Acta Math., 217 (2016), 1–79
DOI: 10.1007/s11511-016-0141-5

Arnold diffusion in arbitrary degrees of freedom and normally hyperbolic invariant cylinders

by

Patrick Bernard
PSL Research University
Paris, France

Vadim Kaloshin
University of Maryland at College Park
College Park, MD, U.S.A.

Ke Zhang
University of Toronto
Toronto, ON, Canada

1. Introduction

On the phase space $\mathbb{T}^n \times B^n$, we consider the Hamiltonian system generated by the C^r time-periodic Hamiltonian

$$H_\varepsilon(\theta, p, t) = H_0(p) + \varepsilon H_1(\theta, p, t), \quad (\theta, p, t) \in \mathbb{T}^n \times B^n \times \mathbb{T},$$

where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, B^n is the unit ball in $\mathbb{R}^n$ around the origin, and $\varepsilon \geqslant 0$ is a small parameter. The equations

$$\dot\theta = \partial_p H_0 + \varepsilon \partial_p H_1 \quad \text{and} \quad \dot p = -\varepsilon \partial_\theta H$$

imply that the momenta p are constant in the case $\varepsilon = 0$. A question of general interest in Hamiltonian dynamics is to understand the evolution of these momenta when $\varepsilon > 0$ is small (see e.g. [1], [2], [3]). In the present paper, we assume that H_0 is convex, and, more precisely,

$$\frac{I}{D} \leqslant \partial_p^2 H_0 \leqslant DI, \tag{1}$$

and prove that a certain form of Arnold's diffusion occur for many perturbations. We assume that $r \geqslant 4$, and denote by $\mathcal{S}^r$ the unit sphere in $C^r(\mathbb{T}^n \times B^n \times \mathbb{T})$.

Theorem 1. *There exist two continuous functions ℓ and ε_0 on $\mathcal{S}^r$, which are positive on an open and dense set $\mathcal{U} \subset \mathcal{S}^r$, and an open and dense subset $\mathcal{V}_1$ of*

$$\mathcal{V} := \{H_0 + \varepsilon H_1 : H_1 \in \mathcal{U} \text{ and } 0 < \varepsilon < \varepsilon_0(H_1)\}$$

such that the following property holds for each Hamiltonian $H\in\mathcal{V}_1$:

There exist an orbit (θ, p) *of* H_ε *and a time* $T\in\mathbb{N}$ *such that*

$$\|p(T)-p(0)\| > \ell(H_1).$$

The key point in this statement is that $\ell(H_1)$ does not depend on $\varepsilon\in]0,\varepsilon_0(H_1)[$. In §1.1, we give a more detailed description of the diffusion path. Moreover, an improved version of the main theorem provides an explicit lower bound on $l(H_1)$ (see Theorem 2.1 and Remark 2.1).

The present work is in large part inspired by the work of Mather [52], [53], [54]. In [52], Mather announced a much stronger version of Arnold diffusion for $n=2$. Our set $\mathcal{V}$ is what Mather called a *cusp residual* set. As in Mather's work, the instability phenomenon thus holds in an open dense subset of a cusp residual set. Our result is, however, quite different. We obtain a much more restricted form of instability, which holds for any $n\geqslant 2$. The restricted character of the diffusion comes from the fact that we do not really solve the problem of double resonance (but only finitely many, independent from ε, double resonances are really problematic). The proof of Mather's result is partially written (see [53]), and he has given lectures about some parts of the proof [54].(1)

The study of Arnold diffusion was initiated by the seminal paper of Arnold, [1], where he describes a diffusion phenomenon on a specific example involving two independent perturbations. A lot of work has then been devoted to describe more general situations where similar constructions could be achieved. A unifying aspect of all these situations is the presence of a normally hyperbolic cylinder, as was understood in [57] and [29]; see also [27], [28], [61], [62], [23], [24], [8]. These general classes of situations have been referred to as *a-priori unstable*.

The Hamiltonian H_ε studied here is, on the contrary, called *a-priori stable*, because no hyperbolic structure is present in the unperturbed system H_0. Our method will, however, rely on the existence of a normally hyperbolic invariant cylinder. The novelty here thus consists in proving that a-priori unstable methods do apply in the a-priori stable case. Application of normal forms to construct normally 3-dimensional hyperbolic invariant cylinders in a-priori stable situation in 3 degrees of freedom had already been discussed in [45] and in [48]. The existence of normally hyperbolic cylinders with a length independent from ε in the a-priori stable case, in arbitrary dimension, have been proved in [10], see also [9]. In the present paper, we obtain an explicit lower bound on the length of such a cylinder. The quantity $\ell(H_1)$ in the statement of Theorem 1 is closely related to this lower bound (see also Remark 2.1). Let us mention some additional works of

(1) After a preliminary version of this paper was completed for $n=2$ the problem of double resonance was solved and existence of a strong form of Arnold diffusion is given in [40].

interest around the problem of Arnold's diffusion: [5], [6], [14], [13], [15], [16], [17], [18], [21], [22], [26], [35], [36], [44], [41], [42], [43], [46], [49], [65], [66], [63], and many others.

1.1. Reduction to normal form

As is usual in the theory of instability, we build our unstable orbits around a resonance. A frequency $\omega\in\mathbb{R}^n$ is said *resonant* if there exists $k\in\mathbb{Z}^{n+1}$, $k\neq 0$, such that $k\cdot(\omega,1)=0$. The set of such integral vectors k forms a submodule Λ of $\mathbb{Z}^{n+1}$, and the dimension of this module (which is also the dimension of the vector subspace of $\mathbb{R}^{n+1}$ it generates) is called the *order*, or the *dimension* of the resonant frequency ω.

In order to apply our proof, we have to consider a resonance of order $n-1$ or, equivalently, of codimension 1. For definiteness and simplicity, we choose once and for all to work with the resonance

$$\omega^s = 0,$$

where

$$\omega = (\omega^s, \omega^f) \in \mathbb{R}^{n-1}\times\mathbb{R}.$$

Similarly, we use the notation

$$\theta = (\theta^s, \theta^f) \in \mathbb{T}^{n-1}\times\mathbb{T} \quad \text{and} \quad p = (p^s, p^f) \in \mathbb{R}^{n-1}\times\mathbb{R},$$

which are the slow and fast variables associated with our resonance (see §2 for definitions). More precisely, we will be working around the manifold defined by the equation

$$\partial_{p^s} H_0(p) = 0$$

in the phase space. In view of (1), this equation defines a C^{r-1} curve Γ in $\mathbb{R}^n$, which can also be described parametrically as the graph of a C^{r-1} function $p_*^s(p^f)\colon \mathbb{R}^{n-1}\to\mathbb{R}$. We will also use the notation $p_*(p^f) := (p_*^s(p^f), p^f))$.

We define the *averaged perturbation* Z corresponding to the *resonance* Γ by

$$Z(\theta^s, p) := \iint H_1(\theta^s, p^s, \theta^f, p^f, t)\, d\theta^f\, dt.$$

If the perturbation $H_1(\theta, p, t)$ is expanded as

$$H_1(\theta, p, t) = H_1(\theta^s, \theta^f, p, t) = \sum_{\substack{k^s\in\mathbb{Z}^{n-1}\\ k^f\in\mathbb{Z}\\ l\in\mathbb{Z}}} h_{[k^s, k^f, l]}(p) e^{2\pi i(k^s\cdot\theta^s + k^f\cdot\theta^f + l\cdot t)},$$

then

$$Z(\theta^s,p)=\sum_{k^s} h_{[k^s,0,0]}(p)e^{2\pi i(k^s\cdot\theta^s)}.$$

Our first generic assumption, which defines the set $\mathcal{U}\subset\mathcal{S}^r$ in Theorem 1, is on the shape of Z. We assume that there exists a subarc $\Gamma_1\subset\Gamma$ such that the following hypotesis holds.

Hypothesis 1. There exists a real number $\lambda\in\left]0,\frac{1}{2}\right[$ such that, for each $p\in\Gamma_1$, there exists $\theta^s_*(p)\in\mathbb{T}^{n-1}$ such that the inequality

$$Z(\theta^s,p)\leqslant Z(\theta^s_*(p),p)-\lambda d^2(\theta^s,\theta^s_*(p)) \tag{$HZ\lambda$}$$

holds for each θ^s.

1.2. Single maximum

Thie condition ($HZ\lambda$) implies that, for each $p\in\Gamma_1$, the averaged perturbation $Z(\theta,p)$ has a unique non-degenerate maximum at $\theta^s_*(p)$. In §1.5 we relax this condition and allow bifurcations from one global maximum to a different one. Note that the set of functions $Z\in C^r(\mathbb{T}^{n-1}\times B^n)$ satisfying Hypothesis 1 on some arc $\Gamma_1\subset\Gamma$ is open and dense for each $r\geqslant 2$. As a consequence, we have that the set $\mathcal{U}$ of functions $H_1\in\mathcal{S}^r$ (the unit sphere in $C^r(\mathbb{T}^n\times B^n\times\mathbb{T})$) whose average Z satisfies Hypothesis 1 on some arc $\Gamma_1\subset\Gamma$ is open and dense in $\mathcal{S}^r$ if $r\geqslant 2$.

The general principle of averaging theory is that the dynamics of H_ε is approximated by the dynamics of the averaged Hamiltonian $H_0+\varepsilon Z$ in a neighborhood of $\mathbb{T}^n\times\Gamma$. The applicability of this principle is limited by the presence of additional resonances, that is points $p\in\Gamma$ such that the remaining frequency $\partial_{p^f}H_0$ is rational. Although additional resonances are dense in Γ, only finitely many of them, called *punctures*, are really problematic. More precisely, denoting by $U_{\varepsilon^{1/3}}(\Gamma_1)$ the $\varepsilon^{1/3}$-neighborhood of Γ_1 in B^n and by $\mathcal{R}(\Gamma_1,\varepsilon,\delta)\subset C^r(\mathbb{T}^n\times B^n\times\mathbb{T})$ the set of functions $R(\theta,p,t)\colon\mathbb{T}^n\times B^n\times\mathbb{T}\to\mathbb{R}$ such that

$$\|R\|_{C^2(\mathbb{T}^n\times U_{\varepsilon^{1/3}}(\Gamma_1)\times\mathbb{T})}\leqslant\delta.$$

We will prove in §2 the following result.

PROPOSITION 1.1. *For each* $\delta\in]0,1[$, *there exist a locally finite subset* $\mathcal{P}_\delta\subset\Gamma$ *and* $\varepsilon_1\in]0,\delta[$, *such that the following holds*:

For each compact arc $\Gamma_1\subset\Gamma$ *disjoint from* $\mathcal{P}_\delta$, *each* $H_1\in\mathcal{S}^r$, *and each* $\varepsilon\subset]0,\varepsilon_1[$, *there exists a* C^r*-smooth canonical change of coordinates*

$$\Phi\colon\mathbb{T}^n\times B\times\mathbb{T}\longrightarrow\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$$

satisfying $\|\Phi - \mathrm{id}\|_{C^0} \leqslant \sqrt{\varepsilon}$ *and such that, in the new coordinates, the Hamiltonian* $H_0 + \varepsilon H_1$ *takes the form*

$$N_\varepsilon = H_0(p) + \varepsilon Z(\theta^s, p) + \varepsilon R(\theta, p, t), \tag{2}$$

with $R \in \mathcal{R}(\Gamma_1, \varepsilon, \delta)$.

The key aspects of this result is that the set $\mathcal{P}_\delta$ is locally finite and independent from ε. Because it is essential to have these properties of $\mathcal{P}_\delta$, the conclusions on the smallness of R are not very strong. Yet they are sufficient to obtain the following result.

THEOREM 1.2. *Let us consider the* C^r *Hamiltonian*

$$N_\varepsilon(\theta, p, t) = H_0(p) + \varepsilon Z(\theta^s, p) + \varepsilon R(\theta, p, t), \tag{3}$$

and assume that $\|Z\|_{C^2} \leqslant 1$ *and that* ($HZ\lambda$) *holds on some arc* $\Gamma_1 \subset \Gamma$ *of the form*

$$\Gamma_1 := \{(p_*(p^f)) : p^f \in [a_-, a_+]\}.$$

Then there exist positive constants δ *and* ε_0*, which depends only on* n*,* H_0*, and* λ*, and such that, for each* $\varepsilon \in]0, \varepsilon_0[$*, the following property holds for an open dense subset of functions* $R \in \mathcal{R}(\Gamma_1, \varepsilon, \delta)$ *(for the* C^r *topology):*

There exists an orbit $(\theta(t), p(t))$ *and an integer* $T \in \mathbb{N}$ *such that*

$$\|p(0) - p_*(a_-)\| < \sqrt{\varepsilon} \quad \textit{and} \quad \|p(T) - p_*(a_+)\| < \sqrt{\varepsilon}.$$

1.3. Derivation of Theorem 1 using Proposition 1.1 and Theorem 1.2

Given $l > 0$, we denote by $\mathcal{D}^r(l)$ the set of C^r Hamiltonians with the following property: There exists an orbit $(\theta(t), p(t))$ and an integer T such that $\|p(T) - p(0)\| > l$. The set $\mathcal{D}^r(l)$ is clearly open.

We now prove the existence of a continuous function ε_0 on $\mathcal{S}^r$ which is positive on $\mathcal{U}$ and such that each Hamiltonian $H_\varepsilon = H_0 + \varepsilon H_1$ with $H_1 \in \mathcal{U}$ and $\varepsilon < \varepsilon_0(H_1)$ belongs to the closure of $\mathcal{D}^r(\varepsilon_0(H_1))$.

For each $H_1 \subset \mathcal{U}$, there exist a compact arc $\Gamma_1 \subset \Gamma$ and a number $\lambda \in \left]0, \frac{1}{4}\right[$ such that the corresponding averaged perturbation Z satisfies Hypothesis 1 on Γ_1 with constant 2λ. We then consider the real δ given by Theorem 1.2 (applied with the parameter λ). By possibly reducing the arc Γ_1, we may assume in addition that this arc is disjoint from the set $\mathcal{P}_\delta$ of punctures for this δ. The following properties then hold:

- the averaged perturbation Z satisfies Hypothesis 1 on Γ_1 with a constant $\lambda' > \lambda$;
- the parameter δ is associated with λ by Theorem 1.2;
- the arc Γ_1 is disjoint from the set $\mathcal{P}_\delta$ of punctures.

We say that $(\Gamma_1, \lambda, \delta)$ is a *compatible set of data* if the second and third points above are satisfied. Then, we denote by $\mathcal{U}(\Gamma_1, \lambda, \delta)$ the set of $H_1 \in \mathcal{S}^r$ which satisfy the first point. This is an open set, and we just proved that the union of all compatible sets of data of these open sets covers $\mathcal{U}$.

With each compatible set of data $(\Gamma_1, \lambda, \delta)$ we associate the positive numbers

$$\ell := \tfrac{1}{2}\|p_- - p_+\|,$$

where $p_\pm$ are the extremities of Γ_1, and $\varepsilon_2(\Gamma_1, \lambda, \delta) := \min\{\varepsilon_1, \tfrac{1}{5}\ell^2, \ell\}$, where ε_1 is associated with δ by Proposition 1.1.

Using a partition of unity, we can build a continuous function ε_0 on $\mathcal{S}^r$ which is positive on $\mathcal{U}$ and have the following property: For each $H_1 \in \mathcal{U}$, there exists a compatible set of data $(\Gamma_1, \lambda, \delta)$ such that $H_1 \in \mathcal{U}(\Gamma_1, \lambda, \delta)$ and $\varepsilon_0(H_1) \leqslant \varepsilon_2(\Gamma_1, \lambda, \delta)$.

For this function ε_0, we claim that each Hamiltonian $H_\varepsilon = H_0 + \varepsilon H_1$ with $H_1 \in \mathcal{U}$ and $0 < \varepsilon < \varepsilon_0(H_1)$ belongs to the closure of $\mathcal{D}^r(\varepsilon_0(H_1))$.

Assuming the claim, we finish the proof of Theorem 1. For $l > 0$, let us denote by $\mathcal{V}(l)$ the open set of Hamiltonians of the form $H_0 + \varepsilon H_1$, where $H_1 \in \mathcal{U}$ satisfies $\varepsilon_0(H_1) > l$ and $\varepsilon \in]0, \varepsilon_0(H_1)[$. The claim implies that $\mathcal{D}(l)$ is dense in $\mathcal{V}(l)$ for each $l > 0$. The conclusion of the theorem (with $l(H_1) := \varepsilon_0(H_1)$) then holds with the open set $\mathcal{V}_1 := \bigcup_{l>0} (\mathcal{V}(l) \cap \mathcal{D}(l))$, which is open and dense in $\mathcal{V} = \bigcup_{l>0} \mathcal{V}(l)$.

To prove the claim, let us consider a Hamiltonian $H_\varepsilon = H_0 + \varepsilon H_1$, with $H_1 \in \mathcal{U}$ and $\varepsilon \in]0, \varepsilon_0(H_1)[$. We take a compatible set of data $(\Gamma_1, \lambda, \delta)$ such that $H_1 \in \mathcal{U}(\Gamma_1, \lambda, \delta)$ and $\varepsilon_0(H_1) \leqslant \varepsilon_2(\Gamma_1, \lambda, \delta)$. We apply Proposition 1.1 to find a change of coordinates Φ which transforms the Hamiltonian $H_0 + \varepsilon H_1$ to a Hamiltonian in the normal form $\Phi^* H_\varepsilon = N_\varepsilon$ with $R \in \mathcal{R}(\Gamma_1, \varepsilon, \delta)$. The change of coordinates Φ is fixed for the sequel of this discussion, as well as ε. By Theorem 1.2, the Hamiltonian N_ε can be approximated in the C^r norm by Hamiltonians $\widetilde{N}_\varepsilon$ admitting an orbit $(\theta(t), p(t))$ such that $p(0) = p_-$ and $p(T) = p_+$ for some $T \in \mathbb{N}$. Let us denote by $\widetilde{H}_\varepsilon := (\Phi^{-1})^* \widetilde{N}_\varepsilon$ the expression in the original coordinates of $\widetilde{N}_\varepsilon$. It can be made arbitrarily C^r-close to H_ε by taking $\widetilde{N}_\varepsilon$ sufficiently close to N_ε. Since $\|\Phi - \mathrm{id}\|_{C^0} \leqslant \sqrt{\varepsilon}$, the extended $\widetilde{H}_\varepsilon$-orbit

$$(x(t), y(t), t \bmod 1) := \Phi(\theta(t), p(t), t \bmod 1)$$

satisfies $\|p(0) - p_-\| \leqslant \sqrt{\varepsilon}$ and $\|p(T) - p_-\| \leqslant \sqrt{\varepsilon}$, hence

$$\|y(T) - y(0)\| \geqslant \|p_+ - p_-\| - 2\sqrt{\varepsilon} > \ell \geqslant \varepsilon_0(H_1).$$

In other words, we have $\widetilde{H}_\varepsilon \in \mathcal{D}(\varepsilon_0(H_1))$. We have proved that H_ε belongs to the closure of $\mathcal{D}(\varepsilon_0(H_1))$. This ends the proof of Theorem 1.

The Hamiltonian in normal form N_ε has the typical structure of what is called an a-priori unstable system under Hypothesis 1. Actually, under the additional assumption that $\|R\|_{C^2}\leqslant\delta$, with δ sufficiently small with respect to ε, the conclusion of Theorem 1.2 would follow from the various works on the a-priori unstable case; see [8], [23], [24], [29], [36], [61], [62]. The difficulty here is the weak hypothesis made on the smallness of R, and, in particular, the fact that ε is allowed to be much smaller than δ.

1.4. Proof of Theorem 1.2

We give a proof based on several intermediate results that will be established in the further sections of the paper. The first step is to establish the existence of a normally hyperbolic cylinder. It is detailed in §3. As a consequence of the difficulties of our situation, we get only a rough control on this cylinder, as was already the case in [10]. Some C^1 norms might blow up when $\varepsilon\to 0$ (see (4)).

The second step consists in building unstable orbits along this cylinder under additional generic assumptions. In the a-priori unstable case, where a regular cylinder is present, several methods have been developed. Which of them can be extended to the present situation is unclear. Here we manage to extend the variational approach of [8], [23], [24] (which are based on Mather's work). We use the framework of [8], but also essentially appeal to ideas from [51] and [24] for the proof of one of the key genericity results. A self-contained proof of the required genericity with many new ingredients is presented in §5.

The second step consists of three main steps:

- along a resonance Γ prove existence of a normally hyperbolic cylinder $\mathcal{C}$ and derive its properties (see Theorem 1.3);
- show that this cylinlder $\mathcal{C}$ contains a family of Mañé sets $\widetilde{\mathcal{N}}(c)$, $c\in\Gamma$, each being of Aubry–Mather type, i.e. a Lipschitz graph over the circle (see Theorem 1.4);
- using the notion of a forcing class [8] to generically construct orbits diffusing along this cylinder $\mathcal{C}$ (see Theorem 1.5).

