^{int}) \subset \mathcal{U} \times \mathbb{T}^3$, and $F^\infty \subset \partial\mathcal{U}$.

7.3. Definition of the lenses. Here we construct the "lenses"—convex sets whose union contains the support of the perturbation β. We construct them on the universal cover of $\mathbb{T}^3$.

Fix a multi-index $\bar n \in \mathbb{Z}_2^n$. For any $I \in V_{\bar n}$ first define the lenses $\mathcal{L}_{\mathbf{n}}$ in terms of coordinates $(\theta_{\bar n}, I_{\bar n}) = (M_{\bar n}\theta, M_{\bar n}^* I)$. For each $I_{\bar n} \in M_{\bar n}^* V_{\bar n}$, the lenses are balls in variables $\theta_{\bar n}$ of radius $\sqrt{\varepsilon_n}$ centered at the grid of points

$$\mathbf{n} \in \widetilde{\Gamma}_{\bar n} = (2, 2, \frac{2}{c_{\bar n}}) \cdot \mathbb{Z}^3. \tag{62}$$

So, we have $\mathcal{L}_{\mathbf{n}} = B(\mathbf{n}, \sqrt{\varepsilon_n})$. This choice is motivated by the following. For $I \in V_{\bar{n}}$ the Hamiltonian H^{int}, written in coordinates $(\theta_{\bar{n}}, I_{\bar{n}})$, has the form (58). Here we have the lenses placed at the points with $(c_{\bar{n}}\tilde{z}_{\bar{n}})$ being an even integer. This permits us to use methods of Sect. 5. In terms of the original coordinates, the lenses for $I \in V_{\bar{n}}$ are convex sets of the form $M_{\bar{n}}^{-1}B(\mathbf{n}, \sqrt{\varepsilon_n})$ placed at the points

$$\mathbf{n} \in \Gamma_{\bar{n}} = M_{\bar{n}}^{-1}\left(2, 2, \frac{2}{c_{\bar{n}}}\right) \cdot \mathbb{Z}^3. \tag{63}$$

Define the deformation as

$$\beta_{\bar{n},\mathbf{n}}(\theta_{\bar{n}}, I_{\bar{n}}, \varepsilon_n) = \zeta\left(\frac{|\theta_{\bar{n}} - \mathbf{n}|}{\varepsilon_n}\right),$$

where, to be specific, we take $\zeta([0, 1/2]) = 1$, $\zeta([1, \infty)) = 0$ with ζ being C^∞-smooth on $\mathbb{R}$, monotone decreasing on $[1/2, 1]$ and even. Then

$$\beta_{\bar{n}}(\theta_{\bar{n}}, I_{\bar{n}}, \varepsilon_n) = \sum_{\mathbf{n}} \beta_{\bar{n},\mathbf{n}}(\theta_{\bar{n}}, I_{\bar{n}}, \varepsilon_n).$$

The deformation is thus $\varepsilon_n^{r+1}\beta_{\bar{n}}(\theta_{\bar{n}}, I_{\bar{n}}, \varepsilon_n)$. In terms of the original coordinates, set $\beta(\theta, I) := \sum_{\bar{n}} \beta_{\bar{n}}(M_{\bar{n}}\theta, M_{\bar{n}}^* I, \varepsilon_n)$.

It is left to describe the lenses for $I \in W_{\bar{n}1}$. Let us first do it in terms of $(\theta_{\bar{n}}, I_{\bar{n}})$. By the formula above, for $I \in V_{\bar{n}1}$, we have defined the lenses in terms of coordinates $(\theta_{\bar{n}1}, I_{\bar{n}1})$. They are balls of radius $\sqrt{\varepsilon_{n+1}}$ centered at the points of the lattice $\widetilde{\Gamma}_{\bar{n}1} = 2\left(1, 1, \frac{1}{c_{\bar{n}1}}\right)\cdot\mathbb{Z}^3$. Written in coordinates $(\theta_{\bar{n}}, I_{\bar{n}})$, $\widetilde{\Gamma}_{\bar{n}1}$ has the form $\widetilde{\Gamma}'_{\bar{n}1} = T_{\bar{n}1}^{-1}\widetilde{\Gamma}_{\bar{n}1}$ ($T_{\bar{n}1}$ in the transformation defined by (55)). One can compute that

$$\widetilde{\Gamma}'_{\bar{n}1} \subset 2\left(\frac{1}{c_{\bar{n}}^2 c_{\bar{n}1}}, \frac{1}{c_{\bar{n}}^2 c_{\bar{n}1}}, \frac{1}{c_{\bar{n}}}\right) \cdot \mathbb{Z}^3.$$

Notice that the third components of the above lattice and those of $\widetilde{\Gamma}_{\bar{n}}$ defined in (62) constitute the same set. This is important for constructing the lenses for $I \in W_{\bar{n}}$.

Recall that, by definition, $I \in V_{\bar{n}}$ is equivalent to $I_{\bar{n}} \in M_{\bar{n}}^* V_{\bar{n}}$, $I \in W_{\bar{n}1}$ is equivalent to $I_{\bar{n}} \in M_{\bar{n}}^* W_{\bar{n}1}$. Denote

$$\widetilde{V}_{\bar{n}} = M_{\bar{n}}^* V_{\bar{n}}, \quad \widetilde{V}_{\bar{n}1} = M_{\bar{n}}^* V_{\bar{n}1}, \quad \widetilde{W}_{\bar{n}1} = M_{\bar{n}}^* W_{\bar{n}1}.$$

Recall that $H^{int}(\theta_{\bar{n}}, I_{\bar{n}})$ has form (58) for $I_{\bar{n}} \in \widetilde{V}_{\bar{n}}$, and form (59) for $I_{\bar{n}} \in \widetilde{V}_{\bar{n}1}$. The Hamiltonian H^{int} for $I_{\bar{n}} \in \widetilde{W}_{\bar{n}1}$ was constructed using the "continuation" procedure of Sect. 6, which does not affect the $(I_{\bar{n}}, z_{\bar{n}})$-part of the Hamiltonian.