1.4.1. Existence and properties of a normally hyperbolic cylinder $\mathcal{C}$

THEOREM 1.3. *Let us consider the C^r Hamiltonian system* (3) *and assume that Z satisfies* ($HZ\lambda$) *on some arc $\Gamma_1\subset\Gamma$ of the form*

$$\Gamma_1 := \{(p_*(p^f)) : p^f \in [a_-, a_+]\}.$$

Then there exist constants $C>1>\varkappa>\delta>0$*, which depend only on* n*,* H_0*, and* λ*, and such that, for each* ε *in* $]0,\delta[$*, the following property holds for each function* $R\in\mathcal{R}(\Gamma_1,\varepsilon,\delta)$*:*

There exists a C^2 *map*

$$(\Theta^s,P^s)(\theta^f,p^f,t)\colon \mathbb{T}\times[a_- -\varkappa\varepsilon^{1/3},a_+ +\varkappa\varepsilon^{1/3}]\times\mathbb{T}\longrightarrow\mathbb{T}^{n-1}\times\mathbb{R}^{n-1}$$

such that the cylinder

$$\mathcal{C}:=\{(\theta^s,p^s)=(\Theta^s,P^s)(\theta^f,p^f,t):p^f\in[a_- -\varkappa\varepsilon^{1/3},a_+ +\varkappa\varepsilon^{1/3}] \text{ and } (\theta^f,t)\in\mathbb{T}\times\mathbb{T}\}$$

is weakly invariant with respect to N_ε *in the sense that the Hamiltonian vector field is tangent to* $\mathcal{C}$*. The cylinder* $\mathcal{C}$ *is contained in the set*

$$W:=\{(\theta,p,t):p^f\in[a_- -\varkappa\varepsilon^{1/3},a_+ +\varkappa\varepsilon^{1/3}], \|\theta^s-\theta^s_*(p^f)\|\leqslant\varkappa \text{ and } \|p^s-p^s_*(p^f)\|\leqslant\varkappa\sqrt{\varepsilon}\},$$

and it contains all the full orbits of N_ε *contained in* W*. We have the estimates*

$$\left\|\frac{\partial\Theta^s}{\partial p^f}\right\|\leqslant C\left(1+\sqrt{\frac{\delta}{\varepsilon}}\right),\qquad \left\|\frac{\partial\Theta^s}{\partial(\theta^f,t)}\right\|\leqslant C(\sqrt{\varepsilon}+\sqrt{\delta}),\tag{4}$$

$$\left\|\frac{\partial P^s}{\partial p^f}\right\|\leqslant C,\qquad \left\|\frac{\partial P^s}{\partial(\theta^f,t)}\right\|\leqslant C\sqrt{\varepsilon},\tag{5}$$

$$\|\Theta^s(\theta^f,p^f,t)-\theta^s_*(p^f)\|\leqslant C\sqrt{\delta}.$$

A similar, weaker, result is proved in [10]. The present statement contains better quantitative estimates. It follows from Theorem 3.1 below, which makes these estimates even more explicit. The terms $\varkappa\varepsilon^{1/3}$ come from the fact that we only estimate R on the $\varepsilon^{1/4}$-neighborhood of Γ_1, see the definition of $\mathcal{R}(\Gamma_1,\varepsilon,\delta)$.

For convenience of notations we extend our system from $\mathbb{T}^n\times B^n\times\mathbb{T}$ to $\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$. It is more pleasant in many occasions to consider the time-1 Hamiltonian flow ϕ and the discrete system that it generates on $\mathbb{T}^n\times\mathbb{R}^n$. We will thus consider the cylinder

$$\mathcal{C}_0=\{(q,p)\in\mathbb{T}^n\times\mathbb{R}^n:(q,p,0)\in\mathcal{C}\}.$$

We will think of this cylinder as being ϕ-invariant, although this is not precisely true, due to the possibility that orbits may escape through the boundaries. If r is large enough, it is possible to prove the existence of a really invariant cylinder closed by KAM-invariant circles, but this is not useful here.

The presence of this normally hyperbolic invariant cylinder is another similarity with the a-priori unstable case. The difference is that we only have rough control on the present cylinders, with some estimates blowing up when $\varepsilon\to0$. As we will see, variational methods can still be used to build unstable orbits along the cylinder. We will use the variational mechanism of [8]. Variational methods for this problem were initiated by Mather; see [50] in an abstract setting. In a quite different direction, they were also used by Bessi to study the Arnold's example of [1]; see [15].

1.4.2. Weak KAM and Mather theory

We will use standard notations of weak KAM and Mather theory, we recall here the most important ones for the convenience of the reader. We mostly use Fathi's presentation in terms of weak KAM solutions, see [31], and also [8] for the non-autonomous case. We consider the Lagrangian function $L(\theta, v, t)$ associated with N_ε (see §4 for the definition) and, for each $c\in\mathbb{R}^n$, the function

$$G_c(\theta_0,\theta_1) := \min_\gamma \int_0^1 (L(\gamma(t),\dot\gamma(t),t) - c\cdot\dot\gamma(t))\,dt,$$

where the minimum is taken on the set of C^1 curves $\gamma\colon[0,1]\to\mathbb{T}^n$ such that $\gamma(0)=\theta_0$ and $\gamma(1)=\theta_1$. It is a classical fact that this minimum exists, and that the minimizers is the projection of a Hamiltonian orbit. A (discrete) *weak* KAM *solution at cohomology* c is a function $u\in C(\mathbb{T}^n,\mathbb{R})$ such that

$$u(\theta) = \min_{v\in\mathbb{R}^n} (u(\theta-v) + G_c(\theta-v,\theta) + \alpha(c)),$$

where $\alpha(c)$ is the only real constant such that such a function u exists. For each curve $\gamma\colon\mathbb{R}\to\mathbb{T}^n$ and each $S<T$ in $\mathbb{Z}$, we thus have the inequalities

$$u(\gamma(T)) - u(\gamma(S)) \leqslant G_c(\gamma(S),\gamma(T)) + (T-S)\alpha(c) \leqslant \int_S^T (L(\gamma(t),\dot\gamma(t),t) - c\cdot\dot\gamma(t) + \alpha(c))\,dt.$$

A curve $\theta\colon\mathbb{R}\to\mathbb{T}^n$ is said to be *calibrated* by u if

$$u(\theta(T)) - u(\theta(S)) = \int_S^T (L(\theta(t),\dot\theta(t),t) - c\cdot\dot\theta(t) + \alpha(c))\,dt,$$

for each $S<T$ in $\mathbb{Z}$. The curve θ is then the projection of a Hamiltonian orbit (θ,p), such an orbit is called a *calibrated orbit*. We denote by

$$\widetilde{\mathcal{I}}(u,c) \subset \mathbb{T}^n \times \mathbb{R}^n$$

the union of all calibrated orbits $(\theta,p)(t)$ of the sets $(\theta,p)(\mathbb{Z})$, or equivalently of the sets $(\theta,p)(0)$. In other words, these are the initial conditions of the orbits which are calibrated by u. By definition, the set $\widetilde{\mathcal{I}}(u,c)$ is invariant under the time-1 Hamiltonian flow φ, it is moreover compact and not empty. We also denote by

$$s\widetilde{\mathcal{I}}(u,c)\subset\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$$

the suspension of $\widetilde{\mathcal{I}}(u,c)$, or in other words the set of points of the form

$$(\theta(t),p(t),t \bmod 1)$$

for each $t\in\mathbb{R}$ and each calibrated orbit (θ,p). The set $s\widetilde{\mathcal{I}}(u,c)$ is compact and invariant under the extended Hamiltonian flow. Note that $s\widetilde{\mathcal{I}}(u,c)\cap\{t=0\}=\widetilde{\mathcal{I}}(u,c)\times\{0\}$. The projection

$$\mathcal{I}(u,c)\subset\mathbb{T}^n$$

of $\widetilde{\mathcal{I}}(u,c)$ on $\mathbb{T}^n$ is the union of points $\theta(0)$, where θ is a calibrated curve. The projection

$$s\mathcal{I}(u,c)\subset\mathbb{T}^n\times\mathbb{T}$$

of $s\widetilde{\mathcal{I}}(u,c)$ on $\mathbb{T}^n\times\mathbb{T}$ is the union of points $(\theta(t),t \bmod 1)$, where $t\in\mathbb{R}$ and θ is a calibrated curve. It is an important result of Mather theory that $s\widetilde{\mathcal{I}}(u,c)$ is a Lipschitz graph above $s\mathcal{I}(u,c)$ (hence $\widetilde{\mathcal{I}}(u,c)$ is a Lipschitz graph above $\mathcal{I}(u,c)$). We finally define the Aubry and Mañé sets by

$$\tilde{\mathcal{A}}(c)=\bigcap_u\widetilde{\mathcal{I}}(u,c),\ \ s\tilde{\mathcal{A}}(c)=\bigcap_u s\widetilde{\mathcal{I}}(u,c),\ \ \widetilde{\mathcal{N}}(c)=\bigcup_u\widetilde{\mathcal{I}}(u,c),\ \ s\widetilde{\mathcal{N}}(c)=\bigcup_u s\widetilde{\mathcal{I}}(u,c), \tag{6}$$

where the unions and the intersections are taken on the set of all weak KAM solutions u at cohomology c. When a clear distinction is needed, we will call the sets $s\tilde{\mathcal{A}}(c)$ and $s\widetilde{\mathcal{N}}(c)$ the suspended Aubry (and Mañé) sets. We denote by $s\mathcal{A}(c)$ and $s\mathcal{N}(c)$ the projections of $s\tilde{\mathcal{A}}(c)$ and $s\widetilde{\mathcal{N}}(c)$ on $\mathbb{T}^n\times\mathbb{T}$, respectively. Similarly, $\mathcal{A}(c)$ and $\mathcal{N}(c)$ are the projections of $\tilde{\mathcal{A}}(c)$ and $\widetilde{\mathcal{N}}(c)$ on $\mathbb{T}^n$, respectively. The Aubry set $\tilde{\mathcal{A}}(c)$ is compact, non-empty and invariant under the time-1 flow. It is a Lipschitz graph above the projected Aubry set $\mathcal{A}(c)$. The Mañé set $\widetilde{\mathcal{N}}(c)$ is compact and invariant. Its orbits (under the time-1 flow) either belong, or are bi-asymptotic, to $\tilde{\mathcal{A}}(c)$.

In [8], an equivalence relation is introduced on the cohomology $H^1(\mathbb{T}^n,\mathbb{R})=\mathbb{R}^n$, called *forcing relation*. It will not be useful for the present exposition to recall the precise definition of this forcing relation. What is important is that, if c and c' belong to the same forcing class, then there exists an orbit (θ,p) and an integer $T\in\mathbb{N}$ such that $p(0)=c$ and $p(T)=c'$. We will establish here that, in the presence of generic additional assumptions, the resonant arc Γ_1 is contained in a forcing class, which implies the conclusion of Theorem 1.2, but also the existence of various types of orbits, see [8, §5] for more details. To prove that Γ_1 is contained in a forcing class, it is enough to prove that each of its points is in the interior of its forcing class. This can be achieved using the mechanisms exposed in [8], called the Mather mechanism and the Arnold mechanism, under appropriate informations on the sets

$$\tilde{\mathcal{A}}(c)\subset\widetilde{\mathcal{I}}(u,c)\subset\widetilde{\mathcal{N}}(c),\quad c\in\Gamma_1.$$

1.4.3. Localization and a graph theorem

The first step is to relate these sets to the normally hyperbolic cylinder $\mathcal{C}_0$ as follows.

THEOREM 1.4. *In the context of Theorem 1.3, we may assume, by possibly reducing the constant* $\delta>0$, *that the following additional property holds for each function* $R\in\mathcal{R}(\Gamma_1,\varepsilon,\delta)$ *with* $\varepsilon\in]0,\delta[$:

For each $c\in\Gamma_1$, *the Mañé set* $\widetilde{\mathcal{N}}(c)$ *is contained in the cylinder* $\mathcal{C}_0$. *Moreover, the restriction of the coordinate map* $\theta^f:\mathbb{T}^n\times\mathbb{R}^n\to\mathbb{T}$ *to* $\widetilde{\mathcal{I}}(u,c)$ *is a bi-Lipschitz homeomorphism for each weak* KAM *solution* u *at cohomology* c.

Proof. The proof is based on estimates on weak KAM solutions that will be established in §4. Let $\varkappa$ be as given by Theorem 1.3. Theorem 4.1 (which is stated and proved in §4) implies that the suspended Mañé set $s\widetilde{\mathcal{N}}(c)$ is contained in the set

$$\{\|\theta^s-\theta^s_*(c^f)\|\leqslant\varkappa : \|p^s-p^s_*(c^f)\|\leqslant\varkappa\sqrt{\varepsilon}\text{ and }|p^f-c^f|\leqslant\varkappa\sqrt{\varepsilon}\},$$

provided $R\in\mathcal{R}(\Gamma_1,\varepsilon,\varkappa^{16})$ and $\varepsilon\in]0,\varepsilon_0[$ (a constant depending on $\varkappa$). As a consequence, this inclusion holds for $R\in\mathcal{R}(\Gamma_1,\varepsilon,\delta)$ and $\varepsilon\in]0,\delta[$, with $\delta=\min(\varkappa^{16},\varepsilon_0)$. The suspended Mañé set $s\widetilde{\mathcal{N}}(c)$ is then contained in the domain called W in the statement of Theorem 1.3. It is thus contained in $\mathcal{C}$, hence $\widetilde{\mathcal{N}}(c)\subset\mathcal{C}_0$.

Let us consider a weak KAM solution u of N_ε at cohomology c and prove the projection part of the statement. Let (θ_i,p_i), $i=1,2$ be two points in $\widetilde{\mathcal{I}}(u,c)$. By Theorem 4.2, we have

$$\|p_2-p_1\|\leqslant 9\sqrt{D\varepsilon}\|\theta_2-\theta_1\|\leqslant 9\sqrt{D\varepsilon}(\|\theta^f_2-\theta^f_1\|+\|\theta^s_2-\theta^s_1\|).$$

Since the points belong to $\mathcal{C}_0$, the last estimate in Theorem 1.3 implies that

$$\|\theta^s_2-\theta^s_1\|\leqslant C\left(1+\sqrt{\frac{\delta}{\varepsilon}}\right)(\|\theta^f_2-\theta^f_1\|+\|p_2-p_1\|).$$

We get

$$\|p_2-p_1\|\leqslant 9C\sqrt{D}\big(2\sqrt{\varepsilon}+\sqrt{\delta}\big)\|\theta^f_2-\theta^f_1\|+9C\sqrt{D}\big(\sqrt{\varepsilon}+\sqrt{\delta}\big)\|p_2-p_1\|.$$

If δ is small enough and $\varepsilon<\delta$, then

$$9C\sqrt{D}\big(\sqrt{\varepsilon}+\sqrt{\delta}\,\big)\leqslant 9C\sqrt{D}\big(2\sqrt{\varepsilon}+\sqrt{\delta}\,\big)\leqslant \tfrac{1}{2},$$

hence

$$\|p_2-p_1\|\leqslant 9C\sqrt{D}\big(2\sqrt{\varepsilon}+\sqrt{\delta}\,\big)\|\theta_2^f-\theta_1^f\|+\tfrac{1}{2}\|p_2-p_1\|,$$

thus

$$\|p_2-p_1\|\leqslant 9C\sqrt{D}\big(4\sqrt{\varepsilon}+2\sqrt{\delta}\,\big)\|\theta_2^f-\theta_1^f\|\leqslant \|\theta_2^f-\theta_1^f\|. \qquad \square$$

1.4.4. Structure of Aubry sets inside the cylinder and existence of diffusing orbits

This last result, in conjunction with the theory of circle homeomorphisms, has strong consequences:

All the orbits of $\tilde{\mathcal{A}}_0(c)$ have the same rotation number $\varrho(c)=(\varrho^f(c),0)$, with $\varrho^f(c)\in\mathbb{R}$. Since the sub-differential $\partial\alpha(c)$ of the convex function α is the rotation set of $\tilde{\mathcal{A}}(c)$, we conclude that the function α is differentiable at each point of Γ_1, with $d\alpha(c)=(\varrho^s(c),0)$.

When $\varrho^s(c)$ is rational, the Mather minimizing measures are supported on periodic orbits.

When $\varrho^s(c)$ is irrational, the invariant set $\tilde{\mathcal{A}}(c)$ is uniquely ergodic. As a consequence, there exists one and only one weak KAM solution (up to the addition of an additive constant), hence $\widetilde{\mathcal{N}}(c)=\tilde{\mathcal{A}}(c)$.

In the irrational case, we will have to consider homoclinic orbits. Such orbits can be dealt with by considering the two-fold covering

$$\xi\colon \mathbb{T}^n\longrightarrow\mathbb{T}^n,$$
$$\theta=(\theta^f,\theta_1^s,\theta_2^s,...,\theta_{n-1}^s)\longmapsto\xi(\theta)=(\theta^f,2\theta_1^s,\theta_2^s,...,\theta_{n-1}^s).$$

The idea of using a covering to study homoclinic orbits comes from Fathi; see [30]. This covering lifts to a symplectic covering

$$\Xi\colon \mathbb{T}^n\times\mathbb{R}^n\longrightarrow\mathbb{T}^n\times\mathbb{R}^n,$$
$$(\theta,p)=(\theta,p^f,p_1^s,p_2^s,...,p_{n-1}^s)\longmapsto\Xi(\theta,p)=\big(\xi(\theta),p^f,\tfrac{1}{2}p_1^s,p_2^s,...,p_{n-1}^s\big),$$

and we define the lifted Hamiltonian $\widetilde{N}=N\circ\Xi$. It is known, see [30], [25], [8], that

$$\tilde{\mathcal{A}}_{H\circ\Xi}(\xi^*c)=\Xi^{-1}(\tilde{\mathcal{A}}_H(c)),$$

where $\xi^* c = \big(c^f, \frac{1}{2}c^s_1, c^s_2, ..., c^s_{n-1}\big)$. On the other hand, the inclusion

$$\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^* c) \supset \Xi^{-1}(\widetilde{\mathcal{N}}_N(c)) = \Xi^{-1}(\tilde{\mathcal{A}}_N(c))$$

is not an equality. More precisely, in the present situation, the set $\tilde{\mathcal{A}}_{N\circ\Xi}(\tilde{c})$ is the union of two disjoint homeomorphic copies of the circle $\tilde{\mathcal{A}}_N(\tilde{c})$, and $\widetilde{\mathcal{N}}_{N\circ\Xi}(\tilde{c})$ contains heteroclinic connections between these copies (which are the liftings of orbits homoclinic to $\tilde{\mathcal{A}}_N(c)$). More can be said if we are allowed to make a small perturbation to avoid degenerate situations. We recall that a metric space is called totally disconnected if its only connected subsets are its points. The hypothesis of total disconnectedness in the following statement can be seen as a weak form of transversality of the stable and unstable manifolds of the invariant circle $\tilde{\mathcal{A}}_N(c)$.

THEOREM 1.5. *In the context of Theorems 1.3 and 1.4, the following property holds for a dense subset of functions $R \in \mathcal{R}(\Gamma_1, \varepsilon, \delta_0)$ (for the C^r topology). Each $c \in \Gamma_1$ is in one of the following cases:*

(1) *$\theta^f(\mathcal{I}(u,c)) \subsetneq \mathbb{T}$ for each weak KAM solution u at cohomology c;*

(2) *$\varrho(c)$ is irrational, $\theta^f(\mathcal{N}_N(c)) = \mathbb{T}$ (hence, $\widetilde{\mathcal{N}}_N(c)$ is an invariant circle), and*

$$\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^* c) \setminus \Xi^{-1}(\widetilde{\mathcal{N}}_N(c))$$

is totally disconnected.

The arc Γ_1 is then contained in a forcing class, and hence the conclusion of Theorem 1.2 holds.

Proof. By general results on Hamiltonian dynamics, the set $\mathcal{R}_1 \subset \mathcal{R}(\Gamma_1, \varepsilon, \delta_0)$ of functions R such that the flow map ϕ does not admit any non-trivial invariant circle of rational rotation number is C^r-dense. This condition holds for example if N is Kupka Smale (in the Hamiltonian sense, see [59] for example).

Since the coordinate map θ^f is a homeomorphism when restricted to $\widetilde{\mathcal{I}}(u,c)$, this set is an invariant circle if $\theta^f(\mathcal{I}(u,c)) = \mathbb{T}$. If $R \in \mathcal{R}_1$, this implies that the rotation number $\varrho^f(c)$ is irrational. In other words, for $R \in \mathcal{R}_1$, condition (1) can be violated only at points c when $\varrho^f(c)$ is irrational, and then $\widetilde{\mathcal{I}}(u,c) = \tilde{\mathcal{A}}(c) = \widetilde{\mathcal{N}}(c)$ is an invariant circle.

When $R \in \mathcal{R}_1$, it is possible to perturb R away from $\mathcal{C}_0$ in such a way that

$$\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^* c) \setminus \Xi^{-1}(\widetilde{\mathcal{N}}_N(c))$$

is totally disconnected for each value of c such that $\widetilde{\mathcal{N}}(c)$ is an invariant circle. This second perturbation procedure is not easy because there are uncountably many such values of c. This is the result of Theorem 5.1. A result of this kind was obtained in [24], here we give a self-contained proof with many new ingredients; see §5.

We now explain, under the additional condition ((1) or (2)), how the variational mechanisms of [8] can be applied to prove that Γ_1 is contained in a forcing class. It is enough to prove that each point $c \in \Gamma_1$ is in the interior of its forcing class. We treat separately the two cases.

In the first case, we can apply the Mather mechanism, see (0.11) in [8, §0.11]. In that paper, the subspace $Y(u,c) \subset \mathbb{R}^n$, defined as the set of cohomology classes of closed 1-forms whose support is disjoint from $\mathcal{I}(u,c)$, is associated with each weak KAM solution u at cohomology c (in [8], the notation $R(\mathcal{G})$ is used). In the present case, we know that the map θ^f restricted to $\widetilde{\mathcal{I}}(u,f)$ is a bi-Lipschitz homeomorphism which is not onto. We conclude that $Y(u,c)=\mathbb{R}^n$. Since this holds for each weak KAM solution u, we conclude that

$$Y(c) := \bigcap_u Y(u,c) = \mathbb{R}^n.$$

The result called Mather mechanism in [8] states that there is a small ball $B \subset Y(c)$ centered at 0 in Y such that the forcing class of c contains $c+B$. In the present situation, we conclude that c is in the interior of its forcing class.

In the second case, we can apply the Arnold's mechanism; see [8, §9]. We work with the Hamiltonian $N \circ \Xi$ lifted to the 2-fold cover. By Proposition (7.3) in [8], it is enough to prove that $\xi^* c$ is in the interior of its forcing class for the lifted Hamiltonian $N \circ \Xi$; this implies that c is in the interior of its forcing class for N.

The preimage $\Xi^{-1}(\widetilde{\mathcal{N}}_N(c))$ is the union of two closed curves $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$. The set $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^* c)$ contains these two curves, as well as a set $\widetilde{\mathcal{H}}_{12}$ of heteroclinic connections from $\widetilde{\mathcal{S}}_1$ to $\widetilde{\mathcal{S}}_2$, and a set $\widetilde{\mathcal{H}}_{21}$ of heteroclinic connections from $\widetilde{\mathcal{S}}_2$ to $\widetilde{\mathcal{S}}_1$. Theorem (9.2) in [8] states that $\xi^* c$ is in the interior of its forcing class, provided $\widetilde{\mathcal{H}}_{12}$ and $\widetilde{\mathcal{H}}_{21}$ are totally disconnnected. Actually, the hypothesis is stated in [8] in a slightly different way, we explain in Appendix B that total disconnectedness actually implies the hypothesis of [8]. We conclude that each $c \in \Gamma_1$ is in the interior of its forcing class. Since Γ_1 is connected, it is contained in a single forcing class. It is then a simple consequence of the definition of the forcing relation, see [8, §5], that the conclusion of Theorem 1.2 holds. This ends the proof of Theorem 1.2, using the results proved in the rest of the paper. □

1.5. Bifurcation points and a longer diffusion path

This section discusses some improvements on Theorems 1 and 1.2. There are two limitations to the size of the resonant arc $\Gamma_1 \subset \Gamma$ to which the above construction can be applied.

The first limitation comes from the assumption that hypothesis ($HZ\lambda$) should hold on Γ_1. Given a resonant arc $\Gamma_2 \subset \Gamma$, it is generic to satisfy this condition on a certain subarc $\Gamma_1 \subset \Gamma_2$, but it is not generic to satisfy ($HZ\lambda$) on the whole of Γ_2. The presence of values of $c \in \Gamma_2$ such that $Z(\cdot, c)$ has two non-degenerate maxima cannot be excluded. In this section, we explain how a modification of the proof of Theorem 1.2 allows us to get rid of this limitation.

The second limitation comes from the normal form theorem, and from the impossibility to incorporate a finite set of additional resonances (punctures) in the domain of our normal forms. This limitation is serious, and bypassing it would require specific work around additional resonances, which will not be discussed here. Some preprints on this issue appeared after the first version of the present work; see [20], [39], [40] (the latter ones being sequels to the present work, and the first one is independent). Here, the best we can achieve is to prove existence of diffusion orbits between two consecutive punctures. The number of punctures is independant from ε, it depends on the parameter δ in Theorem 1.2, which can be computed using the non-degeneracy parameter λ; see Remark 2.1.

In order to get rid of the first limitation, we consider a second hypothesis on Z.

Hypothesis 2. There exists a real number $\lambda > 0$ and two points $\vartheta_1^s, \vartheta_2^s$ in $\mathbb{T}^{n-1}$ such that the balls $B(\vartheta_1^s, 3\lambda)$ and $B(\vartheta_2^s, 3\lambda)$ are disjoint and such that, for each $p \in \Gamma_1$, there exist two local maxima $\theta_1^s(p) \in B(\vartheta_1^s, \lambda)$ and $\theta_2^s(p) \in B(\vartheta_2^s, \lambda)$ of the function $Z(\cdot, p)$ in $\mathbb{T}^{n-1}$ satisfying

$$\partial^2_{\theta^s} Z(\theta_1^s(p), p) \leqslant \lambda I, \quad \partial^2_{\theta^s} Z(\theta_2^s(p), p) \leqslant \lambda I,$$

and

$$Z(\theta^s, p) \leqslant \max\{Z(\theta_1^f(p), p), Z(\theta_2^f(p), p)\} - \lambda(\min\{d(\theta^s - \theta_1^s), d(\theta^s - \theta_2^s)\})^2$$

for each $p \in \Gamma_1$ and each $\theta^s \in \mathbb{T}^{n-1}$.

Given an arc $\Gamma_2 \in \mathbb{R}^n$, the following property is generic in $C^r(\mathbb{T}^{n-1} \times \mathbb{R}^n, \mathbb{R})$:

The arc Γ_2 is a finite union of subarcs such that either Hypothesis 1 or 2 holds on each of these subarcs, with a common constant $\lambda > 0$.

We have the following improvement on Theorem 1.2.

Proposition 1.6. *For the system* (3), *assume that there exists $\lambda > 0$ such that for each $c \in \Gamma_1$, either Hypothesis* 1 *or* 2 *holds for each $c \in \Gamma_1$. Then there exists $\delta > 0$, which depend only on n, H_0, and λ, and such that, for each $\varepsilon \in]0, \delta[$, the following property holds for a dense subset of functions $R \in \mathcal{R}(\Gamma_1, \varepsilon, \delta)$ (for the C^r topology):*

There exist an orbit (θ, p) and an integer $T \in \mathbb{N}$ such that

$$p(0) = p_*(a_-) \quad \text{and} \quad p(T) = p_*(a_+).$$

Proof of Proposition 1.6. We use the same framework as in the proof of Theorem 1.2, so it is enough to prove that each element of Γ_1 is in the interior of its forcing class.

Observe first that Theorem 3.1 can be applied to prove the existence of two invariant cylinders $\mathcal{C}^1$ and $\mathcal{C}^2$ in the extended phase space $\mathbb{T}^n \times \mathbb{R}^n \times \mathbb{T}$. Moreover, we can choose the parameter $\varkappa$ smaller than λ, in such a way that

$$\theta^s(\mathcal{C}_1) \subset B(\vartheta_1^s, 2\lambda) \quad \text{and} \quad \theta^s(\mathcal{C}_2) \subset B(\vartheta_2^s, 2\lambda).$$

As earlier, we denote by $\mathcal{C}_0^1$ and $\mathcal{C}_0^2$ the intersections with the section $\{t=0\}$. By Theorem 4.12, we have

$$\tilde{\mathcal{A}}(c) \subset \mathcal{C}_0^1 \cup \mathcal{C}_0^2$$

for each $c \in \Gamma_1$. Let us now introduce two smooth functions $F_i(\theta^s): \mathbb{T}^{n-1} \to [0,1]$, $i \in \{1,2\}$, with the property that $F_1=1$ in $B(\vartheta_2^s, 2\lambda)$, $F_1=0$ outside of $B(\vartheta_2^s, 3\lambda)$, $F_2=1$ in $B(\vartheta_1^s, 2\lambda)$ and $F_2=0$ outside of $B(\vartheta_1^s, 3\lambda)$

Considering the modified Hamiltonians $N-F_i$ will help the description of the Mather sets of N. One can check by inspection in the proofs (using that F_i does not depend on p) that Theorem 4.1 applies to $N-F_i$, and allows to conclude that the Mañé set $\widetilde{\mathcal{N}}_i(c)$ of $N-F_i$ is contained in $\mathcal{C}_0^i$. Let us denote by $\alpha_i(c)$ the α function of $N-F_i$. These objects are closely related to Mather's local Aubry sets.

LEMMA 1.7. *For each* $c \in \Gamma_1$, *the* $\alpha_i(c)$ *are differentiable at* c, *and*

$$\alpha(c) = \max\{\alpha_1(c), \alpha_2(c)\}.$$

Moreover,

- *if* $\alpha(c)=\alpha_1(c)>\alpha_2(c)$, *then* $\widetilde{\mathcal{N}}(c)=\widetilde{\mathcal{N}}_1(c)$;
- *if* $\alpha(c)=\alpha_2(c)>\alpha_1(c)$, *then* $\widetilde{\mathcal{N}}(c)=\widetilde{\mathcal{N}}_2(c)$;
- *if* $\alpha(c)=\alpha_1(c)=\alpha_2(c)$, *then* $\widetilde{\mathcal{N}}_1(c) \cup \widetilde{\mathcal{N}}_2(c) \subsetneq \widetilde{\mathcal{N}}(c)$.

Proof. The functions $\alpha_i(c)$ are C^1 for the same reason as $\alpha(c)$ is C^1 in the one-peak case.

Since $N-N_i \leqslant N$, we have $\alpha_i(c) \leqslant \alpha(c)$. On the other hand, we know that

$$\alpha(c) = \max_\mu \left(c \cdot \varrho(\mu) - \int (p\, \partial_p N - N)\, d\mu \right),$$

where the maximum is taken over the set of invariant measures μ. Since we know that $\tilde{\mathcal{A}}(c) \subset \mathcal{C}_0^1 \cup \mathcal{C}_0^2$, and since the maximizing measures are supported on the Aubry set, we conclude that each ergodic maximizing measure is supported either on $\mathcal{C}^1$ or on $\mathcal{C}^2$. If the measure is supported in $\mathcal{C}^i$, then we have

$$\alpha_i(c) \geqslant c \cdot \varrho(\mu) - \int (p\, \partial_p N - N + F_i)\, d\mu = c \cdot \varrho(\mu) - \int (p\, \partial_p N - N)\, d\mu = \alpha(c).$$

This proves the equality $\alpha(c)=\max\{\alpha_1(c),\alpha_2(c)\}$.

As is explained in the proof of Theorem 4.12, there are two possibilities for the Mañé set $\widetilde{\mathcal{N}}(c)$: either it is contained in one of the $\mathcal{C}_0^i$, or it intersects both of them, and then also contains connections (because it is necessarily chain transitive).

If the Mañé set $\widetilde{\mathcal{N}}(c)$ intersects $\mathcal{C}_0^i$, then the intersection is a compact invariant set, which thus support an invariant measure. This measure must be maximizing the functional $c\cdot\varrho(\mu)-\int(p\,\partial_p N-N\,d\mu)$, and thus also the functional

$$c\cdot\varrho(\mu)-\int(p\,\partial_p N-N+F_i)\,d\mu.$$

As a consequence, we must have $\alpha(c)=\alpha_i(c)$. □

We can prove, by the variational mechanisms of [8], that a point c is in the interior of its forcing class in the following three cases:

First case: the Mañé $\widetilde{\mathcal{N}}(c)$ set is contained in one of the cylinders $\mathcal{C}_0^i$, and it does not contain any invariant circle. Then the Mather mechanism applies as in the single-peak case, and c is contained in the interior of its forcing class.

Second case: the Mañé set is an invariant circle (then necessarily contained in one of the cylinders $\mathcal{C}_0^i$), it is uniquely ergodic, and $\widetilde{\mathcal{N}}_{N\circ\Xi}(c)\setminus\Xi^{-1}(\widetilde{\mathcal{N}}(c))$ is totally disconnected. Then the Arnold's mechanism applies as in the single peak case, and c is contained in the interior of its forcing class.

Third case: the sets $\widetilde{\mathcal{N}}_i(c)$ are both non-empty and uniquely ergodic, and

$$\widetilde{\mathcal{N}}(c)\setminus(\widetilde{\mathcal{N}}_1(c)\cup\widetilde{\mathcal{N}}_2(c))$$

is totally disconnected. Then the Arnold's mechanism applies directly (without taking a cover), and c is contained in the interior of its forcing class.

Each $c\in\Gamma_1$ is in one of these three cases provided the following set of additional conditions holds:

- the sets $\widetilde{\mathcal{N}}_i(c)$ are uniquely ergodic;
- the equality $\alpha_1(c)=\alpha_2(c)$ has finitely many solutions on Γ_1;
- the set $\widetilde{\mathcal{N}}(c)\setminus(\widetilde{\mathcal{N}}_1(c)\cup\widetilde{\mathcal{N}}_2(c))$ is totally disconnected (and not empty) in case $\alpha_1(c)=\alpha_2(c)$;
- the set $\widetilde{\mathcal{N}}_{N\circ\Xi}(c)\setminus\Xi^{-1}(\widetilde{\mathcal{N}}(c))$ is totally disconnected whenever $\widetilde{\mathcal{N}}(c)$ is an invariant circle.

Let us now explain how these conditions can be imposed by a C^r perturbation of R.

We start by considering a perturbation R_1 of R such that, for each rational number $\varrho\in\mathbb{Q}\setminus\{0\}$, there exists a unique Mather minimizing measure of rotation number ϱ. Such a condition is known to be generic (because it concerns only countably many rotation numbers); see [47], [25], [12], [11].

We then consider a perturbation R_2 of the form R_1-sF_1, with a small $s>0$. It is easy to see that the functions $\alpha_i^2(c), c\in\Gamma_1$, associated with the Hamiltonian $H_0+\varepsilon Z+\varepsilon R_2$ are

$$\alpha_1^2(c)=\alpha_1^1(c) \quad \text{and} \quad \alpha_2^2(c)=\alpha_2^1(c)+s,$$

where $\alpha_i^1(c)$ are the functions associated with $H_0+\varepsilon Z+\varepsilon R_1$. By Sard's theorem, there exist arbitrarily small regular values s of the difference $\alpha_1^1-\alpha_2^1$. If s is such a value, then 0 is a regular value of the difference $\alpha_1^2-\alpha_2^2$, hence the equation $\alpha_1^2(c)=\alpha_2^2(c)$ has only finitely many solutions on Γ. Note that the perturbation is locally constant around the cylinders $\mathcal{C}^i$, hence this second perturbation does not destroy the first property.

We then perform new perturbations supported away from $\mathcal{C}^i$, which preserve the first two properties. The third property is not hard to obtain since it now concerns only finitely many values of c. The last property is obtained using arguments of §5.

We have proved that the Hamiltonian R can be perturbed in such a way that each point of Γ_1 is in the interior of its forcing class. □

2. Normal forms

The goal of the present section is to prove Proposition 1.1 which allows to reduce Theorem 1 to Theorem 1.2. This reduction to the normal form does not use the convexity assumption. We put the initial Hamiltonian H_ε in normal form around a compact subarc Γ_2 of the resonance

$$\Gamma=\{p^s=p_*(p^f)\}=\{p\in\mathbb{R}^n:\partial_{p^s}H_0=0\}.$$

This global normal form is obtained by using mollifiers to glue local normal forms that depends on the arithmetic properties of the frequencies. This allows a simpler proof for instability, as we avoid the need to justify transitions between different local coordinates.

Recall that we study a resonance of order $n-1$ or, equivalently, of codimension 1. The resonance of order $n-1$ is given by a lattice Λ spanned by $n-1$ linearly independent vectors $k_1,...,k_{n-1}\in(\mathbb{Z}^n\setminus\{0\})\times\mathbb{Z}$. Denote by $\theta_j^s=k_j\cdot\theta$, $\omega_j^s=k_j\cdot\nabla H_0(p)$, $j=1,...,n-1$ and $\theta^s=(\theta_1^s,...,\theta_{n-1}^s)$ the slow angles and by $\omega^s=(\omega_1^s,...,\omega_{n-1}^s)$ the slow actions, respectively. Choose a complement angle θ^f so that $(\theta^s,\theta^f)\in\mathbb{T}^{n-1}\times\mathbb{T}$ form a basis.

For $p\in\Gamma$ we have $\omega(p)=(0,\partial_{p^f}H_0(p))$. We say that p has an additional resonance if the remaining frequency $\partial_{p^f}H_0(p)$ is rational. In order to reduce the system to an appropriate normal form, we must remove some additional resonances. More precisely, we denote by $\mathcal{D}(K,s)\subset B$ the set of momenta p such that

- $\|\partial_{p^s} H_0(p)\|\leqslant s$, and
- $|k^f\partial_{p^f}H_0(p)+k^t|\geqslant 3Ks$ for each $(k^f,k^t)\in\mathbb{Z}^2$ satisfying $\max\{|k^f|,|k^t|\}\in]0,K]$.

The following result, which does not use the convexity of H_0, is a refinement of Proposition 1.1:

THEOREM 2.1. (Normal Form) *Let $H_0(p)$ be a C^4 Hamiltonian. For each $\delta\in]0,1[$, there exist positive parameters K_0, ε_0, and β such that, for each C^4 Hamiltonian H_1 with $\|H_1\|_{C^4}\leqslant 1$, each $K\geqslant K_0$, and each $\varepsilon\leqslant\varepsilon_0$, there exists a smooth change of coordinates*

$$\Phi\colon \mathbb{T}^n\times B\times\mathbb{T}\longrightarrow\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$$

satisfying $\|\Phi-\mathrm{id}\|_{C^0}\leqslant\sqrt{\varepsilon}$ and $\|\Phi-\mathrm{id}\|_{C^2}\leqslant\delta$ and such that, in the new coordinates, the Hamiltonian $H_0+\varepsilon H_1$ takes the form

$$N_\varepsilon=H_0(p)+\varepsilon Z(\theta^s,p)+\varepsilon R(\theta,p,t),$$

with $\|R\|_{C^2}\leqslant\delta$ on $\mathbb{T}^n\times\mathcal{D}(K,\beta\varepsilon^{1/4})\times\mathbb{T}$. We can take $K_0=c\delta^{-2}$, $\beta=c\delta^{-1-n}$, and $\varepsilon_0=\delta^{6n+5}/c$, where $c>0$ is some constant depending only on n and $\|H_0\|_{C^4}$.

The proof actually builds a symplectic diffeomorphism $\tilde\Phi$ of $\mathbb{T}^{n+1}\times\mathbb{R}^{n+1}$ of the form

$$\tilde\Phi(\theta,p,t,e)=(\Phi(\theta,p,t),e+f(\theta,p,t))$$

and such that

$$N_\varepsilon+e=(H_\varepsilon+e)\circ\tilde\Phi.$$

We have the estimates $\|\tilde\Phi-\mathrm{id}\|_{C^0}\leqslant\sqrt{\varepsilon}$ and $\|\tilde\Phi-\mathrm{id}\|_{C^2}\leqslant\delta$.

Remark 2.1. (Distance between punctures) On the interval, the distance between two adjacent rationals with denominator at most K is $1/K^2$. Choose $K=K_0$ as in Theorem 2.1, the distance between adjacent punctures is at least $D^{-1}/K^2\geqslant D^{-1}c^{-1}\delta^4$.

The length of Γ_1 is determined by the choice of δ, which can be chosen optimally in Theorem 1.3 and Theorem 4.1. Upon inspection of the proof, it is not difficult to determine that δ can be chosen to a power of λ, which shows the distance between punctures is polynomial in λ.

To prove Theorem 2.1 we proceed in three steps. We first obtain a global normal form N_ε adapted to all resonances. We then show that this normal form takes the desired form on the domain $\mathcal{D}(K,s)$. However, the averaging procedure lowers smoothness, in particular, the technique requires the smoothness $r\geqslant n+5$. To obtain a result that does not require this relation between r and n, we use a smooth approximation trick that goes back to Moser.

2.1. A global normal form adapted to all resonances

We first state a result for autonomous systems. The time-periodic version will come as a corollary. Consider the Hamiltonian $H_\varepsilon(\phi, J)=H_0(J)+\varepsilon H_1(\phi, J)$, where $(\phi, J)\in \mathbb{T}^m\times\mathbb{R}^m$ (later, we will take $m=n+1$). Let $B=\{|J|\leqslant 1\}$ be the unit ball in $\mathbb{R}^m$. Given any integer vector $k\in\mathbb{Z}^m\setminus\{0\}$, let $[k]=\max_i\{|k_i|\}$. To avoid zero denominators in some calculations, we make the unusual convention that $[(0, ..., 0)]=1$. We fix once and for all a bump function $\varrho:\mathbb{R}\to\mathbb{R}$ to be a C^∞ such that

$$\varrho(x)=\begin{cases} 1, & \text{if } |x|\leqslant 1,\\ 0, & \text{if } |x|\geqslant 2,\end{cases}$$

and $0<\varrho(x)<1$ in between. For each $\beta>0$ and $k\in\mathbb{Z}^m$, we define the function $\varrho_k(J)=\varrho(k\cdot\partial_J H_0/\beta\varepsilon^{1/4}[k])$, where $\beta>0$ is a parameter.

Theorem 2.2. *There exists a constant* $c_m>0$, *which depends only on* m, *such that the following holds: given*

- *a* C^4 *Hamiltonian* $H_0(J)$;
- *a* C^r *Hamiltonian* $H_1(\varphi, J)$ *with* $\|H_1\|_{C^r}=1$;
- *parameters* $r\geqslant m+4$, $\delta\in]0,1[$, $\varepsilon\in]0,1[$, $\beta>0$, $K>0$,

satisfying

- $K\geqslant c_m\delta^{-1/(r-m-3)}$;
- $\beta\geqslant c_m(1+\|H_0\|_{C^4})\delta^{-1/2}$;
- $\beta\varepsilon^{1/4}\leqslant\|H_0\|_{C^4}$,

there exists a C^2 *symplectic diffeomorphism* $\Phi:\mathbb{T}^m\times B\to\mathbb{T}^m\times\mathbb{R}^m$ *such that, in the new coordinates, the Hamiltonian* $H_\varepsilon=H_0+\varepsilon H_1$ *takes the form*

$$H_\varepsilon\circ\Phi=H_0+\varepsilon R_1+\varepsilon R_2$$

with

- $R_1=\sum_{k\in\mathbb{Z}^m:|k|\leqslant K}\varrho_k(J)h_k(J)e^{2\pi i(k\cdot\phi)}$, *where* $h_k(J)$ *is the* k*-th coefficient for the Fourier expansion of* H_1;
- $\|R_2\|_{C^2}\leqslant\delta$;
- $\|\Phi-\mathrm{id}\|_{C^0}\leqslant\delta\sqrt{\varepsilon}$ *and* $\|\Phi-\mathrm{id}\|_{C^2}\leqslant\delta$.

If both H_0 *and* H_1 *are smooth, then so is* Φ.

We now prove Theorem 2.2. To avoid cumbersome notation, we will denote by c_m various different constants depending only on the dimension m. We have the following basic estimates about the Fourier series of a function $g(\phi, J)$. Given a multi-index $\alpha=(\alpha_1, ..., \alpha_m)$, we set $|\alpha|=\alpha_1+...+\alpha_m$. We also set $\varkappa_m=\sum_{k\in\mathbb{Z}^m}[k]^{-m-1}$.