By now, the lenses have been defined for $I_{\bar{n}} \in \widetilde{W}_{\bar{n}1} \cap \widetilde{V}_{\bar{n}}$ and for $I_{\bar{n}} \in \widetilde{W}_{\bar{n}1} \cap \widetilde{V}_{\bar{n}1}$. Let $\psi(I_{\bar{n}})$ and $\xi(I_{\bar{n}})$ be smooth functions supported on $\widetilde{W}_{\bar{n}1}$, taking values between 0 and 1,

$$\psi(I_{\bar{n}}) = \begin{cases} 1 & \text{for} \quad I_{\bar{n}} \in \widetilde{V}_{\bar{n}} \cap \widetilde{W}_{\bar{n}1} \\ 0 & \text{for} \quad I_{\bar{n}} \in 2\widetilde{V}_{\bar{n}1} \cap \widetilde{W}_{\bar{n}1}, \end{cases} \qquad \xi(I_{\bar{n}}) = \begin{cases} 1 & \text{for} \quad I_{\bar{n}} \in \widetilde{V}_{\bar{n}1} \cap \widetilde{W}_{\bar{n}1} \\ 0 & \text{for} \quad I_{\bar{n}} \notin 2\widetilde{V}_{\bar{n}1}. \end{cases}$$

For each lens corresponding to $I_{\bar{n}} \in \widetilde{V}_{\bar{n}}$ or $I_{\bar{n}} \in \widetilde{V}_{\bar{n}1}$ we define a whole family of lenses smoothly depending on $I_{\bar{n}} \in \widetilde{W}_{\bar{n}1}$. For each $I_{\bar{n}}$, the corresponding lens $\mathcal{L}_{\mathbf{n}}$ is a round ball in $\theta_{\bar{n}}$-variables, centered at $\mathbf{n}$, of radius $s_{\mathbf{n}}$. Distinguish 3 cases:

— If $\mathbf{n} \in \widetilde{\Gamma}_{\bar{n}} \cap \widetilde{\Gamma}'_{\bar{n}1}$, then for each $I_{\bar{n}} \in \widetilde{W}_{\bar{n}1}$, define the radius $s_{\mathbf{n}}(I_{\bar{n}}) = \sqrt{\varepsilon_n}\psi(I_{\bar{n}}) + \sqrt{\varepsilon_{n+1}}(1 - \psi(I_{\bar{n}}))$.
— If $\mathbf{n} \in \widetilde{\Gamma}_{\bar{n}} \backslash \widetilde{\Gamma}'_{\bar{n}1}$, then define $s_{\mathbf{n}}(I_{\bar{n}}) = \sqrt{\varepsilon_n}\psi(I_{\bar{n}})$.
— If $\mathbf{n} \in \widetilde{\Gamma}'_{\bar{n}1} \backslash \widetilde{\Gamma}_{\bar{n}}$, then define $s_{\mathbf{n}}(I_{\bar{n}}) = \sqrt{\varepsilon_{n+1}}\xi(I_{\bar{n}})$.

The deformation $\beta(\theta, I)$ is defined to be a smooth function, supported on the lenses. Define β separately for each smooth family of lenses discussed above in the following way. Let

$$\beta_{\bar{n},\mathbf{n}}(\theta_{\bar{n}}, I_{\bar{n}}) = \begin{cases} r_n^{2r+2} & \text{for } I_{\bar{n}} \in B\left(\mathbf{n}, \frac{r_{\mathbf{n}}^2(I_{\bar{n}})}{2}\right) \\ 0 & \text{for } I_{\bar{n}} \notin B\left(\mathbf{n}, r_{\mathbf{n}}^2(I_{\bar{n}})\right), \end{cases}$$

where $\mathbf{n}$ is the center of the corresponding lens. Then

$$\beta_{\bar{n}}(\theta_{\bar{n}}, I_{\bar{n}}) = \sum_{\mathbf{n}} \beta_{\bar{n},\mathbf{n}}(\theta_{\bar{n}}, I_{\bar{n}}).$$

Having constructed the lenses and the deformation function, we return to the original coordinates by the transformation $(M^*, M)^{-1}$.

In this section we complete the proof of Theorem 1. Recall that the H-fractal $\overline{F} \subset \mathbb{S}^2$ constructed in Sect. 7 has Hausdorff dimension 2. Recall that $\mathcal{K}$ stands for the set of totally irrational vectors in $\mathbb{R}^3$. We define $F^\infty = \left(\overline{F} \backslash F\right) \cap \mathcal{K}$, see (61). Its Hausdorff dimension is also 2. Finally, define the set

$$\mathcal{F}^\infty = \mathbb{T}^3 \times F^\infty.$$

It has Hausdorff dimension 5. In Sect. 7 we have constructed a Hamiltonian H on the set $\mathcal{U} \times \mathbb{T}^3$, see (61) for the definition of $\mathcal{U}$, and Theorem 3 for the properties of H. In this section we prove that H has a trajectory whose closure contains $\mathcal{F}^\infty$.

Let $\{\varepsilon_n\}_n$ be a monotone decreasing sequence of positive numbers such that $\varepsilon_n \to 0$ with n fast enough. There are several reasons for ε_n's to decay fast enough.

- We would like the supports of the localized perturbations to be pairwise disjoint and ε_n-dependent neighborhoods of F_n to be disjoint from F_{n+j} for any $|j| \geq 2,\ j \geq -n$. Recall that the non-localized perturbation of the original Lagrangian has the form: (18,20) for a single resonance and (38,39) for a double resonance.
- We also need to "fit in" the lenses into a fairly dense grid $\Gamma_{\bar{n}}$.
- We need ε_n to be small enough so that our Lemmas 4 and 8 hold.

Then for each n we shall find $N = N(n)$ (which stands for the number of lenses needed to shadow F_n) and construct a trajectory $\{\gamma^n(t) = (\theta^n, I^n)(t)\}_t$ whose velocity ε_k-approximates F_k for each $1 \leq k \leq n$, i.e., an ε_k-neighborhood of $\{I^n(t)\}_{t\in\mathbb{R}}$ contains F_k.

As $n \to \infty$ and $N = N(n) \to \infty$ we get a sequence of trajectories $\{\gamma^n(t)\}_t$ approximating F^∞ with better and better precision. By construction all $\omega \in F^\infty$, aside from an at most countable set, are totally rational, and dynamics on the corresponding 3-dimensional torus $\mathbb{T}^3_\omega$ is integrable. This implies that if $\{I^n(t)\}_{t\in\mathbb{R}}$ approximates $F^\infty \subset \mathbb{S}^2 \subset \mathbb{R}^3$ with better and better precision, then $\{\gamma^n(t)\}_{t\in\mathbb{R}}$ approximates $\mathcal{F}^\infty = \mathbb{T}^3 \times F^\infty$ with better and better precision.

Variational problems are constructed in such a way that each $\{\gamma^n(t)\}_t$ starts inside of a lens around the origin and its velocity belongs to a compact ball. Thus, we can find a converging subsequence. The limiting trajectory will be the one proving Theorem 1.

To define parameters for the variational problem we need to map from action variables to velocity and back. Let H^{int} be the Hamiltonian constructed in Sect. 7.2. Let $\mathcal{L}^{int} : (\theta, I) \to (\theta, \dot\theta = \partial_I H^{int})$ be the Legendre map associated with H^{int}.

We emphasize that *by construction our Hamiltonian is close to the sum of squares Hamiltonian* $H_0(I) = |I|^2/2$. In the coordinates that we study this Hamiltonian, we have three types of zones:

- close to a double resonance (for actions in $V_{\bar n}$), where the Hamiltonian has the form (59–60),
- close to a single resonance (for actions in $W_{\bar n}$), where by Lemma 11 it is close $H_0(I) - \bar\varepsilon \cos^2 l(\theta)$, where $l(\theta)$ is a linear periodic function of θ, or
- exactly $H_0(I)$.

In all cases the Hamiltonian is convex.

In each of the above cases, Lemma 6 or 10 defines a constant $\sigma_{\bar n}$.

Denote $\delta_n = \min\{\varepsilon_n, \varkappa_n\}$, where $\varkappa_n$ is $\varkappa = \varkappa(\varepsilon)$ from Lemmas 5 and 9 respectively.

Denote $F^{\mathcal{L}} = \mathcal{L}^{int}(F)$ and $F_n^{\mathcal{L}} = \mathcal{L}^{int}(F_n) \subset \mathbb{R}^3$ images of the action fractal F and F_n's in the space of velocities. Since $\dot\theta = \partial_I H^{int}(I)$ is invertible, in what follows we identify F and $F^{\mathcal{L}}$.

8. Shadowing and Construction of Variational Problems

8.1. From a unit vector "of average velocity" v to $\Delta\mathbf{n}$, c, T, type of a resonance, and type of a grid of lenses. In Theorem 3 we have defined the fractal set F, the Hamiltonian $H = H^{int} + \tilde\beta$ in its neighborhood, open sets $V_{\bar n}$ and $W_{\bar n}$ corresponding to single and double resonance, and the corresponding families of lenses.

Suppose we have a unit vector v (think of it as the approximate average velocity of a trajectory of H that we have to find). Suppose v is such that $v = \frac{\partial H}{\partial I}(\theta, I)$ for some I such that $I \in V_{\bar n}$ or $I \in W_{\bar n}$ for some $\bar n$. Theorem 3 gives us the family of lenses corresponding to this neighborhood $V_{\bar n}$ or $W_{\bar n}$. Fix $\mathbf{n}$ and $\mathbf{n}'$—centers of two lenses in this family such that $\Delta\mathbf{n} = \mathbf{n} - \mathbf{n}'$ is large enough and $w = \frac{\Delta\mathbf{n}}{|\Delta\mathbf{n}|}$ is close enough to v (the quantitative part will be made precise later).

For the two centers of lenses, $\mathbf{n}$ and $\mathbf{n}'$, consider the minimizer $\gamma^*(t)$ of the integrable system connecting them and being on the energy surface $\{H^{\mathrm{int}}\} = 1/2$. This minimizer is c-static for some $c = c(\Delta\mathbf{n})$. Denote by $T = T(\Delta\mathbf{n})$ the corresponding time.

Remark 3. Notice that if $\Delta\mathbf{n} = \mathbf{n} - \mathbf{n}'$ is sufficiently large, then the minimizer $\gamma^*(t)$ (which is a piece of a trajectory of the integrable system) is at least $4\sqrt{\varepsilon}_n$-dense on the corresponding invariant torus. This is due to the fact that in the local coordinates $(\theta_{\bar n}, I_{\bar n})$, the $z_{\bar n}$-component of the speed is less or equal to $2\sqrt{\varepsilon_n}$, while the $(x_{\bar n}, y_{\bar n})$-component of the speed is one. We assume that in the following construction $\Delta\mathbf{n}$ are chosen large enough to provide this density.

8.2. Setting up a variational problem. Fix $n \in \mathbb{Z}_+$. The n^{th} order tree F_n consists of a finite collection of resonant segments. Define a sufficiently dense set of points in F_n (which we shall shadow later) as follows: it is δ_0^{3r}-dense on F_1, δ_1^{3r}-dense on $F_2 \setminus F_1, \dots$, and finally, δ_n^{3r}-dense on $F_n \setminus F_{n-1}$. For some $N = N(n)$ select an ordered collection of vectors $\mathcal{V}_n = \{v_j\}_{1 \le j \le N} \subset \mathbb{S}^2$ such that $v_0 = (1, 0, 0)$ and if $v_j \in F_k \setminus F_{k-1}$ for some k, then $|v_j - v_{j+1}| < \delta_k^{3r}$. Moreover, the $2\delta_k^{3r}$-neighborhood of $\mathcal{V}_n$ contains $F_k \setminus F_{k-1}$.

Let $\mathbf{n}_0 = (0, 0, 0)$ and $\mathbf{n}_1 = (\Delta n_x, 0, 2)$ with $\Delta n_x > \delta_0^{-2r-5}$. Based on the algorithm from the previous section we determine a collection of vectors $\{\mathbf{n}_j\}_{0\le j\le N}$ such that $\mathbf{n}_j$ is a center of a certain lens, $\Delta\mathbf{n}_j = \mathbf{n}_{j+1} - \mathbf{n}_j$ is large and directions $\omega_j = \Delta\mathbf{n}_j/|\Delta\mathbf{n}_j|$ and $\omega_{j-1} = \Delta\mathbf{n}_{j-1}/|\Delta\mathbf{n}_{j-1}|$ are close enough. Moreover, for some $1 \le k \le n$ we have that $|\omega_j - v_j| \le 2\delta_k^{3r}$, where $v_j \in \mathcal{V}_n$.