LEMMA 2.3. *For $g(\phi, J)\in C^r(\mathbb{T}^m\times B)$, we have the following.*

(1) *If $l\leqslant r$, we have $\|g_k(J)e^{2\pi i(k\cdot\varphi)}\|_{C^l}\leqslant[k]^{l-r}\|g\|_{C^r}$.*

(2) *Let $g_k(J)$ be a series of functions such that the inequality*

$$\|\partial_{J^\alpha} g_k\|_{C^0}\leqslant M[k]^{-|\alpha|-m-1}$$

holds for each multi-index α with $|\alpha|\leqslant l$, for some $M>0$. Then, we have

$$\left\|\sum_{k\in\mathbb{Z}^m} g_k(J)e^{2\pi i(k\cdot\varphi)}\right\|_{C^l}\leqslant c\varkappa_m M.$$

(3) *Let $\Pi_K^+ g=\sum_{|k|>K} g_k(J)e^{2\pi i(k\cdot\phi)}$. Then, for $l\leqslant r-m-1$, we have*

$$\|\Pi_K^+ g\|_{C^l}\leqslant\varkappa_m K^{m-r+l+1}\|g\|_{C^r}.$$

Proof. (1) Let us assume that $k\neq 0$ and take j such that $k_j=[k]$. Let α and η be two multi-indices such that $|\alpha+\eta|\leqslant l$. Finally, let $b=r-l$, and let β be the multi-index $\beta=(0,...,0,b,0,...,0)$, where $\beta_j=b$. We have

$$g_k(J)e^{2\pi i(k\cdot\varphi)}=\int_{\mathbb{T}^m} g(\theta,J)e^{2\pi i(k,\varphi-\theta)}\,d\theta=\int_{\mathbb{T}^m} g(\theta+\varphi,J)e^{-2\pi i(k\cdot\theta)}\,d\theta,$$

and hence

$$\begin{aligned}\partial_{\varphi^\alpha J^\eta}(g_k(J)e^{2\pi i(k\cdot\varphi)})&=\int_{\mathbb{T}^m}\partial_{\varphi^\alpha J^\eta} g(\theta+\varphi,J)e^{-2\pi i(k\cdot\theta)}d\theta,\\&=\int_{\mathbb{T}^m}\frac{\partial_{\varphi^{\alpha+\beta}J^\eta} g(\theta+\varphi,J)}{(2\pi i k_j)^b}e^{-2\pi i(k\cdot\theta)}\,d\theta.\end{aligned}$$

Since $|\alpha+\beta+\eta|\leqslant r$, we conclude that

$$\|g_k(J)e^{2\pi i(k\cdot\varphi)}\|_{C^l}\leqslant\frac{\|g\|_{C^r}}{(2\pi[k])^b}\leqslant\|g\|_{C^r}[k]^{l-r}.$$

(2) We have

$$\|g_k(J)e^{2\pi i(k\cdot\varphi)}\|_{C^l}\leqslant\left\|\sum_{k\in\mathbb{Z}^m} h_k(J)e^{2\pi i(k\cdot\varphi)}\right\|_{C^l}\leqslant\sum_{k\in\mathbb{Z}^m} c_l|k|^{-r+l}M\leqslant c_l\varkappa_m M$$

(recall that $\varkappa_m=\sum_{k\in\mathbb{Z}^m}|k|^{-m-1}$).

(3) Using (1), we get

$$\begin{aligned}\|\Pi_K^+ g\|_{C^l}&\leqslant\sum_{|k|>K}[k]^{l-r}\|g\|_{C^r}\leqslant\|g\|_{C^r}K^{m-r+l+1}\sum_{|k|>K}[k]^{-m-1}\\&\leqslant\|g\|_{C^r}K^{m-r+l+1}\sum_{k\in\mathbb{Z}^m}[k]^{-m-1}.\end{aligned}$$

□

Proof of Theorem 2.2. Let $\widetilde{G}(\phi, J)$ be the function that solves the cohomological equation

$$\{H_0, \widetilde{G}\} + H_1 = R_1 + R_+,$$

where $R_+ = \Pi_K^+ H_1$. Observing that $\varrho_k(J) = 1$ when $k \cdot \partial_J H_0 = 0$, we have the following explicit formula for G:

$$\widetilde{G}(\varphi, J) = (2\pi i)^{-1} \sum_{|k| \leqslant K} \frac{(1 - \varrho_k(J)) h_k(J)}{k \cdot \partial_J H_0(J)} e^{2\pi i (k \cdot \phi)}$$

where each of the functions $(1 - \varrho_k(J)) h_k(J)/(k \cdot \partial_J H_0)$ is extended by continuity at the points where the denominator vanishes. This function hence takes the value zero at these points. G is well defined thanks to the smoothing terms $1 - \varrho_k$ we introduced, as whenever $k \cdot \partial_J H_0 = 0$ we also have $1 - \varrho_k = 0$ and that term is considered non-present. Since $\widetilde{G}$ as defined above is only C^3, we will consider a smooth approximation

$$G(\varphi, J) = \sum_{|k| \leqslant K} g_k(J) e^{2\pi i (k \cdot \phi)}$$

where $g_k(J)$ are smooth functions which are sufficiently close to

$$\frac{(1 - \varrho_k(J)) h_k(J)}{(2\pi i) k \cdot \partial_J H_0(J)}$$

in the C^3 norm.

Let Φ^t be the Hamiltonian flow generated by εG. Setting

$$F_t = R_1 + R_+ + t(H_1 - R_1 - R_+),$$

we have the standard computation

$$\begin{aligned} \partial_t((H_0 + \varepsilon F_t) \circ \Phi^t) &= \varepsilon \partial_t F_t \circ \Phi^t + \varepsilon \{H_0 + \varepsilon F_t, G\} \circ \Phi^t \\ &= \varepsilon(\partial_t F_t + \{H_0, G\}) \circ \Phi^t + \varepsilon^2 \{F_t, G\} \circ \Phi^t = \varepsilon^2 \{F_t, G\} \circ \Phi^t, \end{aligned}$$

from which it follows that

$$H_\varepsilon \circ \Phi^1 = H_0 + \varepsilon R_1 + \varepsilon R_+ + \varepsilon^2 \int_0^1 \{F_t, G\} \circ \Phi^t \, dt.$$

Let us estimate the C^2 norm of the function $R_2 := R_+ + \varepsilon \int_0^1 \{F_t, G\} \circ \Phi^t \, dt$. It follows from Lemma 2.3 that

$$\|R_+\|_{C^2} \leqslant \varkappa_m K^{-r+m+2} \|H_1\|_{C^r} \leqslant \tfrac{1}{2}\delta.$$

We now focus on the term $\int_0^1 \{F_t, G\}\circ\Phi^t\,dt$. To estimate the norm of F_t, it is convenient to write $F_t=\widetilde{F}_t+(1-t)R_1$, where $\widetilde{F}_t=(1-t)R_+ + tH_1$. Notice that the Fourier expansion of $\widetilde{F}_t$ is simply a constant times that of H_1, Lemma 2.3 then implies that

$$\|\widetilde{F}_t\|_{C^3} \leqslant \sum_{k\in\mathbb{Z}^m} [k]^{3-r}\|H_1\|_{C^r} = \varkappa_m \|H_1\|_{C^r}$$

provided that $r\geqslant m+4$, where we set $\varkappa_m=\sum_{\mathbb{Z}^m}[k]^{-m-1}$.

We now have to estimate the norm of R_1 and G. This requires additional estimates of the smoothing terms ϱ_k as well as the small denominators $k\cdot\partial_J H_0$. We always assume that $l\in\{0,1,2,3\}$ in the following estimates:

- if $\varrho_k(J)\neq 1$ then $|(k\cdot\partial_J H_0)^{-1}|\leqslant \beta^{-1}\varepsilon^{-1/4}|k|^{-1}$;
- $\|(k\cdot\partial_J H_0)^{-1}\|_{C^l}\leqslant c_m\beta^{-l-1}\varepsilon^{-(l+1)/4}\|H_0\|_{C^4}^{l+1}$ on $\{\varrho_k\neq 1\}$;
- $\|\varrho_k(J)\|_{C^l}\leqslant c_m\beta^{-l}\varepsilon^{-l/4}\|H_0\|_{C^4}^{l}$ and $\|1-\varrho_k(J)\|_{C^l}\leqslant c_m\beta^{-l}\varepsilon^{-l/4}\|H_0\|_{C^4}^{l}$.

We have been using the following estimates on the derivative of composition of functions: For $f\colon\mathbb{R}^m\to\mathbb{R}$ and $g\colon\mathbb{R}^m\to\mathbb{R}^m$ we have $\|f\circ g\|_{C^l}\leqslant c_{m,l}\|f\|_{C^l}(1+\|g\|_{C^l}^l)$.

- For each multi-index α, with $|\alpha|\leqslant 3$, we have that

$$\begin{aligned}
&\|\partial_{J^\alpha}((1-\varrho_k(J))h_k(J)(k\cdot\partial_J H_0)^{-1})\|_{C^0}\\
&\qquad\leqslant \sum_{\alpha_1+\alpha_2+\alpha_3=\alpha} \|1-\varrho_k(J)\|_{C^{|\alpha_1|}}\|h_k\|_{C^{|\alpha_2|}}\|(k\cdot\partial_J H_0)^{-1}\|_{C^{|\alpha_3|}(\{\varrho_k\neq 1\})}\\
&\qquad\leqslant c_m \sum_{\alpha_1+\alpha_2+\alpha_3=\alpha}\Big(\beta^{-|\alpha_1|}\varepsilon^{-|\alpha_1|/4}\|H_0\|_{C^4}^{|\alpha_1|}[k]^{-r+|\alpha_2|}\|H_1\|_{C^r}\\
&\qquad\qquad\qquad\qquad\times\beta^{-|\alpha_3|-1}\varepsilon^{-(|\alpha_3|+1)/4}\|H_0\|_{C^4}^{|\alpha_3|+1}\Big)\\
&\qquad\leqslant c_m\beta^{-|\alpha|-1}\varepsilon^{-(|\alpha|+1)/4}[k]^{|\alpha|-r}\|H_0\|_{C^4}^{|\alpha|+1}\|H_1\|_{C^r}.
\end{aligned}$$

In these computations, we have used the hypothesis $\beta\varepsilon^{1/4}\leqslant\|H_0\|_{C^4}$. Since

$$G(\varphi,J)=\sum_{k\in\mathbb{Z}^m}(1-\varrho_k(J))h_k(J)(k\cdot\partial_J H_0)^{-1}e^{2\pi i(k\cdot\varphi)},$$

Lemma 2.3 implies (since $r\geqslant m+1$) the following:

- $\|G\|_{C^l}\leqslant c_m\beta^{-l-1}\varepsilon^{-(l+1)/4}\|H_0\|_{C^4}^{l+1}\|H_1\|_{C^r}\leqslant\varepsilon^{-1}$.

We now turn our attention to $R_1=\sum_{|k|\leqslant K}\varrho_k(J)h_k(J)e^{2\pi i(k\cdot\phi)}$, getting

- $\|h_k\|_{C^l}\leqslant[k]^{l-r}\|H_1\|_{C^r}$.
- $\|\varrho_k h_k\|_{C^l}\leqslant c_m\beta^{-l}\varepsilon^{-l/4}[k]^{-r+l}\|H_0\|_{C^4}^{l}\|H_1\|_{C^r}$.
- $\|R_1\|_{C^l}\leqslant c_m\beta^{-l}\varepsilon^{-l/4}\|H_0\|_{C^4}^{l}\|H_1\|_{C^r}$, provided $r\geqslant m+4$.

We obtain

$$\|F_t\|_{C^l}\leqslant\|R_1\|_{C^l}+\|\widetilde{F}_t\|_{C^l}\leqslant c_m\beta^{-l}\varepsilon^{-l/4}\|H_0\|_{C^4}^{l}\|H_1\|_{C^r}$$

and

$$\|\{F_t,G\}\|_{C^2}\leqslant\sum_{|\alpha_1+\alpha_2|\leqslant 3}\|F_t\|_{C^{|\alpha_1|}}\|G\|_{C^{|\alpha_2|}}\leqslant c_m\beta^{-4}\varepsilon^{-1}\|H_0\|^4_{C^4}\|H_1\|^2_{C^r}.$$

Concerning the flow Φ^t, we observe that $\|\varepsilon G\|_{C^3}\leqslant 1$, and get the following estimates (see, e.g., [28, Lemma 3.15]):

– $\|\Phi^t-\mathrm{id}\,\|_{C^2}\leqslant c_m\varepsilon\|G\|_{C^3}\leqslant c_m\beta^{-4}\|H_0\|^4_{C^4}\|H_1\|_{C^r}\leqslant\delta$;

– $\|\Phi^t-\mathrm{id}\,\|_{C^0}\leqslant c_m\varepsilon\|G\|_{C^1}\leqslant c_m\beta^{-2}\sqrt{\varepsilon}\|H_0\|^2_{C^4}\|H_1\|_{C^2}\leqslant\delta\sqrt{\varepsilon}$.

Finally, we obtain

$$\varepsilon\|\{F_t,G\}\circ\Phi^t\|_{C^2}\leqslant c_m\varepsilon\|\{F_t,G\}\|_{C^2}\|\Phi^t\|^2_{C^2}\leqslant c_m\beta^{-4}\|H_0\|^4_{C^4}\|H_1\|^2_{C^r}\leqslant\tfrac{1}{2}\delta.\qquad\square$$

2.2. Normal form away from additional resonances

We now return to our non-autonomous system and apply Theorem 2.2 around the resonance under study. With the non-autonomous Hamiltonian

$$H_\varepsilon(\theta,p,t)=H_0(p)+\varepsilon H_1(\theta,p,t)\colon\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}\longrightarrow\mathbb{R}$$

we associate the autonomous Hamiltonian

$$\widetilde{H}_e(\varphi,J)=H_0(I)+e+\varepsilon H_1(\theta,I,t)\colon\mathbb{T}^{n+1}\times\mathbb{R}^{n+1}\longrightarrow\mathbb{R},$$

where $\varphi=(\theta,t)$ and $J=(I,e)$. We denote the frequencies $\omega\in\mathbb{R}^{n+1}$ by $\omega=(\omega^f,\omega^s,\omega^t)\in\mathbb{R}^{n-1}\times\mathbb{R}\times\mathbb{R}$, and define the set

$$\Omega(K,s):=\{\omega\in\mathbb{R}^{n+1}:\|\omega^s\|>s\text{ and }|k^f\omega^f+k^t\omega^t|\geqslant 3sK\text{ for all }(k^s,k^t)\in\mathbb{Z}^2_K\},$$

where $\mathbb{Z}^2_K$ denotes the set of pairs (k^f,k^t) of integers such that $0<\max\{k^f,k^t\}\leqslant K$. Note that

$$\mathcal{D}(K,s)=\{p\in\mathbb{R}^n:(\partial_pH_0(p),1)\in\Omega(K,s)\}.$$

COROLLARY 2.4. *There exists a constant $c_n>0$, which depends only on n, such that the following holds. Given*

- *a C^4 Hamiltonian $H_0(p)$,*
- *a C^r Hamiltonian $H_1(\theta,p,t)$ with $\|H_1\|_{C^r}=1$,*
- *parameters $r\geqslant n+5$, $\delta\in]0,1[$, $\varepsilon\in]0,1[$, $\beta>0$, $K>0$,*

satisfying

- $K\geqslant c_n\delta^{-1/(r-n-4)}$,
- $\beta\geqslant c_n(1+\|H_0\|_{C^4})\delta^{-1/2}$,
- $\beta\varepsilon^{1/4}\leqslant\|H_0\|_{C^4}$,

there exists a C^2 symplectic diffeomorphism $\tilde\Phi$ of $\mathbb{T}^{n+1}\times\mathbb{R}^{n+1}$ such that, in the new coordinates, the Hamiltonian $H_\varepsilon=H_0+\varepsilon H_1$ takes the form

$$N_\varepsilon=H_0+\varepsilon Z+\varepsilon R,$$

with

- $\|R\|_{C^2}\leqslant\delta$ *on* $\mathbb{T}^n\times\mathcal{D}(K,\beta\varepsilon^{1/4})\times\mathbb{T}$,
- $\|\tilde\Phi-\mathrm{id}\,\|_{C^0}\leqslant\delta\sqrt{\varepsilon}$ *and* $\|\tilde\Phi-\mathrm{id}\,\|_{C^2}\leqslant\delta$.

The symplectic diffeomorphism $\tilde\Phi$ is of the form

$$\tilde\Phi(\theta,p,t,e)=(\Phi(\theta,p,t),e+f(\theta,p,t))$$

where Φ is a diffeomorphism of $\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$ fixing the last variable t. The maps $\tilde\Phi$ and Φ are smooth if H_0 and H_1 are.

Proof. We apply Theorem 2.2 with $\widetilde H_\varepsilon$, $m=n+1$ and $\tilde\delta=\frac12\delta$. We get a diffeomorphism $\tilde\Phi$ of $\mathbb{T}^{n+1}\times\mathbb{R}^{n+1}$ as the time-1 flow of the Hamiltonian G. By inspection in the proof of Theorem 2.2, we observe that G does not depend on e, which implies that $\tilde\Phi$ has the desired form. We have

$$\widetilde H_\varepsilon\circ\tilde\Phi=\widetilde H_0(J)+\varepsilon\widetilde R_1+\varepsilon\widetilde R_2,$$

where $\|\widetilde R_2\|_{C^2}\leqslant\frac12\delta$ and

$$\widetilde R_1(\theta,p,t)=\sum_{[k]\leqslant K}\varrho\biggl(\frac{k^f\cdot\partial_{p^f}H_0+k^s\partial_{p^s}H_0+k^t}{\beta\varepsilon^{1/4}[k]}\biggr)g_k(p)e^{2\pi ik\cdot(\theta,t)}.$$

Let us compute this sum under the assumption that $p\in\mathcal{D}(K,\beta\varepsilon^{1/4})$ (or equivalently, that $(\partial_pH_0,1)\in\Omega(K,\beta\varepsilon^{1/4})$). We have

$$\left|\frac{k^f\cdot\partial_{p^f}H_0}{\beta\varepsilon^{1/4}[k]}\right|\leqslant1,$$

and hence

$$\varrho\biggl(\frac{k^f\cdot\partial_{p^f}H_0+k^s\partial_{p^s}H_0+k^t}{\beta\varepsilon^{1/4}[k]}\biggr)=1$$

for k such that $k^s=0=k^t$. For the other terms, we have, by definition of $\Omega(K,s)$,

$$\left|\frac{k^s\partial_{p^s}H_0+k^t}{\beta\varepsilon^{1/4}[k]}\right|\geqslant\left|\frac{k^s\partial_{p^s}H_0+k^t}{\beta\varepsilon^{1/4}K}\right|\geqslant3,$$

and hence

$$\left|\frac{k^f\cdot\partial_{p^f}H_0+k^s\partial_{p^s}H_0+k^t}{\beta\varepsilon^{1/4}[k]}\right|\geqslant2$$

and these terms vanish in the expansion of $\widetilde{R}_1$. We conclude that

$$\widetilde{R}_1(\theta,p,t)=\sum_{k^f\in\mathbb{Z}^{n-1},[k^f]\leqslant K} g_{(k_f,0,0)}(p)e^{2\pi i k^f\cdot\theta^f}$$

hence $\widetilde{R}_1=Z-\Pi^+_K(Z)$, with the notation of Lemma 2.3. Finally $\widetilde{H}_\varepsilon\circ\tilde{\Phi}=\widetilde{H}_0+\varepsilon Z+\varepsilon R_2$ with $R_2=\widetilde{R}_2-\Pi^+_K Z$. From Lemma 2.3, we see that

$$\|\Pi^+_K Z\|_{C^2}\leqslant c_n K^{m+3-r}\|Z\|_{C^r}\leqslant c_n K^{m+3-r}\|H_1\|_{C^r}\leqslant c_n K^{m+3-r}\leqslant \tfrac{1}{2}\delta.$$

On the other hand, $\|\widetilde{R}_2\|_{C^2}\leqslant\frac{1}{2}\delta$, and hence $\|R_2\|_{C^2}\leqslant\delta$. □

2.3. Smooth approximation

We finally remove the restriction on r and obtain a smooth change of coordinates. If $r<n+5$, we use Lemma 2.5 below to approximate H_1 by an analytic function $H_1^*:=S_\tau H_1$ (with a parameter τ that will be specified later).

LEMMA 2.5. ([60]) *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a C^r function, with $r\geqslant 4$. Then, for each $\tau>0$, there exists an analytic function $S_\tau f$ such that*

$$\|S_\tau f-f\|_{C^3}\leqslant c(n,r)\|f\|_{C^3}\tau^{r-3}\quad \textit{and}\quad \|S_\tau f\|_{C^s}\leqslant c(n,r)\|f\|_{C^s}\tau^{-(s-r)}$$

for each $s>r$, where $c(n,r)$ is a constant which depends only on n and r.