Now, based on the grid $\mathcal{V}_n$ inside F_n, we choose a collection of centers of lenses $\{\mathbf{n}_j\}_{0\le j\le M}$, (each $\mathbf{n}_j$ is a center of the lens $\mathcal{L}_j := \mathcal{L}_{\mathbf{n}_j}$). Based on the algorithm from Sect. 8.1 we determine sequences of vectors $c_j = c(\Delta n_j)$ and time durations $T_j = T(\Delta n_j)$. Based on c_j and the corresponding Hamiltonian $H = H(\Delta\mathbf{n}_j)$ we determine a closed one-form $\eta_j = \eta_{c_j}$ and $\alpha_j = \alpha(c_j)$. This allows us to define the c-Lagrangian $L_{c_j}(\theta, \dot\theta) = L(\theta, \dot\theta) - \eta_c(\theta)\cdot\dot\theta + \alpha(c)$ as in (22) and (42) for single and double resonances respectively. Recall that $A_{c_j}^{\tau_j}(\theta_j, \theta_{j+1}) = \inf_\gamma \int_0^{\tau_j} L_{c_j}(\gamma(t), \dot\gamma(t))\, dt$ over C^1 curves starting at θ_j at time 0 and ending at θ_{j+1} at time τ_j.

We assume that the sequence c_j for $j = -M, \dots, M$ is such that the corresponding sequence of actions lies in the consecutive neighborhoods: $W_{-N}, V_{-N}, W_{-N+1}, \dots V_{N-1}, W_N$ for $N = N(M)$.

Finally, for each V_j and W_j, there exists a corresponding σ_j from Lemma 6 and 10, respectively.

Recall that we defined in (24):

$$S_j = S(\mathbf{n}_j, w_j) = \{\theta : \theta \in \mathcal{L}_j,\ (\theta \mod 2)\cdot w_j = 0\}$$

be the 2-dimensional disk concentric with $\mathcal{L}_j$ of the same radius as $\mathcal{L}_j$. Denote

$$\mathcal{I}_M = \{(c_j, \mathbf{n}_j, \sigma_j, S_j, T_j), -M \le j \le M\}$$

and

$$Q_M = \{(\theta, \tau) = (\theta_{-M}, \dots, \theta_M, \tau_{-M}, \dots, \tau_M) :\ \theta_j \in S_j, |\tau_j - T_j| \le \sigma_j\}.$$

For a fixed sequence of times $\tau = (\tau_j)_{j=-M}^M$, we consider the following *preliminary* variational problem:

$$\mathcal{A}_c(\theta, \tau) = \sum_{j=-M}^{M} A_{c_j}^{\tau_j}(\theta_j, \theta_{j+1}), \tag{64}$$

where we minimize inside the hypercube Q_M.

Lemma 15. *Fix $M < \infty$. Let the set $\mathcal{I}_M$ satisfy the conditions above. Then the minimum*

$$\mathcal{M}_{\mathcal{I}_M} = \min_{Q_M} \mathcal{A}_c(\theta, T)$$

is attained in the interior of the hypercube Q_M. The value of $\mathcal{M}_{\mathcal{I}_M}$ is positive and finite. Moreover, there is a shadowing trajectory γ of the Euler-Lagrange flow of (4) such that $\gamma(t_j)$ passes through the sections S_j, $j = -M, \dots, M$, and the value of the Hamiltonian H_ε on this trajectory is close to $\frac{1}{2}$.

Proof. First of all we justify the existence of a solution to (64). As we proved in Sects. 4 and 5, the action $A_{c_j}^{\tau_j}(\theta_j, \theta_{j+1})$ is continuous in c_j, τ_j and θ's. Since the hypercube Q_M is compact, a minimum is attained. Consider a "two leg variational problem":

$$\min_{\theta_j \in S_j,\ |\tau_{j-1}-T_{j-1}|<\sigma_{j-1},\ |\tau_j - T_j|<\sigma_j,} A_{c_{j-1}}^{\tau_{j-1}}(\theta_{j-1}, \theta_j) + A_{c_j}^{\tau_j}(\theta_j, \theta_{j+1}),$$

where $\theta_{j-1} \in S_{j-1}$ and $\theta_{j+1} \in S_{j+1}$ are fixed. In suffices to show that the minimum with respect to θ_j, τ_j and τ_{j+1} is interior. Consider two cases depending on whether c_j corresponds to the single or double resonance. □

The single resonance case. Suppose that c_j corresponds to $I \in V_{\bar{n}}$ for some $\bar{n}$. By Lemma 1, we have smooth dependence of c on direction $\omega = \Delta\mathbf{n}/|\Delta\mathbf{n}|$. We choose $\Delta\mathbf{n}_j$ so that

- Both $\Delta\mathbf{n}_j$ and the corresponding T_j are large enough to apply Lemmas 5 and 6.
- Directions $\omega_j = \Delta\mathbf{n}_j/|\Delta\mathbf{n}_j|$ and $\omega_{j-1} = \Delta\mathbf{n}_{j-1}/|\Delta\mathbf{n}_{j-1}|$ are close enough so that the corresponding c_j and c_{j-1} are at most $\varkappa_j/2$ close, where κ_j is from Lemma 5.

The interior property of the minimizer follows from Lemmas 5 and 6

The double resonance case. As in the previous case, c_{j-1} can be determined by $\Delta\mathbf{n}_{j-1}$. But there is a discontinuity, discussed in Sect. 5, as we approach a double resonance locally written $\dot{y} = \dot{z} = 0$. This case subdivides into two sub-cases:

a) bring h_y down to zero and increase h_z away from zero (h_y and h_z are energies of the corresponding components).
b) change sign of h_y, e.g., from positive to negative keeping h_z practically zero;

Consider the case a). In the limit as $h_y \to 0^+$ from above with $\Delta\mathbf{n}_z = 2$ the limiting value of c is (c_x, c_y^+, c_z^+). Similarly the limit of $h_z \to 0^+$ from above with $\Delta\mathbf{n}_y = 2$ the limiting value of c is (c_x, c_y^+, c_z^+). Thus, there is no discontinuity and we proceed in the same way as in the single resonance:

- $\Delta\mathbf{n}_j$ is large enough for application of Lemmas 9 and 10.
- directions $\omega_j = \Delta\mathbf{n}_j/|\Delta\mathbf{n}_j|$ and $\omega_{j-1} = \Delta\mathbf{n}_{j-1}/|\Delta\mathbf{n}_{j-1}|$ are close enough so that the corresponding c_j and c_{j-1} are at most $\varkappa/2$ close.