In order to obtain a smooth change of variables, it is also convenient to approximate $H_0(p)$ in $C^4(B)$ by a smooth $H_0^*(p)$ (using a standard mollification). We then apply Corollary 2.4 to the Hamiltonian

$$H_\varepsilon^*:=H_0^*+\varepsilon H_1^*=H_0^*+\varepsilon_2 H_2$$

with $H_2=H_1^*/\|H_1^*\|_{C^{r_2}}$, $\varepsilon_2=\varepsilon\|H_1^*\|_{C^{r_2}}$, and some parameters $r_2\geqslant r$ and $\delta_2\leqslant\delta$ to be specified later. We find a smooth change of coordinates $\tilde{\Phi}$ such that

$$\widetilde{H}_\varepsilon^*\circ\tilde{\Phi}=\widetilde{H}_0^*+\varepsilon_2 Z_2+\varepsilon_2 R_2=\widetilde{H}_0^*+\varepsilon Z^*+\varepsilon\|H_1^*\|_{C^{r_2}}R_2$$

and $\|R_2\|_{C^2}\leqslant\delta_2$, where $Z_2(\theta^s,p)=\int H_2\,d\theta^f\,dt$ and $Z^*(\theta^s,p)=\int H_1^*\,d\theta^f\,dt$. As usual, we have denoted by $\widetilde{H}_\varepsilon^*$ and $\widetilde{H}_0^*$ the autonomized Hamiltonians $\widetilde{H}_\varepsilon^*=H_\varepsilon^*+e$ and $\widetilde{H}_0^*=H_0^*+e$, respectively. With the same map $\tilde{\Phi}$, we obtain

$$\widetilde{H}_\varepsilon\circ\tilde{\Phi}=\widetilde{H}_0+\varepsilon Z+\varepsilon R$$

with

$$R = \|H_1^*\|_{C^{r_2}} R_2 + (Z - Z^*) + (H_1^* - H_1)\circ\Phi + \frac{\widetilde{H}_0^* - \widetilde{H}_0 + (\widetilde{H}_0 - \widetilde{H}_0^*)\circ\tilde{\Phi}}{\varepsilon}.$$

In the expression above, the map Φ is the trace on the (θ, p, t) variables of the map $\tilde{\Phi}$. Choosing $\tau = \delta_2^{1/(r_2-3)}$, and assuming that $\|H_0^* - H_0\|_{C^4} \leqslant \varepsilon\delta/c(n,4)$ we get

– $\|H_1^* - H_1\|_{C^3} \leqslant c(n, r_2)\delta_2^{(r-3)/(r_2-3)}$,
– $\|H_1^*\|_{C^{r_2}} \leqslant c(n, r_2)\delta_2^{-(r_2-r)/(r_2-3)}$,
– $\|Z^* - Z\|_{C^2} \leqslant \|H_1^* - H_1\|_{C^2} \leqslant c(n, r_2)\delta_2^{(r-3)/(r_2-3)}$,
– $\|\tilde{\Phi} - \mathrm{id}\,\|_{C^2} \leqslant \delta_2 \leqslant \delta \leqslant 1$,
– $\|(H_1^* - H_1)\circ\Phi\|_{C^2} \leqslant c(n, r_2)\|H_1^* - H_1\|_{C^2}(\|\Phi\|_{C^2} + \|\Phi\|_{C^2}^2) \leqslant c(n,5)\|H_1^* - H_1\|_{C^2}$,
– $\|(\widetilde{H}_0 - \widetilde{H}_0^*)\circ\tilde{\Phi}\|_{C^2} \leqslant \delta/c(n, r_2)$,

and finally

$$\|R\|_{C^2} \leqslant c(n, r_2)\delta_2^{(r-3)/(r_2-3)} + \frac{\delta}{c(n, r_2)}.$$

We now set

$$\delta_2 = \frac{\delta^{(r_2-3)/(r-3)}}{c(n, r_2)} \leqslant \frac{\delta}{2}$$

and get $\|R\|_{C^2} \leqslant \delta$. To apply Corollary 2.4 as we just did, we need the following conditions to hold on the parameters:

– $K \geqslant c(n, r_2)\delta^{(r_2-3)/(r-3)(r_2-n-4)}$, which implies $K \geqslant c_n\delta_2^{-1/(r-n-4)}$,
– $\beta \geqslant c(n, r_2)(2 + \|H_0\|_{C^4})\delta^{-(r_2-3)/2(r-3)}$ which implies $\beta \geqslant c_n(1 + \|H_0^*\|_{C^4})\delta_2^{-1/2}$,
– $\beta\varepsilon^{1/4} \leqslant (1 + \|H_0\|_{C^4})\delta^{(r_2-r)/4(r-3)}$ which implies $\beta\varepsilon_2^{1/4} \leqslant \|H_0^*\|_{C^4}$.

We apply the above discussion with $r_2 = 2n+5$ and get Theorem 2.1. Note the estimate

$$\|\,\mathrm{id} - \tilde{\Phi}\|_{C^0} \leqslant \delta_2\sqrt{\varepsilon_2} \leqslant \delta_2^{1-(r_2-r)/2(r_2-3)}\sqrt{\varepsilon} \leqslant \sqrt{\varepsilon}.$$

3. Normally hyperbolic cylinders

In this section, we study the C^2 Hamiltonian

$$N_\varepsilon(\theta, p, t) = H_0(p) + \varepsilon Z(\theta^s, p) + \varepsilon R(\theta, p, t).$$

In the above notations we denote by $p_*^s(p^f) \in \mathbb{R}^{n-1}$ the solution of the equation

$$\partial_{p^s} H_0(p_*^s(p^f), p^f) = 0.$$

We recall also the notation $p_*(p^f) := (p_*^s(p^f), p^f)$ from the introduction. We assume that $\|Z\|_{C^3} \leqslant 1$, and that $D^{-1}I \leqslant \partial_{pp}^2 H_0 \leqslant D\,I$ for some $D \geqslant 1$. To simplify the notation, we will be using the $O(\cdot)$ notation, where $f = O(g)$ means $|f| \leqslant Cg$ for a constant C independent of ε, λ, δ, r, a^-, and a^+. We will not be keeping track of the parameter D, which is considered fixed throughout the paper.

Given parameters

$$\lambda \in]0,1] \quad \text{and} \quad a^- < a^+,$$

we assume that for each $p^f \in [a^-, a^+]$ there exists a local maximum $\theta^s_*(p^f)$ of the map $\theta^s \mapsto Z(\theta^s, p_*(p^f))$, and that θ^s_* is a C^2 function of p^f. We assume in addition that

$$-I \leqslant \partial^2_{\theta^s\theta^s} Z(\theta^s_*(p^f), p_*(p^f)) \leqslant -\lambda I \tag{7}$$

for each $p^f \in [a^-, a^+]$, where as before I is the identity matrix. We shall at some occasions lift the map θ^s_* to a C^2 map taking values in $\mathbb{R}^{n-1}$ without changing its name.

THEOREM 3.1. *The following conclusion holds if* $b \in]0,1[$ *is a sufficiently small constant (how small does not depend on the parameters* ε, λ, δ, a^-, a^+*): If the parameters* $\lambda \in]0,1]$, $a^- < a^+$, ε, *and* δ *satisfy*

$$0 < \varepsilon < b\lambda^{9/2} \quad \text{and} \quad 0 \leqslant \delta < b\lambda^{5/2},$$

if $\|R\|_{C^2} \leqslant \delta$, *on the open set*

$$\{(\theta, p, t) : p^f \in]a^-, a^+[\ \text{and} \ \|p^s - p^s_*(p^f)\| < \varepsilon^{1/2}\}, \tag{8}$$

and if (7) *holds for each* $p^f \in [a^-, a^+]$, *then there exists a* C^2 *map*

$$(\Theta^s, P^s)(\theta^f, p^f, t) : \mathbb{T} \times [a^- + \sqrt{\delta\varepsilon}, a^+ - \sqrt{\delta\varepsilon}\,] \times \mathbb{T} \longrightarrow \mathbb{T}^{n-1} \times \mathbb{R}^{n-1}$$

such that the cylinder

$$\mathcal{C} = \{(\theta^s, p^s) = (\Theta^s, P^s)(\theta^f, p^f, t) : p^f \in [a^- + \sqrt{\delta\varepsilon}, a^+ - \sqrt{\delta\varepsilon}] \ \text{and} \ (\theta^f, t) \in \mathbb{T} \times \mathbb{T}\}$$

is weakly invariant with respect to N_ε *in the sense that the Hamiltonian vector field is tangent to* $\mathcal{C}$. *The cylinder* $\mathcal{C}$ *is contained in the set*

$$V := \{(\theta, p, t) : p^f \in [a^- + \sqrt{\delta\varepsilon}, a^+ - \sqrt{\delta\varepsilon}], \|\theta^s - \theta^s_*(p^f)\| \leqslant b^{1/5}\lambda^{3/2} \\ \text{and} \ \|p^s - p^s_*(p^f)\| \leqslant b^{1/5}\lambda^{3/2}\varepsilon^{1/2}\},$$

and it contains all the full orbits of N_ε *contained in* V. *We have the estimates*

$$\|\Theta^s(\theta^f, p^f, t) - \theta^s_*(p^f)\| \leqslant O(\lambda^{-1}\delta + \lambda^{-3/4}\sqrt{\varepsilon}),$$
$$\|P^s(\theta^f, p^f, t) - p^s_*(p^f)\| \leqslant \sqrt{\varepsilon}\, O(\lambda^{-3/4}\delta + \lambda^{-1/2}\sqrt{\varepsilon}),$$

$$\left\|\frac{\partial \Theta^s}{\partial p^f}\right\| = O\left(\frac{\lambda^{-2}\sqrt{\varepsilon} + \lambda^{-5/4}\sqrt{\delta}}{\sqrt{\varepsilon}}\right), \qquad \left\|\frac{\partial \Theta^s}{\partial(\theta^f, t)}\right\| = O(\lambda^{-2}\sqrt{\varepsilon} + \lambda^{-5/4}\sqrt{\delta}),$$
$$\left\|\frac{\partial P^s}{\partial p^f}\right\| = O(1), \qquad \left\|\frac{\partial P^s}{\partial(\theta^f, t)}\right\| = O(\sqrt{\varepsilon}\,).$$

Notice that the domain V is contained in the domain (8) where the assumption on R is made.

Proof of Theorem 1.3. We derive Theorem 1.3 from Theorem 3.1 as follows. We assume that Hypothesis ($HZ\lambda$) holds on

$$\Gamma_1 := \{p_*(p^f) : p^f \in [a_-, a_+]\}.$$

Then the inequality

$$-I \leqslant \partial^2_{\theta^s\theta^s} Z(\theta^s_*(p^f), p_*(p^f)) \leqslant -2\lambda I$$

holds for $p^f \in [a_-, a_+]$. Since $\|Z\|_{C^3} \leqslant 1$, the inequality

$$-I \leqslant \partial^2_{\theta^s\theta^s} Z(\theta^s, p) \leqslant -\lambda I$$

holds for each (θ^s, p) in the λ-neighborhood of $(\theta^s_*(a_-), p_*(a_-))$. The inequality

$$Z(\theta^s, p_*(a_-)) \leqslant Z(\theta^s_*(a_-), p_*(a_-)) - \lambda d^2(\theta^s, \theta^s_*(a_-))$$

implies that the function $Z(\,\cdot\,, p_*(p^f))$ has a global maximum $\theta^s_*(p^f)$, which is contained in the ball $B(\theta^s_*(a_-), \lambda)$, provided $|p^f - a_-| \leqslant b\lambda^3$ and b is small enough. By a similar reasoning at a_+, we extend the map $p^f \mapsto \theta^s_*(p^f)$ to the interval $[a_- - b\lambda^3, a_+ + b\lambda^3]$ in such a way that, for each p^f in this interval, the point $\theta^s_*(p^f)$ is a local (and even global) maximum of the function $Z(\,\cdot\,, p_*(p^f))$ which satisfies the inequalities

$$-I \leqslant \partial^2_{\theta^s\theta^s} Z(\theta^s_*(p^f), p_*(p^f)) \leqslant -\lambda I.$$

Taking a small $b > 0$, we set $\varkappa = b^{1/5}\lambda^{3/2}$ and $\delta = b^3\lambda^9$. Assuming as in the statement of Theorem 1.3 that the estimate $\|R\|_{C^2} < \delta$ holds on $\mathbb{T}^n \times U_{\varepsilon^{1/3}} \times \mathbb{T}$, hence on

$$\big\{(\theta, p, t) : p^f \in \big]a_- - \tfrac{1}{2}\varepsilon^{1/3}, a_+ + \tfrac{1}{2}\varepsilon^{1/3}\big[\text{ and } \|p^s - p^s_*(p^f)\| < \tfrac{1}{2}\varepsilon^{1/3}\big\},$$

and that $\varepsilon \in \big]0, \delta\big[$, we apply Theorem 3.1 on the interval

$$[a^-, a^+] := \big[a_- - \tfrac{1}{2}\varepsilon^{1/3}, a_+ + \tfrac{1}{2}\varepsilon^{1/3}\big] \subset [a_- - b\lambda^3, a_+ + b\lambda^3].$$

If b (hence $\varkappa$) is small enough, then we have the inclusion

$$[a^- + \sqrt{\varepsilon\delta}, a^+ - \sqrt{\varepsilon\delta}\,] \supset [a_- - \varkappa\varepsilon^{1/3}, a_+ + \varkappa\varepsilon^{1/3}]. \qquad \square$$

The proof of Theorem 3.1 occupies the rest of the section.

The Hamiltonian flow admits the following equation of motion:

$$\begin{cases} \dot\theta^s = \partial_{p^s} H_0 + \varepsilon \partial_{p^s} Z + \varepsilon \partial_{p^s} R, \\ \dot p^s = -\varepsilon \partial_{\theta^s} Z - \varepsilon \partial_{\theta^s} R, \\ \dot\theta^f = \partial_{p^f} H_0 + \varepsilon \partial_{p^f} Z + \varepsilon \partial_{p^f} R, \\ \dot p^f = -\varepsilon \partial_{\theta^f} R, \\ \dot t = 1. \end{cases} \tag{9}$$

The Hamiltonian structure of the flow is not used in the following proof.

It is convenient in the sequel to lift the angular variables to real variables and to consider the above system as defined on $\mathbb{R}^{n-1}\times\mathbb{R}^{n-1}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$. We will see this system as a perturbation of the model system

$$\dot\theta^s = \partial_{p^s} H_0, \quad \dot p^s = -\varepsilon\partial_{\theta^s} Z, \quad \dot\theta^f = \partial_{p^f} H_0, \quad \dot p^f = 0, \quad \dot t = 1. \tag{10}$$

The graph of the map

$$(\theta^f, p^f, t) \longmapsto (\theta^s_*(p^f), p^s_*(p^f))$$

on $\mathbb{R}\times]a^-, a^+[\times\mathbb{R}$ is obviously invariant for the model flow. For each fixed p^f, the point $(\theta^s_*(p^f), p^s_*(p^f))$ is a hyperbolic fixed point of the partial system

$$\dot\theta^s = \partial_{p^s} H_0(p^s, p^f), \quad \dot p^s = -\varepsilon\partial_{\theta^s} Z(\theta^s, p^s, p^f),$$

where p^f is seen as a parameter. This hyperbolicity is the key property we will use, through the theory of normally hyperbolic invariant manifolds. It is not obvious to apply this theory here because the model system itself depends on ε, and because we have to deal with the problem of non-invariant boundaries. We will however manage to apply the quantitative version exposed in Appendix A.

We perform some changes of coordinates in order to put the system in the framework of Appendix A. These coordinates appear naturally from the study of the model system as follows. We set

$$B(p^f) := \partial^2_{p^s p^s} H_0(p_*(p^f)) \quad \text{and} \quad A(p^f) := -\partial^2_{\theta^s\theta^s} Z(\theta^s_*(p^f), p_*(p^f)).$$

If we fix the variable p^f and consider the model system in (θ^s, p^s), we observed that this system has a hyperbolic fixed point at $(\theta^s_*(p^f), p^s_*(p^f))$. The linearized system at this point is

$$\dot\theta^s = B(p^f)\, p^s, \quad \dot p^s = \varepsilon A(p^f)\, \theta^s.$$

To put this system under a simpler form, it is useful to consider the matrix

$$T(p^f) := (B^{1/2}(p^f)(B^{1/2}(p^f)A(p^f)B^{1/2}(p^f))^{-1/2}B^{1/2}(p^f))^{1/2},$$

which is symmetric, positive definite, and satisfies $T^2(p^f)A(p^f)T^2(p^f)=B(p^f)$, as can be checked by a direct computation. We finally introduce the symmetric positive definite matrix

$$\Lambda(p^f) := T(p^f)A(p^f)T(p^f) = T^{-1}(p^f)B(p^f)T^{-1}(p^f).$$

In the new variables

$$\xi = T^{-1}(p^f)\theta^s + \varepsilon^{-1/2}T(p^f)p^s \quad \text{and} \quad \eta = T^{-1}(p^f)\theta^s - \varepsilon^{-1/2}T(p^f)p^s,$$

the linearized system is reduced to the following block-diagonal form:

$$\dot{\xi} = \varepsilon^{1/2}\Lambda(p^f)\xi, \quad \dot{\eta} = -\varepsilon^{1/2}\Lambda(p^f)\eta,$$

see [10] for more details. This leads us to introduce the following set of new coordinates for the full system

$$x = T^{-1}(p^f)(\theta^s - \theta^s_*(p^f)) + \varepsilon^{-1/2}T(p^f)(p^s - p^s_*(p^f)),$$
$$y = T^{-1}(p^f)(\theta^s - \theta^s_*(p^f)) - \varepsilon^{-1/2}T(p^f)(p^s - p^s_*(p^f)),$$

$$I = \varepsilon^{-1/2}p^f \quad \text{and} \quad \Theta = \gamma\theta^f,$$

where γ is a parameter which will be taken later equal to $\delta^{1/2}$. Note that

$$\theta^s = \theta^s_*(\varepsilon^{1/2}I) + \tfrac{1}{2}T(\varepsilon^{1/2}I)(x+y) \quad \text{and} \quad p^s = p^s_*(\varepsilon^{1/2}I) + \tfrac{1}{2}\varepsilon^{1/2}T^{-1}(\varepsilon^{1/2}I)(x-y).$$

LEMMA 3.2. *We have* $\Lambda(p^f) \geqslant \sqrt{\lambda/D}\, I$ *for each* $p^f \in [a^-, a^+]$.

Proof. The matrix Λ is symmetric, hence it satisfies $\Lambda \geqslant \lambda_* I$, where $\lambda_* > 0$ is its smallest eigenvalue. The real number λ_* is then an eigenvalue of the matrix

$$\begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{bmatrix} \quad \text{which is similar to} \quad \begin{bmatrix} 0 & B \\ A & 0 \end{bmatrix}.$$

Since both A and B are square matrices of equal size, we conclude that λ_*^{-2} is an eigenvalue of $A^{-1}B^{-1}$. Since $\|A^{-1}\| \leqslant \lambda^{-1}$ and $\|B^{-1}\| \leqslant D$, we have

$$\lambda_*^{-2} \leqslant \|A^{-1}B^{-1}\| \leqslant D\lambda^{-1}.$$

We conclude that $\lambda_* \geqslant \sqrt{\lambda/D}$. □

The links between the various parameters ε, δ, γ, λ, and ϱ which appear in the computations below will be specified later. We will however assume from the beginning that

$$\delta \leqslant \varrho \leqslant \lambda < 1, \quad \sqrt{\varepsilon} \leqslant \varrho < 1, \quad \text{and} \quad 0 < \gamma \leqslant \lambda < 1.$$

Let us first collect some estimates that will be useful to see that the system (9) is indeed a perturbation of the model system.

LEMMA 3.3. *We have the estimates*

$$\|T\| = O(\lambda^{-1/4}), \quad \|T^{-1}\| = O(1), \quad \|\partial_{p^f} T\| \leqslant O(\lambda^{-5/4}), \quad \|\partial_{p^f} T^{-1}\| \leqslant O(\lambda^{-3/4}),$$
$$\|\partial_{p^f} \theta^s_*\| \leqslant O(\lambda^{-1}), \quad \|p^s_*\|_{C^2} = O(1), \quad \|\theta^s - \theta^s_*\| \leqslant O(\lambda^{-1/4}\varrho), \quad \|p^s - p^s_*\| \leqslant O(\varepsilon^{1/2}\varrho),$$

where $\varrho = \max\{\|x\|, \|y\|\}$.

Proof. We recall that

$$T = (B^{1/2}(B^{1/2}AB^{1/2})^{-1/2}B^{1/2})^{1/2} \quad \text{and} \quad T^{-1} = (B^{-1/2}(B^{1/2}AB^{1/2})^{1/2}B^{-1/2})^{1/2}.$$

As $D^{-1}I \leqslant B \leqslant D\,I$ and $\lambda I \leqslant A \leqslant I$, we obtain that $\|T\| \leqslant O(\lambda^{-1/4})$ and that $\|T^{-1}\| \leqslant O(1)$. To estimate the derivative of T, we consider the map $F: M \mapsto M^{1/2}$ defined on positive symmetric matrices. It is known that that

$$dF_M \cdot N = \int_0^\infty e^{-tM^{1/2}} N e^{-tM^{1/2}}\, dt.$$

To verify this one can diagonalize M, perform integration, and match terms in $(M^{1/2} + \varepsilon dF_M \cdot N)(M^{1/2} + \varepsilon dF_M \cdot N) = M + \varepsilon N + O(\varepsilon^2)$. This implies that

$$\|dF_M\| \leqslant \tfrac{1}{2}\|M^{1/2}\|^{-1} \leqslant \tfrac{1}{2}\|M^{-1/2}\|.$$

As a consequence, if $M(p_f)$ is a positive symetric matrix depending on p_f, we have

$$\|\partial_{p^f} M\| \leqslant \tfrac{1}{2}\|M^{-1/2}\|\,\|\partial_{p^f} M\|.$$

We apply this bound several times to estimate $\partial_{p^f} T$ and $\partial_{p^f} T^{-1}$. In our situation, we have $\partial_{p^f} A = O(1)$ and $\partial_{p^f} B = O(1)$. Using $M = A$ and B, we get $\partial_{p^f}(A^{1/2}) = O(\lambda^{-1/2})$ and $\partial_{p^f}(B^{1/2}) = O(1)$, respectively. Using $M = B^{1/2}AB^{1/2}$, we get

$$\partial_{p^f}[(B^{1/2}AB^{1/2})^{1/2}] = O(\lambda^{-1/2}),$$

and then

$$\begin{aligned}\|\partial_{p^f}[T^{-1}]\| &\leqslant \|(B^{-1/2}(B^{1/2}AB^{1/2})^{1/2}B^{-1/2})^{-1/2}\|\,\|\partial_{p^f}[B^{-1/2}(B^{1/2}AB^{1/2})^{1/2}B^{-1/2}]\| \\ &= O(\lambda^{-1/4})O(\lambda^{-1/2}) = O(\lambda^{-3/4}).\end{aligned}$$

Recalling that

$$\|\partial_{p^f}(M^{-1})\| \leqslant \|M^{-1}\|^2\, \|\partial_{p^f} M\|,$$

we obtain (with $M = T^{-1}$)

$$\|\partial_{p^f} T\| \leqslant \|T\|^2\, \|\partial_{p^f}[T^{-1}]\| \leqslant \partial_{p^f}(M^{1/2}) = O(\lambda^{-5/4}).$$

The other estimates are straightforward. □

COROLLARY 3.4. *Let $\widetilde{V}$ be the image in the (x, y, I, Θ, t) coordinates of the domain called V in the statement. We have*

$$\widetilde{V} \subset \{x : \|x\| \leqslant b^{1/6}\lambda^{5/4}\} \times \{y : \|y\| \leqslant b^{1/6}\lambda^{5/4}\} \times \mathbb{R} \times \left[\frac{a^-}{\sqrt{\varepsilon}} + \sqrt{\delta}, \frac{a^+}{\sqrt{\varepsilon}} - \sqrt{\delta}\right] \times \mathbb{R},$$
$$\widetilde{V} \supset \{x : \|x\| \leqslant 2b^{1/4}\lambda^{7/4}\} \times \{y : \|y\| \leqslant 2b^{1/4}\lambda^{7/4}\} \times \mathbb{R} \times \left[\frac{a^-}{\sqrt{\varepsilon}} + \sqrt{\delta}, \frac{a^+}{\sqrt{\varepsilon}} - \sqrt{\delta}\right] \times \mathbb{R}$$

provided b is small enough.