Consider the case b). We have a discontinuity: in the limit as $h_y \to 0^+$ from above with $\Delta\mathbf{n}_z = 2$ the limiting value of c is (c_x, c_y^+, c_z^+), while in the limit as $h_y \to 0^-$ from below with $\Delta\mathbf{n}_z = 2$ the limiting value of c is $(c_x, -c_y^+, c_z^+)$ (see Sect. 5.1). There are various ways to design transition through a double resonance and overcome discontinuity in c. We choose the one of slow varying c.

As long as $\Delta\mathbf{n}_j$ has its y-component $\Delta\mathbf{n}_j^y > 2$, we proceed as in previous cases. Decreasing Δn^y component to 2 corresponds to reducing the c_y-component of c to c_y^+ (see Sect. 5.1). Suppose $\Delta\mathbf{n}_j$ is such that its $\Delta\mathbf{n}_j^y = 2$. We select a repeated collection of $\Delta\mathbf{n}$:

$$\Delta\mathbf{n}_j = \Delta\mathbf{n}_{j+1} = \cdots = \Delta\mathbf{n}_{j+k-1}, \quad \Delta\mathbf{n}_{j+k} = (\Delta n_x, 0, 2),$$
$$\Delta\mathbf{n}_{j+k+1} = \Delta\mathbf{n}_{j+k+1} = \cdots = \Delta\mathbf{n}_{j+2k-1}, \quad \text{with } \ \Delta n^y_{j+k+1} = -2,$$

for some integer k such that $\rho = [3c_y^+/k] < \varkappa$ with $[\cdot]$ being the integer part. Then we select a collection of c's as follows

$$c_j = (c_x, c_y^+, c_z^+), \quad c_{j+s} = (c_x, c_y^+ - s\rho/3, c_z^+), \quad \text{for } \ s < k, \quad c_{j+k} = (c_x, 0, c_z^+),$$
$$c_{j+k+s} = (c_x, s\rho/3, c_z^+), \quad \text{for } \ s < k, \quad c_{j+2k} = (c_x, -c_y^+, c_z^+).$$

In order to be able to apply Lemmas 9 and 10 we do the following trick. We need to prove that the sum of actions (9) has an interior minimum for $\mathcal{L}_{i-1}$, $\mathcal{L}_i$, and $\mathcal{L}_{i+1}$ satisfying the above conditions. Recall that in (41) we define a closed one form η_c for c's with $c_y \geq c_y^+$. Define then $\eta_j = \eta_{c_j}$. Transform the collection of c_{j+s}, $0 < s < k$, into a collection of closed one-forms $\Delta\eta_{j+s}$, $0 < s < k$ such that $\eta_{c_{j+s+1}} = \eta_{c_j} + \Delta\eta_{j+s}$, and the cohomology class $[\eta_{c_{j+s}}]_{\mathbb{T}^3}$ equals c_{j+s}. Due to (41), each $\Delta\eta_{j+s}$ is a constant one-form with only second non-zero component equal to $s\rho/3$. This implies that the minimizers of the sum

$$\int_0^{T_{j+s}} L_{c_j}(\gamma_{j+s}, \dot{\gamma}_{j+s})\, dt + \int_0^{T_{j+s+1}} L_{c_{j+1}}(\gamma_{j+s+1}, \dot{\gamma}_{j+s+1})\, dt$$

and the minimizers of a similar sum with Lagrangians $L_{c_j}(\gamma_{j+s}, \dot{\gamma}_{j+s})$ and $L_{c_{j+1}}(\gamma_{j+s}, \dot{\gamma}_{j+s})$ replaced by $L_{c_{j+s}}(\gamma_{j+s}, \dot{\gamma}_{j+s})$ and $L_{c_{j+s+1}}(\gamma_{j+s}, \dot{\gamma}_{j+s})$, respectively, are the same. Indeed, dependence on the intermediate point θ_{j+s} disappears. This implies that Lemma 9 applies for each $j+s$ with $0 < s < k$. The construction for $k < s < 2k$ is similar as we have symmetry $c_y \to -c_y$. In the case $k = s$ and $\Delta n_y = 0$, we use u_0^+ and u_0^- as y-component of u (see the last sentence of Lemma 9).

This implies the interior minimum of (64), and proves the first part of the lemma. Now we prove the existence of the shadowing trajectory. The curve $\gamma(t)$ which corresponds to the minimizer of the c-action (64), satisfies the Euler-Lagrange equation on each segment (τ_j^*, τ_{j+1}^*). However, its one-sided derivatives do not have to match at the endpoints $t = \tau_j^*$ of the neighboring intervals. We will show that in the vicinity of this pseudo-solution there exists a true solution. To prove this, we need to modify the variational problem (64).

The idea of this modification is a fairly standard tool (see e.g. [KL1]). A part of the proof of the present lemma is contained in [KL1] verbatim, and we chose to make a precise reference rather than rewrite it here.

Let us modify the Lagrangians L_{c_j} and $L_{c_{j+1}}$ into L_{η_j} and $L_{\eta_{j+1}}$ in such a way that the new Lagrangians match inside of connecting lenses, and at the same time we still have an interior minimum.