From now on, we work on the region

$$p^f \in [a^-, a^+], \quad \|x\| \leqslant \varrho, \quad \|y\| \leqslant \varrho.$$

In view of Lemma 3.3, this region is contained in the (image in the new coordinates of the) domain where the inequality $\|R\|_{C^2} \leqslant \delta$ was assumed.

LEMMA 3.5. *The equations of motion in the new coordinates take the form*

$$\dot{x} = -\sqrt{\varepsilon}\Lambda(\sqrt{\varepsilon}I)x + \varepsilon^{1/2} O(\lambda^{-1/4}\delta + \lambda^{-3/4}\varrho^2) + O(\varepsilon),$$
$$\dot{y} = \sqrt{\varepsilon}\Lambda(\sqrt{\varepsilon}I)y + \varepsilon^{1/2} O(\lambda^{-1/4}\delta + \lambda^{-3/4}\varrho^2) + O(\varepsilon),$$
$$\dot{I} = O(\sqrt{\varepsilon}\delta),$$

where $\varrho = \max\{\|x\|, \|y\|\}$ is assumed to satisfy $\varrho \leqslant \lambda$. The expression for $\dot{\Theta}$ is not useful here.

Proof. The last part of the statement is obvious. We prove the part concerning $\dot{x}$, the calculations for $\dot{y}$ are exactly the same. In the original coordinates the vector field (9) can be written as

$$\dot{\theta}^s = B(p^f)(p^s - p^s_*(p^f)) + O(\|p^s - p^s_*(p^f)\|^2) + O(\varepsilon),$$
$$\dot{p}^s = \varepsilon A(p^f)(\theta^s - \theta^s_*(p^f)) + O(\varepsilon\|\theta^s - \theta^s_*(p^f)\|^2) + O(\varepsilon\delta).$$

As a consequence, we have

$$\begin{aligned}\dot{x} = & T^{-1}B(p^s - p^s_*) + \varepsilon^{1/2}TA(\theta^s - \theta^s_*)\\ & + T^{-1}\cdot O(\|p^s - p^s_*\|^2 + \varepsilon) + \varepsilon^{1/2}T\cdot O(\|\theta^s - \theta^s_*\|^2 + \delta)\\ & + (\partial_{p^f}T^{-1})\dot{p}^f(\theta^s - \theta^s_*) + \varepsilon^{-1/2}(\partial_{p^f}T)\dot{p}^f(p^s - p^s_*)\\ & - T^{-1}(\partial_{p^f}\theta^s_*)\dot{p}^f - \varepsilon^{-1/2}T(\partial_{p^f}p^s_*)\dot{p}^f.\end{aligned}$$

We use the estimates of Lemma 3.3 to simplify (recall also that $\dot{p}^f = O(\varepsilon\delta)$):

$$\begin{aligned}\dot{x} = & T^{-1}B(p^s - p^s_*) + \varepsilon^{1/2}TA(\theta^s - \theta^s_*)\\ & + O(\varepsilon\varrho^2 + \varepsilon) + O(\varepsilon^{1/2}\lambda^{-3/4}\varrho^2 + \varepsilon^{1/2}\lambda^{-1/4}\delta)\\ & + O(\lambda^{-1}\varepsilon\delta\varrho) + O(\lambda^{-5/4}\varepsilon\delta\varrho) + O(\lambda^{-1}\varepsilon\delta + \lambda^{-1/4}\varepsilon^{1/2}\delta).\end{aligned}$$

□

Lemma 3.6. *In the new coordinate system (x, y, Θ, I, t), the linearized system is given by the matrix*

$$L = \begin{bmatrix}\sqrt{\varepsilon}\Lambda & 0 & 0 & 0 & 0\\ 0 & -\sqrt{\varepsilon}\Lambda & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix} + O(\sqrt{\varepsilon}\delta\lambda^{-1/4}\gamma^{-1} + \sqrt{\varepsilon}\lambda^{-3/4}\varrho + \varepsilon\lambda^{-5/4} + \sqrt{\varepsilon}\gamma),$$

where $\varrho = \max\{\|x\|, \|y\|\}$.

Proof. Most of the estimates below are based on Lemma 3.3. In the original coordinates, the matrix of the linearized system is:

$$\tilde{L} = \begin{bmatrix}O(\varepsilon) & \partial^2_{p^s p^s}H_0 + O(\varepsilon) & 0 & \partial^2_{p^f p^s}H_0 + O(\varepsilon) & 0\\ -\varepsilon\partial^2_{\theta^s\theta^s}Z & O(\varepsilon) & 0 & O(\varepsilon) & 0\\ O(\varepsilon) & O(1) & 0 & O(1) & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix} + O(\delta\varepsilon),$$

In our notations we have

$$\tilde{L} = \begin{bmatrix}O(\varepsilon) & B + O(\varepsilon + \sqrt{\varepsilon}\varrho) & 0 & \partial^2_{p^f p^s}H_0 + O(\varepsilon) & 0\\ \varepsilon A + O(\varepsilon\lambda^{-1/4}\varrho) & O(\varepsilon) & 0 & O(\varepsilon) & 0\\ O(\varepsilon) & O(1) & 0 & O(1) & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix} + O(\delta\varepsilon),$$

In the new coordinates, the matrix is the product

$$L=\left[\frac{\partial(x,y,\Theta,I,t)}{\partial(\theta^s,p^s,\theta^f,p^f,t)}\right]\cdot\tilde{L}\cdot\left[\frac{\partial(\theta^s,p^s,\theta^f,p^f,t)}{\partial(x,y,\Theta,I,t)}\right].$$

We have

$$\left[\frac{\partial(\theta^s,p^s,\theta^f,p^f,t)}{\partial(x,y,\Theta,I,t)}\right]=\begin{bmatrix}\frac{1}{2}T & \frac{1}{2}T & 0 & O(\sqrt{\varepsilon}\lambda^{-1}) & 0\\ \frac{1}{2}\sqrt{\varepsilon}T^{-1} & -\frac{1}{2}\sqrt{\varepsilon}T^{-1} & 0 & \sqrt{\varepsilon}\partial_{p^f}p^s_*+O(\varepsilon\lambda^{-3/4}\varrho) & 0\\ 0 & 0 & \gamma^{-1} & 0 & 0\\ 0 & 0 & 0 & \sqrt{\varepsilon} & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix},$$

and hence

$$\tilde{L}\left[\frac{\partial(\theta^s,p^s,\theta^f,p^f,t)}{\partial(x,y,\Theta,I,t)}\right]=O(\gamma^{-1}\delta\varepsilon)$$
$$+\begin{bmatrix}\frac{1}{2}\sqrt{\varepsilon}BT^{-1}+O(\varepsilon\lambda^{-1/4}) & -\frac{1}{2}\sqrt{\varepsilon}BT^{-1}+O(\varepsilon\lambda^{-1/4}) & 0 & O(\varepsilon\lambda^{-3/4}\varrho+\varepsilon^{3/2}\lambda^{-1}) & 0\\ \frac{1}{2}\varepsilon AT+O(\varepsilon\lambda^{-1/2}\varrho) & \frac{1}{2}\varepsilon AT+O(\varepsilon\lambda^{-1/2}\varrho) & 0 & \varepsilon^{3/2}O(\lambda^{-5/4}\varrho+\lambda^{-1}) & 0\\ O(\sqrt{\varepsilon}) & O(\sqrt{\varepsilon}) & 0 & O(\sqrt{\varepsilon}) & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}.$$

This expression is the result of a tedious, but straightforward, computation. Let us just detail the computation of the coefficient on the first line, fourth row, which contains an important cancellation:

$$\sqrt{\varepsilon}\partial^2_{p^sp^s}H_0\partial_{p^f}p^s_*+\sqrt{\varepsilon}\partial^2_{p^fp^s}H_0+O(\varepsilon\lambda^{-3/4}\varrho+\varepsilon^{3/2}\lambda^{-1})$$
$$=\sqrt{\varepsilon}\partial_{p^f}(\partial_{p^s}H_0(p_*(p^f)))+O(\varepsilon\lambda^{-3/4}\varrho+\varepsilon^{3/2}\lambda^{-1})=O(\varepsilon\lambda^{-3/4}\varrho+\varepsilon^{3/2}\lambda^{-1}).$$

We now write

$$\left[\frac{\partial(x,y,\Theta,I,t)}{\partial(\theta^s,p^s,\theta^f,p^f,t)}\right]=\begin{bmatrix}T^{-1} & \varepsilon^{-1/2}T & 0 & O(\varepsilon^{-1/2}\lambda^{-1/4}) & 0\\ T^{-1} & -\varepsilon^{-1/2}T & 0 & O(\varepsilon^{-1/2}\lambda^{-1/4}) & 0\\ 0 & 0 & \gamma & 0 & 0\\ 0 & 0 & 0 & \varepsilon^{-1/2} & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix},$$

and compute that

$$L=\begin{bmatrix}\sqrt{\varepsilon}\Lambda+O(\sqrt{\varepsilon}\lambda^{-3/4}\varrho) & O(\sqrt{\varepsilon}\lambda^{-3/4}\varrho) & 0 & O(\varepsilon\lambda^{-5/4}) & 0\\ O(\sqrt{\varepsilon}\lambda^{-3/4}\varrho) & -\sqrt{\varepsilon}\Lambda+O(\sqrt{\varepsilon}\lambda^{-3/4}\varrho) & 0 & O(\varepsilon\lambda^{-5/4}) & 0\\ O(\sqrt{\varepsilon}\gamma) & O(\sqrt{\varepsilon}\gamma) & 0 & O(\sqrt{\varepsilon}\gamma) & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$
$$+O(\sqrt{\varepsilon}\delta\lambda^{-1/4}\gamma^{-1}).$$

□

In order to prove the existence of a normally hyperbolic invariant strip (for the lifted system), we apply Proposition A.7 to the system in coordinates (x, y, Θ, I, t). More precisely, with the notations of appendix A, we set $u=x$, $s=y$, $c_1=(\Theta, t)$, $c_2=I$, and consider the domain

$$\Omega = \mathbb{R}^2 \times \Omega^{c_2} = \mathbb{R}^2 \times \left[\frac{a^-}{\sqrt{\varepsilon}} + \sqrt{\delta}, \frac{a^+}{\sqrt{\varepsilon}} - \sqrt{\delta}\right].$$

We fix

$$\gamma = \sqrt{\delta}, \quad \alpha = \sqrt{\frac{\varepsilon\lambda}{4D}}, \quad \sigma = \sqrt{\delta}, \tag{11}$$

observe that $\sqrt{\varepsilon}\Lambda \geqslant 2\alpha I$, by Lemma 3.2. We assume, as in the statement of the theorem, that $0<\varepsilon<b\lambda^{9/2}$ and that $0\leqslant\delta<b\lambda^{5/2}$. We apply Proposition A.7 with $B^u=\{u:\|u\|\leqslant\varrho\}$ and $B^s=\{s:\|s\|\leqslant\varrho\}$ under the constraint

$$b^{-1/4}(\lambda^{-3/4}\delta + \lambda^{-1/2}\sqrt{\varepsilon}) \leqslant \varrho \leqslant b^{1/6}\lambda^{5/4}, \tag{12}$$

provided $b\in]0,1[$ is small enough. Observe that, if b is small enough, the inequalities

$$b^{-1/4}(\lambda^{-3/4}\delta + \lambda^{-1/2}\sqrt{\varepsilon}\,) \leqslant 2b^{1/4}\lambda^{7/4} \leqslant b^{1/6}\lambda^{5/4}$$

holds under our assumptions on the parameters, hence values of ϱ satisfying (12) do exist. It is easy to check under our assumptions on the parameters that such values of ϱ exist. Let us check the isolating block condition under the condition (12). By Lemma 3.5, we have

$$\dot{x}\cdot x \geqslant 2\alpha\|x\|^2 - \|x\|\; O(\varepsilon^{1/2}\lambda^{-1/4}\delta + \varepsilon^{1/2}\lambda^{-3/4}\varrho^2 + \varepsilon)$$

if $x\in B^u$ and $y\in B^s$. If, in addition, $\|x\|=\varrho$, then

$$\lambda^{-3/4}\delta \leqslant b^{1/4}\|x\|, \quad \lambda^{-5/4}\varrho^2 \leqslant b^{1/6}\|x\|, \quad \sqrt{\varepsilon/\lambda} \leqslant b^{1/4}\|x\|,$$

and hence

$$\dot{x}\cdot x \geqslant 2\alpha\|x\|^2 - \|x\|^2 b^{1/6} O(\sqrt{\varepsilon\lambda}\,) \geqslant \alpha\|x\|^2,$$

provided b is small enough. Similarly, $\dot{y}\cdot y \leqslant -\alpha\|y\|^2$ on $B^u\times\partial B^s$, provided b is small enough. Concerning the linearized system, we have

$$L_{uu} = \sqrt{\varepsilon}\Lambda + O(\sqrt{\varepsilon}\delta\lambda^{-1/4}\gamma^{-1} + \sqrt{\varepsilon}\lambda^{-3/4}\varrho + \varepsilon\lambda^{-5/4} + \sqrt{\varepsilon}\gamma) = \sqrt{\varepsilon}\Lambda + O(b^{1/6}\sqrt{\varepsilon\lambda}) \geqslant \alpha I,$$
$$L_{ss} = -\sqrt{\varepsilon}\Lambda + O(b^{1/6}\sqrt{\varepsilon\lambda}) \leqslant -\alpha I$$

on $B^u \times B^s \times \Omega_r$. These inequalities holds when b is small enough because $\sqrt{\varepsilon}\Lambda \geqslant 2\alpha I$ and $\sqrt{\varepsilon\lambda} \leqslant O(\alpha)$. Finally, still with the notations of Proposition A.7, we can take

$$\begin{aligned} m &= O\left(\sqrt{\varepsilon}\delta\lambda^{-1/4}\gamma^{-1} + \sqrt{\varepsilon}\lambda^{-3/4}\varrho + \varepsilon\lambda^{-5/4} + \sqrt{\varepsilon}\gamma + \frac{\sqrt{\varepsilon}\delta}{\sigma}\right) \\ &= \sqrt{\varepsilon\lambda}\, O(\sqrt{\delta}\lambda^{-3/4} + \varrho\lambda^{-5/4} + \sqrt{\varepsilon}\lambda^{-7/4}) = \sqrt{\varepsilon\lambda}\, O(b^{1/6}). \end{aligned} \tag{13}$$

If b is small enough, we have $16m < \alpha$, and hence

$$K \leqslant \frac{2m}{\alpha} < \frac{1}{8},$$

and Proposition A.7 can be applied. The invariant strip obtained from the proof of Proposition A.7 does not depend on the choice of ϱ, as long as (12) holds. It contains all the full orbits contained in

$$\{x : \|x\| \leqslant b^{1/6}\lambda^{5/4}\} \times \{y : \|y\| \leqslant b^{1/6}\lambda^{5/4}\} \times \mathbb{R} \times \left[\frac{a^-}{\sqrt{\varepsilon}} + \sqrt{\delta}, \frac{a^+}{\sqrt{\varepsilon}} - \sqrt{\delta}\right] \times \mathbb{R} \supset \widetilde{V},$$

where $\widetilde{V}$ is the image in the new coordinates of the domain V defined in the statement of Theorem 3.1 and where the last inclusion holds provided b is small enough, as follows from Corollary 3.4. So our invariant strip contains all the full orbits contained in $\widetilde{V}$. On the other hand, we can take $\varrho = 2b^{1/4}\lambda^{7/4}$, and since

$$\{x : \|x\| \leqslant 2b^{1/4}\lambda^{7/4}\} \times \{y : \|y\| \leqslant 2b^{1/4}\lambda^{7/4}\} \times \mathbb{R} \times \left[\frac{a^-}{\sqrt{\varepsilon}} + \sqrt{\delta}, \frac{a^+}{\sqrt{\varepsilon}} - \sqrt{\delta}\right] \times \mathbb{R} \subset \widetilde{V}$$

(still for b small enough, by Corollary 3.4), our invariant strip is contained in $\widetilde{V}$.

The possibility of taking $\varrho = b_1^{-1/4}(\lambda^{-3/4}\delta + \lambda^{-1/2}\sqrt{\varepsilon}\,)$ now implies that the cylinder is actually contained in the domain where

$$\|x\|, \|y\| \leqslant b_1^{-1/4}(\lambda^{-3/4}\delta + \lambda^{-1/2}\sqrt{\varepsilon}\,).$$

Moreover, with this choice of ϱ and using that $K = O(m/\sqrt{\varepsilon\lambda}\,)$, we can obtain an improved estimate of the Lipschitz constant K (notation from the appendix):

$$\begin{aligned} K &= O(\sqrt{\delta}\lambda^{-3/4} + \varrho\lambda^{-5/4} + \sqrt{\varepsilon}\lambda^{-7/4}) \\ &= O(\sqrt{\delta}\lambda^{-3/4} + b_1^{-1/4}\delta\lambda^{-2} + b_1^{-1/4}\sqrt{\varepsilon}\lambda^{-7/4} + \sqrt{\varepsilon}\lambda^{-7/4}) \\ &= O(\sqrt{\delta}\lambda^{-3/4} + \sqrt{\delta}\lambda^{-1} + b_1^{-1/4}\sqrt{\varepsilon}\lambda^{-7/4}) \\ &= O(\sqrt{\delta}\lambda^{-1} + b_1^{-1/4}\sqrt{\varepsilon}\lambda^{-7/4}). \end{aligned}$$

Observe finally that, since the system is $(1/\gamma)$-periodic in Θ and 1-periodic in t, so is the invariant strip given by Proposition A.7. We have obtained the existence of a C^1 map

$$w^c=(w^c_u,w^c_s)\colon(\Theta,I,t)\in\mathbb{R}\times\left[\frac{a^-}{\sqrt{\varepsilon}}+\sqrt{\delta},\frac{a^+}{\sqrt{\varepsilon}}-\sqrt{\delta}\right]\times\mathbb{R}\longrightarrow\mathbb{R}^{n-1}\times\mathbb{R}^{n-1}$$

which is $2K$-Lipschitz, $(1/\gamma)$-periodic in Θ, 1-periodic in t, and the graph of which is tangent to the vector field. Our last task is to return to the original coordinates by setting

$$\begin{aligned}\Theta^s(\theta^f,p^f,t)&=\theta^s_*(p^f)+\tfrac{1}{2}T(p^f)\cdot(w^c_u+w^c_s)(\gamma\theta^f,\varepsilon^{-1/2}p^f,t),\\ P^s(\theta^f,p^f,t)&=p^s_*(p^f)+\tfrac{1}{2}\sqrt{\varepsilon}T^{-1}(p^f)\cdot(w^c_u-w^c_s)(\gamma\theta^f,\varepsilon^{-1/2}p^f,t).\end{aligned}\tag{14}$$

All the estimates stated in Theorem 3.1 follow directly from these expressions, and from the fact that $\|dw^c\|\leqslant 2K$. This concludes the proof of Theorem 3.1.

4. Localization and Mather's projected graph theorem

We study the system in normal form $N_\varepsilon=H_0+\varepsilon Z+\varepsilon R$ of Theorem 1.2 from the point of view of Mather theory at a fixed cohomology $c\in\mathbb{R}^n$ such that $\partial_{p^s}H_0(c)=0$ (or in other words such that $c\in\Gamma$). We assume that $\|Z\|_{C^2}\leqslant 1$, and that $\|R\|_{C^2}\leqslant\delta$ on $\{\|p-c\|<\varepsilon^{1/3}\}$. We continue to assume (1), and, for simplicity, we assume that D is large enough and ε small enough for the following inequality to also hold:

$$\frac{1}{D}I\leqslant\partial^2_pN_\varepsilon\leqslant DI.$$

Most of our statement depend on the shape of the function $Z_c\colon\theta^s\longmapsto Z(\theta,c)$. We will most of the time assume that $(HZ\lambda)$ holds at c: There exists θ^s_* such that

$$Z(\theta^s,c)\leqslant Z(\theta^s_*,c)-\lambda d^2(\theta^s,\theta^s_*).$$

We will rewrite this inequality as

$$\widehat{Z}_c(\theta^s)\leqslant-\lambda d^2(\theta^s,\theta^s_*),$$

with the notation $\widehat{Z}_c=Z_c-\max Z_c$. Later in §4.4, we also consider the double peak case, which is not necessary for the proof of Theorem 1.2, but is very natural. Our first statement localizes the Mañé set.