Consider exact one-forms $\Delta\eta_j : T\mathbb{R}^3 \to \mathbb{R}$,

$$\Delta\eta_j(\theta) = \begin{cases} 0 & \text{for } |\mathbf{n}_j - \theta| \geq r_j \\ \eta_{c_{j+1}}(\theta) - \eta_{c_j}(\theta) & \text{for } |\mathbf{n}_j - \theta| \leq 0.5\, r_j \\ (1 - \mu(|\mathbf{n}_j - \theta|))(\eta_{c_{j+1}}(\theta) - \eta_{c_j}(\theta)) & \text{for } 0.5\, r_j \leq |\mathbf{n}_j - \theta| \leq 0.8\, r_j \end{cases} \tag{65}$$

where μ is a smooth nondecreasing function whose support is $(0.5, \infty)$ and which is identically one on $(0.8, \infty)$, values r_j are determined from the size of lenses. Define a function b_j such that $b_j(\mathbf{n}_j) = 0$ and $\nabla b_j = \eta_j + \Delta\eta_j := \tilde{\eta}_j$. Now we modify

$$L_{\tilde{\eta}_j}(\gamma_{j+1}, \dot{\gamma}_{j+1}) = L_{c_j}(\gamma_{j+1}, \dot{\gamma}_{j+1}) + \Delta\eta_j(\gamma_{j+1}) \cdot (\dot{\gamma}_{j+1}).$$

Then instead of minimizing (64) we shall minimize

$$\min_{\theta_j \in S_j,\ |\tau_j - T_j| < N} \sum_{1 \leq j \leq 10} \int_0^{\tau_j} L_{\tilde{\eta}_j}(\gamma(t), \dot{\gamma}(t))\, dt. \tag{66}$$

This leads to comparing

$$\int_0^{T_j} L_{\tilde\eta_j}(\gamma_j,\dot\gamma_j)dt + \int_0^{T_{j+1}} L_{\tilde\eta_{j+1}}(\gamma_{j+1},\dot\gamma_{j+1})\,dt$$
$$= \int_0^{T_j} L_0(\gamma_j,\dot\gamma_j)\,dt + \int_0^{T_{j+1}} L_0(\gamma_{j+1},\dot\gamma_{j+1})\,dt$$
$$+(b_{j+1}(\theta_{j+1}) - b_{j+1}(\theta_j)) + (b_j(\theta_j) - b_j(\theta_{j-1})).$$

By construction b-terms depending on θ_j cancel, thus, the second expression is independent of θ_j and the minimum of the first line and the total sum in θ_j is the same.

The above relation between the two actions implies that an interior minimum of the action (66) corresponds to a solution of the Euler-Lagrange equation. This implication is proved in [KL1], p. 423.

It is left to verify that depth of the corresponding local minima of (64), claimed by Lemma 15, survives this modification. Notice that the local minima are obtained using Lemmas 5 and 9.

The difference between ω_{j+1} and ω_j is bounded by $2\delta_k^{3r}$. the dependence of c on ω is analytic for single resonance (see Sect. 4.1, Lemma 1). Problem of the regularity of the dependence of c on ω for the double resonance (see Sect. 5.1) reduces to that of the dependence of c on ω for the pendulum. Express both of them in terms of the energy h of the pendulum: $c = c(h)$ and $\omega = \omega(h)$. Then $\partial_\omega c = \partial_h c/\partial_h\omega$. Using the calculations of the Appendix we have that

$$\partial_h c \sim \mathcal{T}(h) \qquad \text{and} \qquad \partial_h\omega = \left(\frac{1}{\mathcal{T}(h)}\right)' = -\frac{\mathcal{T}'(h)}{\mathcal{T}^2(h)}.$$

Since $\mathcal{T}(h) \sim -2^{-3/2}(\ln h + 0.5\,\ln\varepsilon)/\sqrt{\varepsilon}$ and $\mathcal{T}'(h) \sim \sqrt{2}/(\sqrt{\pi}h\sqrt{\varepsilon})$, $\partial_\omega c \to 0$ as $\omega \to 0$. Thus, for some $C > 0$ independent of ε, the bound $|\omega_{j+1} - \omega_j| < 2\delta_k^{3r}$ implies $|c_{j+1} - c_j| < C\delta_k^{3r}$.

To see the regularity of the dependence of η_c on c for the single resonance we use Lemma 1, and for the double resonance use explicit formulas in Example 2, Sect. 3.3 and (41). This shows that the difference $\Delta\eta_j$ is $C\delta_k^{3r}$-small (see (65)).

This, in its turn, shows that difference between L_{η_j} and L_{c_j} is $C\delta_k^{3r}$-small. Thus, if each pair of neighbors in the sum (64) minimum

$$A_{c_{j-1}}^{\tau_{j-1}}(\theta_{j-1},\theta_j) + A_{c_j}^{\tau_j}(\theta_j,\theta_{j+1})$$

has θ-depth $\varepsilon_j^{2k+2}/2$. Thus, if approximation L_{η_j} of L_{c_j} is $C\delta_k^{3r} \le C\varepsilon_k^{3r}$-small, then

$$A_{\eta_{j-1}}^{\tau_{j-1}}(\theta_{j-1},\theta_j) + A_{\eta_j}^{\tau_j}(\theta_j,\theta_{j+1})$$

also has an inner minimum. For completion of arguments see [KL1], p. 423–424.

8.3. Trajectory passing through infinitely many lenses. Let γ_m be the trajectory that passes through lenses $\mathcal{L}_{-m},\dots\mathcal{L}_m$, whose existence and ordering was established above. Consider a sequence $(\gamma_m)_{m=1}^\infty$ of such trajectories. In the lens $\mathcal{L}_0$ consider the sequence of points $p_m = \gamma_m \cap S_0$. Let p^* be a limit point of this sequence which exists by compactness. Recall that all the constructed trajectories have energies H close to $\frac{1}{2}$

(say, $|H(\gamma_m) - \frac{1}{2}| \leq 0.1$). Therefore, the corresponding space of velocities is compact. Hence, there exists a subsequence of $(\gamma_m)_{m=1}^{\infty}$ such that for each γ in this subsequence (with an appropriate shift of time) we have: $\gamma(0) = p^*$, $\dot{\gamma}(0) = v^*$. By construction, the trajectory $\gamma(t)$ of the Euler-Lagrange flow with these initial conditions will pass through all the lenses.

8.4. Density in a set of Hausdorff dimension 5. In Sect. 8.2 we have chosen a sequence of centers of lenses in such a way that the trajectory of the integrable system between each two lenses on the fixed energy surface fills the corresponding invariant torus, call it $\mathbb{T}^3(\Delta\mathbf{n})$, $4\sqrt{\varepsilon}_n$-densely (where n corresponds to the segment of the H-fractal); see Remark 3. Of course, the deformed system H may not have an invariant torus. But the piece of the trajectory of H through the same lenses, call it γ_H is close to this torus, in the sense that this torus lies in the $8\sqrt{\varepsilon}_n$-neighborhood of γ_H.

Now, fix any $\delta > 0$ and let $I \in F^{\infty}$. We shall show that there is a piece of the trajectory γ, constructed above, such that the torus $I \times \mathbb{T}^3$ lies in the δ-neighborhood of this piece. Since $I \in \overline{F} \setminus F$, there is a segment of F, call it $l_{\bar{n}}$, such that $l_{\bar{n}}$ lies in the $\delta/2$-neighborhood of I, and $\varepsilon_n < \delta^2/200$. Continuity arguments complete the proof.

8.5. Proof of Theorem 2. The proof of Theorem 2 is quite similar, and we do not present a detailed account of it. The only difference is the following. The projection of the two chosen invariant tori (i.e., tori with totally irrational rotation numbers $\vec{\omega}'$ and ω'') to the energy surface $\{H_\varepsilon = 1/2\}$ are two points, I' and I''. We construct a connected "path" consisting of a countable number of segments, each segment being the intersection of the energy surface $\{H_\varepsilon = 1/2\}$ with an appropriate resonant plain passing through the origin. This "path" converges to I' in one direction, and to I'' in the other direction. The Hamiltonian H_1 vanishes on the end points of the "path". Modulo this difference, the construction of H_1 is the same as in the proof of Theorem 1.

Acknowledgements. The first author acknowledges useful discussions with Ke Zhang. V. K. was supported by NSF grant DMS-1101510. We would like to thank the referee for his thorough reading of the manuscript and many useful remarks.

9. Appendix: Auxiliary Pendulum Bounds

Here we prove the asymptotic formulas for the period of the standard pendulum that were used above. Consider the standard pendulum with one degree of freedom,[4]

$$h_z = \frac{\dot{z}^2}{2} - \varepsilon \cos^2 \frac{\pi z}{2}.$$

For all $h_z > 0$ trajectories are periodic with period, denoted by $\mathcal{T}(h_z)$, given by

$$\mathcal{T}(h_z) = \int_0^2 \frac{dz}{\sqrt{2(h_z + \varepsilon \cos^2 \frac{\pi z}{2})}}. \tag{67}$$

[4] One could get rid of the coefficient ε by $\sqrt{\varepsilon}$ time rescaling.

For small h_z/ε we have the following asymptotic behavior of $\mathcal{T}(h_z)$:

$$\mathcal{T}(h_z) = -\frac{\ln c_1 \frac{h_z}{\varepsilon}}{c_0\sqrt{\varepsilon}}(1+o(1)), \tag{68}$$

where $c_0 = 2^{3/2}$, $c_1 = 4/\pi^2$. The asymptotic behavior of $\mathcal{T}'(h_z)$ for small h_z/ε is given by

$$\mathcal{T}'(h_z) = \frac{1}{c_2 h_z\sqrt{\varepsilon}}(1+o(1)), \tag{69}$$

where $c_2 = \sqrt{\pi}/\sqrt{\varepsilon}$.

In order to verify (68), notice that for small h_z the main contribution comes from a neighborhood of $z = 1$. There one can approximate $\cos^2 \frac{\pi z}{2} \simeq \left(\frac{\pi(z-1)}{2}\right)^2$. Recall the table integral:

$$\int \frac{dx}{\sqrt{x^2+a^2}} = \ln\left(x + \sqrt{x^2+a^2}\right).$$

For small a integrating over $[-1, 1]$ we get $\ln \frac{\sqrt{1+a^2}+1}{\sqrt{1+a^2}-1} \sim -\frac{\ln a}{2}(1+o(1))$. Thus, denoting $\frac{2\sqrt{h_z}}{\pi\sqrt{\varepsilon}}$ by a, and making the change of variable $x = (z-1)$, we see that the dominant part of $\mathcal{T}(h_z)$ is $-\frac{\ln \frac{4h_z}{\pi^2\varepsilon}}{2^{3/2}\pi\sqrt{\varepsilon}}$. Since the contribution outside of $[-1, 1]$ can only increase the constant in front of the leading term, we get (68). Here we verify (69). The derivative of $\mathcal{T}(h_z)$ is

$$\mathcal{T}'(h_z) = -\int_0^2 \frac{dz}{\sqrt{2(h_z + \varepsilon\cos^2\frac{\pi z}{2})}^{\,3}}.$$

Again the main contribution comes from a neighborhood of $z = 1$ and we can approximate $\cos^2 \frac{\pi z}{2} \simeq \left(\frac{\pi(z-1)}{2}\right)^2$. Recall the table integrals

$$\int \frac{dx}{\sqrt{a^2+x^2}^{\,3}} = \frac{x}{a^2\sqrt{x^2+a^2}}.$$

For small a integrating over $[-1, 1]$ we get $\frac{2}{a^2\sqrt{1+a^2}} \sim \frac{2}{a^2}(1+o(1))$. Again, setting $a = \frac{2\sqrt{h_z}}{\pi\sqrt{\varepsilon}}$ and $x = (z-1)$, we see that the dominant part of $\mathcal{T}'(h_z)$ is $\frac{\sqrt{2}}{\sqrt{\pi}h_z\sqrt{\varepsilon}}$. Since the contribution outside of $[-1, 1]$ can only increase the constant in front of the leading term, we get (69).