THEOREM 4.1. *In the single peak case* (*when* $(HZ\lambda)$ *holds at* c), *if* $\delta>0$ *is small enough with respect to* n, D, λ, *and* ε *is small enough with respect to* n, D, λ, δ, *then the Mañé set at cohomology* c *of the Hamiltonian* N_ε *satisfies*

$$s\widetilde{\mathcal{N}}(c)\subset B(\theta^s_*,\delta^{1/5})\times\mathbb{T}\times B(c,\sqrt{\varepsilon}\delta^{1/16})\times\mathbb{T}\subset\mathbb{T}^{n-1}\times\mathbb{T}\times\mathbb{R}^n\times\mathbb{T}.$$

This statement is proved in §4.2. Our second statement is a quantitative version of the celebrated Mather Lipschitz graph theorem, it does not rely on any particular assumption on Z, besides $\|Z\|_{C^2}\leqslant 1$:

THEOREM 4.2. *For each weak* KAM *solution* u *of* N_ε *at cohomology* c, *the set* $\widetilde{\mathcal{I}}(u,c)\subset\mathbb{T}^n\times\mathbb{R}^n$ *is contained in a* $9\sqrt{D\varepsilon}$*-Lipschitz graph above* $\mathbb{T}^n$.

This theorem is proved in §4.3. We will always assume in this section that δ is sufficiently small with respect to n, H_0 and λ, and that ε is sufficiently small with respect to n, H_0, λ and δ.

4.1. Some inequalities

We will denote by N the Hamiltonian N_ε and by L the associated Lagrangian function, which is defined by

$$L(\theta,v,t)=\max_{p\in\mathbb{R}^n}(p\cdot v-N(\theta,p,t)).$$

The function L is then C^2, and the maps

$$(\theta,p,t)\longmapsto\partial_pN(\theta,p,t)\quad\text{and}\quad(\theta,v,t)\longmapsto\partial_vL(\theta,v,t)$$

are diffeomorphisms of $\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T}$, which are inverses of each other. The maximum in the definition of L is reached at $p=\partial_vL(\theta,v,t)$. Since $I/D\leqslant\partial_{pp}N\leqslant DI$, we have

$$\frac{I}{D}\leqslant\partial_{vv}L\leqslant DI.$$

We will also denote by $L_0(v)$ the Lagrangian associated with H_0, or more explicitly $L_0(v):=\sup_p(p\cdot v-H_0(p))$. It satisfies

$$\frac{I}{D}\leqslant\partial_{vv}L_0\leqslant DI.$$

LEMMA 4.3. *For each* $\varrho\in[4D\varepsilon,\varepsilon^{1/4}]$, *the image of the open set* $\mathbb{T}^n\times B(c,\varrho)\times\mathbb{T}$ *under the diffeomorphism* ∂_pN *contains the set*

$$\mathbb{T}^n\times B\biggl(\partial_pH_0(c),\frac{\varrho}{2D}-2\varepsilon\biggr)\times\mathbb{T}.$$

In particular, if ε *is small enough, the image of* $\mathbb{T}^n\times B(c,\varepsilon^{1/4})\times\mathbb{T}$ *contains*

$$\mathbb{T}^n\times B\biggl(c,\frac{\varepsilon^{1/4}}{4D}\biggr)\times\mathbb{T}.$$

Proof. In view of the estimate $\partial_p^2 H \geqslant I/D$, each of the applications $p \mapsto \partial_p N(\theta, p, t)$ sends the ball $B(c, r)$ to a set which contains the ball $B(\partial_p N(\theta, c, t), r/2D)$. Since

$$|\partial_p N(\theta, c, t) - \partial_p H_0(c)| \leqslant \varepsilon + \varepsilon\delta \leqslant 2\varepsilon,$$

we conclude that the image contains $B(\partial_p H_0(c), \varrho/2D - 2\varepsilon)$. □

LEMMA 4.4. *The estimates*

$$\|\partial_{\theta v} L\|_{C^0} \leqslant 2D\varepsilon \quad \textit{and} \quad \|\partial_{\theta\theta} L\|_{C^0} \leqslant 3\varepsilon$$

hold on $\mathbb{T}^n \times B(c, \varepsilon^{1/3}/4D) \times \mathbb{T}$.

Proof. Note first that the estimates

$$\|\partial_{\theta p} H\| \leqslant 2\varepsilon \quad \text{and} \quad \|\partial_{\theta\theta} H\| \leqslant 2\varepsilon$$

hold on the domain $\mathbb{T}^n \times B(c, \varepsilon^{1/3}) \times \mathbb{T}$, which contains the image of

$$\mathbb{T}^n \times B\left(c, \frac{\varepsilon^{1/3}}{4D}\right) \times \mathbb{T}$$

under $\partial_v L$. Observing that $\partial_\theta L = -\partial_\theta N(\theta, \partial_v L_\varepsilon(\theta, v))$, which implies that

$$\partial_{v\theta} L(\theta, v, t) = -\partial_{p\theta} N_\varepsilon(\theta, \partial_v L(\theta, v, t), t)\partial_{vv} L(\theta, v, t)$$

we deduce that $\|\partial_{\theta v} L\| \leqslant 2D\varepsilon$ on $\mathbb{T}^n \times B(c, \varepsilon^{1/3}/4D) \times \mathbb{T}$. The equality

$$\partial_{\theta\theta} L(\theta, v, t) = -\partial_{\theta\theta} N(\theta, \partial_v L(\theta, v, t), t) - \partial_{p\theta} N(\theta, \partial_v L(\theta, v, t), t)\partial_{\theta v} L(\theta, v, t),$$

implies that $\|\partial_{\theta\theta} L\| \leqslant 2\varepsilon + (2\varepsilon)(2D\varepsilon)$ on $\mathbb{T}^n \times B(c, \varepsilon^{1/3}/4D) \times \mathbb{T}$. □

LEMMA 4.5. *We have the estimate*

$$|L(\theta, v, t) - (L_0(v) - \varepsilon Z(\theta^s, c))| \leqslant 2\varepsilon\delta$$

if $|v - \partial_p H_0(c)| < \varepsilon^{1/3}/4D$.

Proof. On the domain $\{|p - c| < \varepsilon^{1/3}\}$, we have

$$|N(\theta, p, t) - (H_0(p) + \varepsilon Z(\theta^s, c))| \leqslant \varepsilon^{5/4} + \varepsilon\delta \leqslant 2\varepsilon\delta.$$

If $|v - \partial_p H_0(c)| < \varepsilon^{1/3}/4D$, then, by Lemma 4.3,

$$L(\theta, v, p) = \sup_{|p-c|<\varepsilon^{1/3}} [p \cdot v - N(\theta, p, t)]$$

and, by Lemma 4.3 applied with $R \equiv 0$ and $Z(\theta^s, p) \equiv Z(\theta^s, c)$,

$$L_0(v) - \varepsilon Z(\theta^s, c) = \sup_p (p \cdot v - H_0(p) - \varepsilon Z(\theta^s, c)) = \sup_{|p-c|<\varepsilon^{1/3}} (p \cdot v - H_0(p) - \varepsilon Z(\theta^s, c)). \quad \square$$

Let us now estimate the value $\alpha(c)$ of the Mather function of N. We use the notation $Z_c(\theta^s):=Z(\theta^s,c)$.

LEMMA 4.6. *The value $\alpha(c)$ of the Mather function of N satisfies*

$$|\alpha(c)-(H_0(c)+\varepsilon\max_{\theta^s} Z_c)|\leqslant 2\varepsilon\delta.$$

The reason behind this inequality is that the value $\alpha(c)$ of the Hamiltonian $H_0+\varepsilon Z_c$ is $H_0(c)+\varepsilon\max_{\theta^s} Z_c$.

Proof. On one hand, we have

$$\alpha(c)\leqslant \max_{(t,\theta)} N_\varepsilon(t,\theta,c)\leqslant H_0(c)+\varepsilon\max_{\theta^s} Z_c+\varepsilon \max_{(t,\theta)\in\mathbb{T}^{n+1}} R(\theta,c,t)\leqslant H_0(c)+\varepsilon\max_{\theta^s} Z_c+\varepsilon\delta.$$

For the other inequality, we use that $\partial_{p^s}H_0=0$. We consider the Haar measure μ of the torus $\mathbb{T}\times\{\theta^s_*(c)\}\times\{\partial H_0(c)\}\times\mathbb{T}$, where $\theta^s_*(c)$ is any point maximizing Z_c. This measure is not necessarily invariant under the Lagrangian flow of L, but it is invariant under the Lagrangian flow of L_0-Z_c (because $\partial_{p^s}H_0=0$) hence it is closed, which means that

$$\int(\partial_t f+\partial_\theta f\cdot v)\, d\mu(\theta,v,t)=0$$

for each smooth function $f(t,\theta)$. See [4] and [32] (both inspired by [47]) for the notion of closed measures. Each closed measure μ has a rotation vector $\varrho(\mu):=\int v\, d\mu(\theta,v,t)\in\mathbb{R}^n$, and its action is not less than $c\cdot\varrho(\mu)-\alpha(c)$. Here, we have $\varrho(\mu)=\partial_p H_0(c)$, and hence

$$\begin{aligned}\alpha(c)\geqslant c\cdot\partial_p H_0(c)-\int L\, d\mu &= c\cdot\partial_p H_0(c)-L_0(\omega)+\varepsilon Z_c(\theta^s_*(c))-2\varepsilon\delta\\ &= H_0(c)+\varepsilon\max_{\theta^s} Z_c-2\varepsilon\delta.\end{aligned}$$

□

LEMMA 4.7. *If ε is small enough (with respect to D and δ), we have the estimates*

$$L(\theta,v,t)-c\cdot v+\alpha(c)\geqslant\frac{\|v-\partial H_0(c)\|^2}{4D}-\varepsilon\widehat{Z}_c(\theta^s)-4\varepsilon\delta, \tag{15}$$

$$L(\theta,v,t)-c\cdot v+\alpha(c)\leqslant D\|v-\partial H_0(c)\|^2-\varepsilon\widehat{Z}_c(\theta^s)+4\varepsilon\delta, \tag{16}$$

for each $(\theta,v,t)\in\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{R}$, where $\widehat{Z}_c(\theta^s):=Z(\theta^s,c)-\max_{\theta^s} Z(\theta^s,c)$.

Proof. It is a direct computation:

$$\begin{aligned}L(\theta,v,t)&\geqslant c\cdot v-N(\theta,c,t)+\frac{\|v-\partial_p N(\theta,c,t)\|^2}{2D}\\ &\geqslant c\cdot v-H_0(c)-\varepsilon Z_c(\theta^s)-\varepsilon\delta+\frac{(\|v-\partial_p H_0(c))\|-2\varepsilon)^2}{2D}\\ &\geqslant c\cdot v-\alpha(c)+\varepsilon(\max_{\theta^s} Z_c-Z_c(\theta^s))-3\varepsilon\delta+\tfrac{\|v-\partial_p H_0(c))\|^2}{4D}-16\varepsilon^2,\end{aligned}$$

$$\begin{aligned}L(\theta,v,t)&\leqslant c\cdot v-N(\theta,c,t)+\tfrac12 D\|v-\partial_p N(\theta,c,t)\|^2\\ &\leqslant c\cdot v-H_0(c)-\varepsilon Z_c(\theta^s)+\varepsilon\delta+\tfrac12 D(\|v-\partial_p H_0(c))\|+2\varepsilon)^2\\ &\leqslant c\cdot v-\alpha(c)+\varepsilon(\max_{\theta^s} Z_c-Z_c(\theta^s))+3\varepsilon\delta+D\|v-\partial_p H_0(c))\|^2+8D\varepsilon^2.\end{aligned}$$

□

It is useful to consider suspended weak KAM solutions. Recall that we defined weak KAM solutions associated with a Lagrangian L at cohomolgy c as functions u on $\mathbb{T}^n$ such that, for each $t\in\mathbb{N}$,

$$u(\theta)=\inf_{\gamma}\biggl(u(\gamma(0))+\int_0^t (L(\gamma(s),\dot\gamma(s),s)-c\cdot\dot\gamma(s)+\alpha(c))\,ds\biggr),$$

where the infimum is taken over the set of C^1 curves $\gamma\colon\mathbb{R}\to\mathbb{T}^n$ such that $\gamma(t)=\theta$. We can similarly define suspended weak KAM solutions as functions $u\colon\mathbb{T}^n\times\mathbb{T}\longrightarrow\mathbb{R}$ such that

$$u(\theta,T \bmod 1)=\inf_{\gamma}\biggl(u(\gamma(S),S \bmod 1)+\int_S^T (L(\gamma(t),\dot\gamma(t),t)+c\cdot\dot\gamma(t))\,dt\biggr),$$

for all real times $S\leqslant T$, where the infimum is taken over the space of C^1 curves $\gamma\colon[S,T]\to\mathbb{T}^n$ such that $\gamma(T)=\theta$. There is a bijection between suspended weak KAM solutions $u(\theta,t)$ and genuine weak KAM solutions: Each suspended weak KAM solution $u(\theta,t)$ restricts to a genuine weak KAM solution $u(\theta)=u(\theta,0)$, and each genuine weak KAM solution $u(\theta)$ is the restriction of a unique suspended weak KAM solution $u(\theta,t)$ which can be defined by

$$u(\theta,t \bmod 1)=\inf_{\gamma}\biggl(u(\gamma(0))+\int_0^t (L(\gamma(s),\dot\gamma(s),s)+c\cdot\dot\gamma(s)+\alpha(c))\,ds\biggr),$$

for each $t>0$, where the infimum is taken on C^1 curves $\gamma\colon\mathbb{R}\longrightarrow\mathbb{T}^n$ such that $\gamma(t)=\theta$. We shall use the same notation for a weak KAM solution u and the associated suspended weak KAM solution. Curves γ calibrated by the weak KAM solutions $u(\theta)$ are also calibrated by the corresponding suspended weak KAM solution in the sense that

$$u(\gamma(t_2),t_2 \bmod 1)-u(\gamma(t_1),t_1 \bmod 1)=\int_{t_1}^{t_2}(L(\gamma(s),\dot\gamma(s),s)+c\cdot\dot\gamma(s)+\alpha(c))\,ds$$

for each time interval $[t_1,t_2]$. Let us now estimate the oscillation $\operatorname{osc} u:=\max u-\min u$ of suspended weak KAM solutions. We consider a convex subset $\Omega\subset\mathbb{T}^{n-1}$, meaning that it is the projection of a convex subset $\widetilde\Omega$ of $\mathbb{R}^{n-1}$, of diameter less than $2\sqrt{n}$.

LEMMA 4.8. *Let $u(\theta,t)$ be a suspended weak KAM solution of N at cohomology c. Given two points $(\theta_1,t_1),(\theta_2,t_2)\in\mathbb{T}\times\Omega\times\mathbb{T}$, we have*

$$u(\theta_2,t_2)-u(\theta_1,t_1)\leqslant 10\sqrt{nD\varepsilon(m+4\delta)},$$

where $m:=-\inf_\Omega\widehat Z_c$. We can in particular take $\Omega=\mathbb{T}^{n-1}$, then $m\leqslant1$ and we conclude that $\operatorname{osc} u\leqslant 10\sqrt{2nD\varepsilon}$.

Proof. We have $0\geqslant \widehat{Z}_c\geqslant -m$ on Ω. We take two points (θ_i,t_i), $i=1,2$, in the domain $\mathbb{T}\times\Omega\times\mathbb{T}$, and consider the curve

$$\theta(t)=\theta_1+(t-\tilde t_1)\frac{\tilde\theta_2-\tilde\theta_1+[(T+\tilde t_2-\tilde t_1)\partial H_0(c)]}{T+\tilde t_2-\tilde t_1},$$

where $T\in\mathbb{N}$ is a parameter to be fixed later, $\tilde t_i\in[0,1[$, and $\tilde\theta_i\in[0,1[\times\widetilde\Omega$ are representatives of the angular variables t_i and θ_i, and $[\omega]\in\mathbb{Z}^n$ is the component-wise integral part of ω. Note that $\theta(\tilde t_1)=\theta_1$ and $\theta(\tilde t_2+T)=\theta_2$, hence

$$\begin{aligned}
u(\theta_2,t_2)-u(\theta_1,t_1) &\leqslant \int_{\tilde t_1}^{\tilde t_2+T}(L(\theta(t),\dot\theta(t),t)-c\cdot\dot\theta(t)+\alpha(c))\,dt\\
&\leqslant \int_{\tilde t_1}^{\tilde t_2+T}(D\|\dot\theta-\partial H_0(c)\|^2-\varepsilon\widehat{Z}_c(\theta^s(t))+4\varepsilon\delta)\,dt\\
&\leqslant \int_{\tilde t_1}^{\tilde t_2+T}\left(\frac{9Dn}{(T+\tilde t_2-\tilde t_1)^2}+\varepsilon m+4\varepsilon\delta\right)dt\\
&\leqslant \frac{9Dn}{T+\tilde t_2-\tilde t_1}+(T+\tilde t_2-\tilde t_1)\varepsilon(m+4\delta).
\end{aligned}$$

This inequality holds for all $T\in\mathbb{N}$, in particular, we can choose $T\in\mathbb{N}$ so that

$$2\sqrt{\frac{nD}{\varepsilon(m+4\delta)}}\leqslant T+\tilde t_2-\tilde t_1\leqslant 3\sqrt{\frac{nD}{\varepsilon(m+4\delta)}}$$

and obtain $u(\theta_2,t_2)-u(\theta_1,t_1)\leqslant 10\sqrt{nD\varepsilon(m+4\delta)}$. □

4.2. Localization of the invariant sets

We prove Theorem 4.1. It is enough to prove that the inclusion

$$s\widetilde{\mathcal{I}}(u,c)\subset B(\theta^s_*(c),\delta^{1/5})\times\mathbb{T}\times B(c,\sqrt{\varepsilon}\delta^{1/16})\times\mathbb{T}$$

holds for each (suspended) weak KAM solution u. We fix such a solution $u(\theta,t)$ and prove the inclusion. The following preliminary localization, which does not use any assumption on the shape of Z, implies that the set $s\widetilde{\mathcal{I}}(u,c)$ is contained (when ε is small enough) in the domain $\{\|p-c\|<\varepsilon^{1/3}\}$ where the assumption $\|R\|_{C^2}\leqslant\delta$ is made.

Lemma 4.9. *Let $(\theta,p)\colon[t_1,t_2]\to\mathbb{T}^n\times\mathbb{R}^n$ be an orbit calibrated by u. If t_1 and t_2 are such that $t_2-t_1\geqslant\varepsilon^{-1/2}$, then*

$$\|p(t)-c\|\leqslant C\sqrt{\varepsilon}$$

for each $t\in[t_1,t_2]$, where C is a constant which depends on n and D. In particular,

$$s\widetilde{\mathcal{I}}(u,c)\subset\mathbb{T}^n\times B(c,C\sqrt{\varepsilon})\times\mathbb{T}\subset\mathbb{T}^n\times B(c,\varepsilon^{1/3})\times\mathbb{T}.$$

Proof. We denote by C_i various positive constants which depend on n and D. Since $\widehat{Z}_c \leqslant 0$, we have

$$L(\theta, v, t) \geqslant \frac{\|v - \partial_p H_0(c)\|^2}{4D} - 4\varepsilon\delta.$$

As a consequence, $L(\theta, v, t) \geqslant 20\varepsilon\sqrt{nD}$ if $\|v - \partial_p H_0(c)\| \geqslant C_1\sqrt{\varepsilon}$. In view of Lemma 4.3, we thus have

$$L(\theta(t), \dot{\theta}(t), t) \geqslant 20\sqrt{nD}\varepsilon$$

for each t such that $\|p(t) - c\| \geqslant C_2\sqrt{\varepsilon}$. Since θ is a calibrated curve, we have

$$\int_{t_1'}^{t_2'} L(\theta(t), \dot{\theta}(t), t)\, dt \leqslant \operatorname{osc} u$$

for each $[t_1', t_2'] \subset [t_1, t_2]$. In particular, by Lemma 4.8 we have $20\varepsilon\sqrt{nD}(t_2 - t_1) > \operatorname{osc} u$. Therefore, there exists a time $t_0 \in [t_1, t_2]$ such that $\|p(t_0) - c\| = C_2\sqrt{\varepsilon}$. Let $t_3 \in [t_1, t_2]$ be the time maximizing $\|p(t) - c\|$. We assume for definiteness that $t_3 \geqslant t_0$, and that $\|p(t) - c\| \geqslant C_2\sqrt{\varepsilon}$ for each $t \in [t_0, t_3]$ (otherwise we reduce the interval). The equations of motion imply that $\|\dot{p}\| \leqslant 2\varepsilon$ on $[t_0, t_3]$, hence $t_3 \geqslant t_0 + (\|p(t_3) - c\| - C_2\sqrt{\varepsilon})/2\varepsilon$, and, using the above lower bound on $L(\theta(t), \dot{\theta}(t), t)$,

$$20\sqrt{nD\varepsilon} \geqslant \operatorname{osc} u \geqslant \int_{t_0}^{t_3} L(\theta(t), \dot{\theta}(t), t)\, dt \geqslant 10\sqrt{nD}(\|p(t_3) - c\| - C_2\sqrt{\varepsilon}\,),$$

which implies that $\|p(t_3) - c\| \leqslant (2 + C_2)\sqrt{\varepsilon}$. □

We now assume that $Z(\theta^s, c) \leqslant Z(\theta^s_*, c) - \lambda d^2(\theta^s, \theta^s_*)$, or, equivalently, that

$$\widehat{Z}_c(\theta^s) \leqslant -\lambda d^2(\theta^s, \theta^s_*),$$

and prove the horizontal part of Theorem 4.1, or more precisely that

$$s\mathcal{I}(u, c) \subset \mathbb{T} \times B(\theta^s_*(c), \delta^{1/5}) \times \mathbb{T}. \tag{17}$$

We consider the domain $\Omega = B(\theta^s_*, 4\sqrt{\delta/\lambda}\,)$. On this domain, we have $-8\delta/\lambda \leqslant \widehat{Z}_c$, hence, by Lemma 4.8, the oscillation of u on $\mathbb{T} \times \Omega \times \mathbb{T}$ satisfies

$$\operatorname{osc}_{\mathbb{T}\times\Omega\times\mathbb{T}} u \leqslant 40\sqrt{\frac{nD\varepsilon\delta}{\lambda}}.$$

For $\theta^s \notin \Omega$, we have

$$L(\theta, v, t) - c\cdot v - \alpha(c) \geqslant \frac{\|v - \partial_p H_0(c)\|^2}{4D} + \frac{\lambda\varepsilon d^2(\theta^s, \theta^s_*)}{2} \geqslant \frac{\|v^s\|^2}{4D} + \lambda\varepsilon\frac{d^2(\theta^s, \theta^s_*)}{2},$$

by Lemma 4.7. Let $\theta:\mathbb{R}\to\mathbb{T}^n$ be a curve calibrated by u, and let $[t_1,t_2]$ be an excursion of θ^s outside of Ω, meaning that $d(\theta^s(t),\theta^s_*)>4\sqrt{\delta/\lambda}$ for each $t\in]t_1,t_2[$, and that $d(\theta^s(t_1),\theta^s_*)=4\sqrt{\delta/\lambda}=d(\theta^s(t_2),\theta^s_*)$. We have the inequalities

$$40\sqrt{\frac{nD\varepsilon\delta}{\lambda}}\geqslant\int_{t_1}^{t_2}(L(\theta(t),\dot\theta(t),t)-c\cdot\dot\theta(t)+\alpha(c))\,dt$$
$$\geqslant\int_{t_1}^{t_2}\left(\frac{\|\dot\theta^s(t)\|^2}{4D}+\lambda\varepsilon\frac{d^2(\theta^s(t),\theta^s_*(c))}{2}\right)dt.$$

If the curve $\theta^s(t)$ is not contained in $B(\theta^s_*,\delta^{1/5})$ for $t\in[t_1,t_2]$, then there exists a time interval $[t_3,t_4]\subset[t_1,t_2]$ such that $d(\theta(t),\theta^s_*)>\frac12\delta^{1/5}$ for $t\in[t_3,t_4]$, $d(\theta(t_3),\theta^s_*)=\frac12\delta^{1/5}=d(\theta(t_4),\theta^s_*)$, and $\max_{t\in[t_3,t_4]}d(\theta(t),\theta^s_*)>\delta^{1/5}$. We then have $\int_{t_3}^{t_4}\|\dot\theta^s(t)\|\,dt\geqslant\delta^{1/5}$, and hence

$$40\sqrt{\frac{nD\varepsilon\delta}{\lambda}}\geqslant\int_{t_1}^{t_2}\left(\frac{\|\dot\theta^s(t)\|^2}{4D}+\lambda\varepsilon\frac{d^2(\theta^s(t),\theta^s_*(c))}{2}\right)dt$$
$$\geqslant\int_{t_3}^{t_4}\left(\frac{\|\dot\theta^s(t)\|^2}{4D}+\frac{\lambda\varepsilon d^2(\theta^s(t),\theta^s_*(c))}{2}\right)dt$$
$$\geqslant\frac{1}{4D(t_4-t_3)}\left(\int_{t_3}^{t_4}\|\dot\theta^s(t)\|\,dt\right)^2+\frac{\lambda\varepsilon(t_4-t_3)\delta^{2/5}}{8}$$
$$\geqslant\frac{1}{4D(t_4-t_3)}\delta^{2/5}+\frac{\lambda\varepsilon\delta^{2/5}(t_4-t_3)}{8}$$
$$\geqslant\frac{\sqrt{\lambda\varepsilon}}{8\sqrt{D}}\delta^{2/5},$$

which is a contradiction when δ is small enough with respect to n, D, and λ. We have proved (17).