References

[Ar] Arnold, V.: Instabilities in dynamical systems with several degrees of freedom. Sov. Math. Dokl. **5**, 581–585 (1964)

[Ar1] Arnold., V.: A stability problem and ergodic properties of classical dynamical systems (Russian). In: *Proceedings of the International Congress of Mathematicians*, (Moscow, 1966). Berlin: Intel. Math. Union, 1968, pp. 387–392

[Ar2] Arnold, V.: *Mathematical methods of classical mechanics,* 2^{nd} edition, Graduate Texts in Mathematics, **60**, Berlin-Heidleberg-New York: Springer, 1989
[Be] Bernard, P.: The dynamics of pseudographs in convex hamiltonian systems. J. Amer. Math. Soc. **21**(3), 615–669 (2008)
[Be2] Bernard, P.: Symplectic aspects of mather theory. Duke Math. J. **136**((3), 401–420 (2007)
[Bs1] Bessi, U.: An approach to arnold's diffusion through the calculus of vartiations. Nonlinear Anal. **26**(6), 1115–1135 (1996)
[Bs2] Bessi, U.: Arnold's diffusion with two resonances. J. Diff. Eqs. **137**(2), 211–239 (1997)
[Bi] Birkhoff, G.D.: *Collected Math Papers.* Vol. **2**, New York: Dover, 1968, p. 462–465
[Bo1] Bounemoura, A.: An example of instability in high-dimensional hamiltonian systems. International Mathematical Research Notices **2012**(3), 685–716 (2012)
[Bo2] Bounemoura, A.: *Stabilité et instabilité des systèmes hamiltoniens presque-intérables.* 2010, Thèse de doctorat de l'Université Paris Sud
[BK] Bourgain, J., Kaloshin, V.: On diffusion in high-dimensional hamiltonian systems. J. Funct. Anal. **229**(1), 1–61 (2005)
[CY] Cheng, C.-Q., Yan, J.: Existence of diffusion orbits in a priori unstable hamiltonian systems. J. Diff. Geom. **67**(3), 457–517 (2004)
[CI] Contreras G., Itturiaga R.: Global Minimizers of Autonomous Lagrangians, CIMAT, Mexico, 2000
[DLS] Delshams, A., de la Llave, R., Seara, T.: A geometric mechanism for diffusion in hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model. Mem. Amer. Math. Soc. **179**(844), 141 (2006)
[Do] Douady, R.: Stabilité ou instabilité des points fixes elliptiques. (french) [stability or instability of elliptic fixed points]. Ann. Sci. Scole Norm. Sup. (4) **21**(1), 1–46 (1988)
[E] Ehrenfest, P.T.: *The Conceptual Foundations of the Statistical Approach in Mechanics.* Ithaca, NY: Cornell University Press, 1959
[Fal] Falconer, K.: *Fractal Geometry.* New York: Wiley, 2003
[Fa] Fathi, A.: *Weak KAM theory.* Cambridge Studies in Advanced Mathematics, Cambridge: Cambridge University Press
[Fa1] Fathi, A.: Théorème kam faible et théorie de mather sur les systèmes lagrangiens. (french) [a weak kam theorem and mather's theory of lagrangian systems]. C. R. Acad. Sci. Paris Sèr. I Math. **324**(9), 1043–1046 (1997)
[FM] Fathi, A., Mather, J.: Failure of convergence of the lax-oleinik semigroup in the time periodic case. Bull. Soc. Math. de France **128**, 473–483 (2000)
[Fe] Fermi, E.: Dimonstrazione che in generale un sistema mecanico e' quasi ergodico. Nuovo Cimento **25**, 267–269 (1923)
[G] Gallavotti, G.: Fermi and Ergodic problem. Preprint 2001
[GL] Gidea, M., de la Llave, R.: Topological methods in the instability problem of hamiltonian systems. Discrete Contin. Dyn. Syst. **14**(2), 295–328 (2006)
[H] Herman, M.: Some open problems in dynamics. In: *Proceedings of the International Congress of Mathematicians*, Vol. II (Berlin, 1998). Doc. Math. 1998, Extra Vol. II, Berlin: Int. Math. Union, 1998, pp. 797–808
[KL1] Kaloshin, V., Levi, M.: An example of arnold diffusion for near-integrable hamiltonians. Bull. Amer. Math. Soc. (N.S.) **45**(3), 409–427 (2008)
[KL2] Kaloshin, V., Levi, M.: Geometry of arnold diffusion. SIAM Rev. **50**(4), 702–720 (2008)
[KLS] Kaloshin, V., Levi, M., Saprykina, M.: *Arnold diffusion in a pendulum lattice*, arXiv:1104.0580v1 [math.DS]
[KMV] Kaloshin, V., Mather, J., Valdinoci, E.: Instability of resonant totally elliptic points of symplectic maps in dimension 4. Analyse complexe, systemes dynamiques, sommabilite des series divergentes et theories galoisiennes. II. Asterisque No. **297**, 79–116 (2004)
[KZZ] Kaloshin, V., Zhang, K., Zheng, Y.: Almost dense orbit on energy surface. In: *Proceedings of the XVIth ICMP*, Prague, River Edge, NJ: World Scientific, 2010, pp. 314–322
[LM] Lochak, P., Marco, J.-P.: Diffusion times and stability exponents for nearly integrable analytic systems. cent. Eur. J. Math. **3**(3), 342–397 (2005)
[MS] Marco, J.-P., Sauzin, D.: Stability and instability for gevrey quasi-convex near-integrable hamiltonian systems. Publ. Math. Inst. Hautes Etudes Sci. No. **96**, 199–275 (2003)
[Ma] Mather, J.: Action minimizing invariant measures for positive definite lagrangian systems. Math. Z. **207**(2), 169–207 (1991)
[Ma1] Mather, J.: Arnold diffusion. I: announcement of results. J. Math. Sci. **124**(5), 5275–5289 (2004)
[Ma2] Mather, J.: Arnold diffusion. II, preprint, 2006, 160pp
[MDS] McDuff, D., Salamon, D.: *Introduction to Symplectic Topology*, 2^{nd} edition, Oxford: Oxford Math. Monogr, 1998
[Moe] Moeckel, R.: Transition tori in the five-body problem. J. Diff. Eqs. **129**, 290–314 (1996)

[T] Treschev, D.: Evolution of slow variables in a priori unstable hamiltonian systems. Nonlinearity **17**(5), 1803–1841 (2004)
[Zha] Zhang, K.: Speed of arnold diffusion for analytic hamiltonians. Inventiones **186**(2), 255–290 (2011)
[Zhe] Zheng, Y.: *Arnold diffusion in the planar Five body problem.* Preprint, 2009

Communicated by G. Gallavotti