We can now prove a better vertical localization of the set $s\widetilde{\mathcal{I}}(u,c)$ than was obtained in Lemma 4.9. On the domain $\mathbb{T}\times B(\theta^s_*,\delta^{1/5})\times\mathbb{T}$, we have $\widehat{Z}_c\geqslant-\frac12\delta^{2/5}$. We deduce from Lemma 4.8 that

$$10\delta^{1/5}\sqrt{nD\varepsilon}\geqslant u(\theta(t_2),t_2)-u(\theta(t_1),t_1)=\int_{t_1}^{t_2}(L(\theta(t),\dot\theta(t),t)-c\cdot\dot\theta(t)+\alpha(c))\,dt$$

for each curve $\theta:\mathbb{R}\to\mathbb{T}^n$ calibrated by u and each time interval $[t_1,t_2]$. We can choose the time interval $[t_1,t_2]$ as a maximal excursion outside of $\{\|p-c\|<\frac12\sqrt{\varepsilon}\delta^{1/16}\}$. On $[t_1,t_2]$, we have $\|\dot\theta-\partial_pH_0(c)\|\geqslant\sqrt{\varepsilon}\delta^{1/16}/5D$ (by Lemma 4.3), and hence

$$L(\theta,\dot\theta,t)-c\cdot\dot\theta+\alpha(c)\geqslant\frac{\varepsilon\delta^{1/8}}{100D^2}-4\varepsilon\delta\geqslant\frac{\varepsilon\delta^{1/8}}{200D}.$$

We thus have

$$\frac{(t_2-t_1)\varepsilon\delta^{1/8}}{200D}\leqslant10\delta^{1/5}\sqrt{nD\varepsilon},$$

and so $2\varepsilon(t_2-t_1)\leqslant\frac12\sqrt{\varepsilon}\delta^{1/16}$ (if δ is small enough). As $\|\dot p\|\leqslant2\varepsilon$ and $\|p(t_1)-c\|=\frac12\sqrt{\varepsilon}\delta^{1/16}$, we conclude that $\|p(t)-c\|\leqslant\sqrt{\varepsilon}\delta^{1/16}$ on $[t_1,t_2]$. This ends the proof of Theorem 4.1.

4.3. The Lipschitz constant

We prove Theorem 4.2. We will work here with weak KAM solutions rather than suspended weak KAM solutions. We recall the concept of semi-concave function on $\mathbb{T}^n$. A function $u: \mathbb{T}^n \to \mathbb{R}$ is called K-semi-concave if the function

$$x \longmapsto u(x) - \tfrac{1}{2}K\|x\|^2$$

is concave on $\mathbb{R}^n$, where u is seen as a periodic function on $\mathbb{R}^n$. It is equivalent to require that, for each $\theta \in \mathbb{T}^n$, there exists a linear form l on $\mathbb{R}^n$ such that the inequality

$$u(\theta+y) \leqslant u(\theta) + l\cdot y + \tfrac{1}{2}K\|y\|^2$$

holds for each $y \in \mathbb{R}^n$. It is sufficient to check that, for each θ, there exists l such that this inequality holds for $\|y\| \leqslant 1$. We will need the following regularity result of Fathi; see [31].

LEMMA 4.10. *Let u_1 and u_2 be K-semi-concave functions, and let $\mathcal{I} \subset \mathbb{T}^n$ be the set of points where the sum u_1+u_2 is minimal. Then the functions u_1 and u_2 are differentiable at each point of $\mathcal{I}$, and the differential $x \mapsto du_1(x)$ is $6K$-Lipschitz on $\mathcal{I}$.*

The weak KAM solutions of cohomology c are the functions $u: \mathbb{T}^n \to \mathbb{R}$ such that

$$u(\theta) := \min_{\gamma}\left(u(\gamma(0)) + \int_0^T (L(\gamma(t), \dot\gamma(t), t) - c\cdot\dot\gamma(t) + \alpha(c))\,dt\right),$$

for each $T \in \mathbb{N}$, where the minimum is taken over the set of C^1 curves $\gamma: [0,T] \to \mathbb{T}^n$ satisfying the final condition $\gamma(T) = \theta$.

PROPOSITION 4.11. *For each $c \in \mathbb{R}^n$, each weak* KAM *solution u at cohomology c is $\frac{3}{2}\sqrt{D\varepsilon}$-semi-concave.*

Proof. Given $T \in \mathbb{N}$ and $\theta \in \mathbb{T}^n$, there exists a curve $\Theta: [0,T] \longrightarrow \mathbb{T}^n$ such that $\Theta(T) = \theta$ which is calibrated by u, which means that

$$u(\theta) = u(\Theta(0)) + \int_0^T (L(t, \Theta(t), \dot\Theta(t)) - c\cdot\dot\Theta(t) + \alpha(c))\,dt.$$

We assume that $T \geqslant \varepsilon^{-1/2}$, which implies by Lemma 4.9 that $\|p(t)-c\| \leqslant C\sqrt{\varepsilon}$, for a constant C independant of ε and δ. We deduce that $\|\dot\Theta - \partial_p H_0(c)\| \leqslant C\sqrt{\varepsilon}$ (with a higher constant C) for each $t \in [0,T]$. We lift Θ (and the point $\theta = \Theta(T)$) to a curve in $\mathbb{R}^n$ without changing its name, and consider, for each $x \in \mathbb{R}^n$, the curve

$$\Theta_x(t) := \Theta(t) + \frac{tx}{T},$$

so that $\Theta_x(T)=\theta+x$. Each of the curves Θ_x, $\|x\|\leqslant 1$, satisfy $\|\dot\Theta_x-\partial_p H_0(c)\|\leqslant C\sqrt{\varepsilon}\leqslant\varepsilon^{1/3}$ (provided ε is small enough). We have the inequality

$$u(\theta+x)-u(\theta)\leqslant\int_0^T\left(L(\Theta_x(t),\dot\Theta_x(t),t)-L(\Theta(t),\dot\Theta(t),t)-\frac{cx}{T}\right)dt.$$

Use Lemma 4.4, we get

$$\begin{aligned}L(\Theta_x(t),\dot\Theta_x(t),t)\leqslant L(\Theta(t),\dot\Theta(t),t)+\partial_\theta L(\Theta(t),\dot\Theta(t),t)\cdot\frac{tx}{T}\\ +\partial_v L(\Theta(t),\dot\Theta(t),t)\cdot\frac{x}{T}+\frac{3\varepsilon}{2}\left|\frac{tx}{T}\right|^2+2D\varepsilon t\left|\frac{x}{T}\right|^2+\frac{D}{2}\left|\frac{x}{T}\right|^2.\end{aligned}\tag{18}$$

Using the Euler–Lagrange equation and integrating by parts, we conclude that

$$u(\theta+x)-u(\theta)\leqslant(c+\partial_v L(T,\Theta(T),\dot\Theta(T)))\cdot x+\left(\frac{\varepsilon T}{2}+D\varepsilon+\frac{D}{2T}\right)|x|^2$$

for each $T\in\mathbb{N}$, $T\geqslant\varepsilon^{-1/2}$. Taking $T\in\left[\sqrt{D/\varepsilon},\sqrt{2D/\varepsilon}\,\right]$, we obtain

$$u(\theta+x)-u(\theta)\leqslant(c+\partial_v L(T,\Theta(T),\dot\Theta(T)))\cdot x+\tfrac{3}{2}\sqrt{D\varepsilon}|x|^2$$

for each $x\in\mathbb{R}^n$, $\|x\|\leqslant 1$. This ends the proof of the semi-concavity. □

Proof of Theorem 4.2. Let u be a weak KAM solution, and let $\breve u$ be the conjugated dual weak KAM solution. Then the set $\widetilde{\mathcal{I}}(u,c)$ can be characterized as follows: Its projection $\mathcal{I}(u,c)$ on $\mathbb{T}^n$ is the set where $u=\breve u$, and

$$\widetilde{\mathcal{I}}(u,c)=\{(x,c+du(x)):x\in\mathcal{I}(u,c)\}.$$

Since $-\breve u$ is semi-concave, it is a consequence of Lemma 4.10 that the differential $du(x)$ exists for $x\in\mathcal{I}(u,c)$. Moreover, we can prove exactly as in Proposition 4.11 that $-\breve u$ is $\frac{3}{2}D\varepsilon$-semi-concave. Lemma 4.10 then implies that the map $x\mapsto du(x)$ is $9\sqrt{D\varepsilon}$-Lipschitz on $\mathcal{I}(u,c)$. □

4.4. Double peak case

We now localize the Aubry and Mañé sets in the more general case where ($HZ\lambda$) is replaced by

$$\widehat{Z}_c(\theta^s)\leqslant-\lambda\min\{d(\theta^s-\theta^s_1),d(\theta^s-\theta^s_2)\}^2.$$

It is natural to relax ($HZ\lambda$) in this way because, for a generic family of functions $\widehat{Z}_c$, $c\in\Gamma$, there exist values of c for which $\widehat{Z}_c$ has two degenerate maxima. Note that Theorem 4.2 is still valid in this case, its proof does not use ($HZ\lambda$). On the other hand, Theorem 4.1 is replaced by the following result.

THEOREM 4.12. *If $\delta>0$ is small enough with respect to n, D and λ, and if ε is small enough with respect to n, D, λ and δ, then the Aubry set at cohomology c of the Hamiltonian N_ε satisfies*

$$s\tilde{\mathcal{A}}(c)\subset(B(\theta_1^s,\delta^{1/5})\cup B(\theta_2^s,\delta^{1/5}))\times\mathbb{T}\times B(c,\sqrt{\varepsilon}\delta^{1/16})\times\mathbb{T}\subset\mathbb{T}^{n-1}\times\mathbb{T}\times\mathbb{R}^n\times\mathbb{T}.$$

If, moreover, the projection $\theta^s(s\mathcal{A}(c))\subset\mathbb{T}^{n-1}$ is contained in one of the (disjoint) balls $B(\theta_i^s,\delta^{1/5})$, then the projection $\theta^s(s\mathcal{N}(c))\subset\mathbb{T}^{n-1}$ of the Mañé set is contained in the same ball $B(\theta_i^s,\delta^{1/5})$.

Proof. We assume that $\theta_1^s\neq\theta_2^s$, and that δ is small enough for the balls $B(\theta_i^s,2\delta^{1/5})$, $i=1,2$, to be disjoint. We first show that

$$\theta^s(s\mathcal{A}(c))\subset B(\theta_1^s,\delta^{1/5})\cup B(\theta_2^s,\delta^{1/5}).$$

As in the single peak case, we set $r_1=4\sqrt{\delta/\lambda}$, and observe that

$$L(\theta,v,t)-c\cdot v-\alpha(c)\geqslant\frac{\|v^s\|^2}{4D}+\frac{\lambda\varepsilon\min\{d(\theta^s-\theta_1^s),d(\theta^s-\theta_2^s)\}^2}{2}$$

for $\theta^s\notin B(\theta_1^s,r_1)\cup B(\theta_2^s,r_1)$. The θ^s component of each orbit of the Aubry set spends a finite amount of time outside of $B(\theta_1^s,r_1)\cup B(\theta_2^s,r_1)$. Each maximal orbit segment outside of this union connects $B(\theta_i^s,r_1)$ to $B(\theta_j^s,r_1)$ for some $i\in\{1,2\}$ and some $j\in\{1,2\}$. Exactly as in the single peak case, the orbits segments connecting $B(\theta_i^s,r_1)$ to itself are contained in $B(\theta_i^s,\delta^{1/5})$. So the claim holds, provided there exists no orbit segment in $s\mathcal{A}(c)$ connecting $B(\theta_i^s,r_1)$ to $B(\theta_j^s,r_1)$ with $i\neq j$.

Assume for example that there exists an orbit segment $\theta:[t_1,t_2]\to\mathbb{T}^n$ connecting $B(\theta_1^s,r_1)$ to $B(\theta_2^s,r_1)$. Then, given any suspended weak KAM solution u, the same action estimates as in the single peak case imply that

$$u(\theta(t_2),t_2)-u(\theta(t_1),t_1)\geqslant\frac{\sqrt{\lambda\varepsilon}}{8\sqrt{D}}\delta^{2/5}.$$

As the Aubry set is chain recurrent, there must exist an orbit segment $\check{\theta}:[\check{t}_1,\check{t}_2]\to\mathbb{T}^n$ connecting $B(\theta_2^s,r_1)$ to $B(\theta_1^s,r_1)$, and we have

$$u(\check{\theta}(\check{t}_2),\check{t}_2)-u(\check{\theta}(\check{t}_1),\check{t}_1)\geqslant\frac{\sqrt{\lambda\varepsilon}}{8\sqrt{D}}\delta^{2/5}.$$

By using Lemma 4.8 with $\Omega=B(\theta_1^s,r_1)$ and $\Omega=B(\theta_2^s,r_1)$, we get that

$$u(\check{\theta}(\check{t}_2),\check{t}_2)-u(\theta(t_1),t_1)\leqslant 40\sqrt{\frac{nD\varepsilon\delta}{\lambda}}\quad\text{and}\quad u(\theta(t_2),t_2)-u(\check{\theta}(\check{t}_1),\check{t}_1)\leqslant 40\sqrt{\frac{nD\varepsilon\delta}{\lambda}}.$$

All these inequalities together imply that

$$40\sqrt{\frac{nD\varepsilon\delta}{\lambda}} \geqslant \frac{\sqrt{\lambda\varepsilon}}{8\sqrt{D}}\delta^{2/5},$$

which does not hold if δ is small enough. This contradiction proves that no excursion connecting $B(\theta_1^s, r_1)$ to $B(\theta_2^s, r_1)$ can exist in the Aubry set. Note that we have used the chain recurrence of the Aubry set, and that the conclusion does not in general apply to the Mañé set. We have proved that

$$s\mathcal{A}(c) \subset \mathbb{T}\times(B(\theta_1^s, \delta^{1/5})\cup B(\theta_2^s, \delta^{1/5}))\times\mathbb{T}.$$

The vertical part of the localization follows exactly as in the single peak case.

In general, such a localization does not hold for the Mañé set, which may contain connections from one of the regions $\mathbb{T}\times B(\theta_i^s, \delta^{1/5})\times\mathbb{T}$ to the other (but, in view of the calculations above, not in both direction). If such a connection exists, then its α-limit is contained in one of the domains $\mathbb{T}\times B(\theta_i^s, \delta^{1/5})\times\mathbb{T}$, say $\mathbb{T}\times B(\theta_1^s, \delta^{1/5})\times\mathbb{T}$, and its ω-limit is contained in the other domain $\mathbb{T}\times B(\theta_2^s, \delta^{1/5})\times\mathbb{T}$. Recalling that the α and ω limits of the Mañé set are contained in the Aubry set, we conclude that each of the intersections

$$s\mathcal{A}(c)\cap(\mathbb{T}\times B(\theta_i^s, \delta^{1/5})\times\mathbb{T})$$

is non-empty. This proves the last part of the statement □

5. Non-degeneracy of the barrier functions

In this section we prove the following result.

THEOREM 5.1. *In the context of Theorem 1.5, by possibly taking a smaller* δ_0, *for a residue set of* $R\in\mathcal{R}=\mathcal{R}(r,\varepsilon,\delta_0)$ *the following hold: for any* $c\in\Gamma_1$ *such that* $\varrho(c)$ *is irrational and* $\theta^f(\mathcal{N}_N(c))=\mathbb{T}$, *the set* $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\Xi^{-1}(\widetilde{\mathcal{N}}_N(c))$ *is totally disconnected.*

This is a delicate perturbation problem, and a version of it for *a-priori unstable* systems appeared in [24] and was discussed in [51]. In this section we give a self-contained proof with many new ingredients.

5.1. Outline of the proof

In this section we prove Theorem 5.1 assuming some statements that are proved in later subsections. Let L denote the Lagrangian associated with N.

• We define $\mathcal{R}_1\subset\mathcal{R}(r,\varepsilon,\delta)$ to be the set of R such that $\theta^f(\mathcal{N}_N(c))\neq\mathbb{T}$ whenever $\varrho^f(c)$ is rational. The set $\mathcal{R}_1$ is a residue subset of $\mathcal{R}$. We also abuse notation and denote by $\mathcal{R}_1$ the set of Hamiltonians of the form $N=H_0+\varepsilon Z+\varepsilon R$, $R\in\mathcal{R}_1$.

- We define

$$\Gamma_*(N)=\{c\in\Gamma_1:\theta^f(\mathcal{N}_N(c))=\mathbb{T}\},$$

according to the previous item, for $N\in\mathcal{R}_1$ and $c\in\Gamma_*(N)$, we necessarily have $\varrho^f(c)$ irrational. In particular, $\mathcal{A}_N(c)=\mathcal{N}_N(c)$ contains a unique static class. In view of the upper semi-continuity of the Mañé set, $\Gamma_*(N)$ is a compact subset of Γ_1.

- If $N\in\mathcal{R}_1$ and $c\in\Gamma_*(N)$, then the Aubry set $\tilde{\mathcal{A}}_{N\circ\Xi}(\xi^*c)=\Xi^{-1}\tilde{\mathcal{A}}_N(c)$ contains exactly two static classes denoted $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$ (with projections $\mathcal{S}_1$ and $\mathcal{S}_2$). Then the Mañé set is the disjoint union

$$\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)=\widetilde{\mathcal{S}}_1\cup\widetilde{\mathcal{S}}_2\cup\widetilde{\mathcal{H}}_{12}\cup\widetilde{\mathcal{H}}_{21}, \tag{19}$$

where $\widetilde{\mathcal{H}}_{12}$ (and $\widetilde{\mathcal{H}}_{21}$) is the set of heteroclinic orbits from $\widetilde{\mathcal{S}}_1$ to $\widetilde{\mathcal{S}}_2$ (and vice versa). Projections are denoted $\mathcal{H}_{12}$ and $\mathcal{H}_{21}$. Note that $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\Xi^{-1}\widetilde{\mathcal{N}}_N(c)=\widetilde{\mathcal{H}}_{12}\cup\widetilde{\mathcal{H}}_{21}$. We will also use the notation $\widetilde{\mathcal{S}}_i(N,c)$ and $\widetilde{\mathcal{H}}_{ij}(N,c)$ when discussing the dependence on N and c.

- For $N\in\mathcal{R}_1$ and $c\in\Gamma_*(N)$, the static classes $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$ determine two elementary forward and two backward weak KAM solutions

$$h(\zeta_1,\cdot),\;\; h(\zeta_2,\cdot),\quad h(\cdot,\zeta_1),\;\; h(\cdot,\zeta_2),\quad \zeta_i\in\mathcal{S}_i,\;\; i=1,2,$$

where the barrier functions are evaluated for $N\circ\Xi$ and ξ^*c. The associated pseudographs are denoted $\mathcal{E}_i(N,c)$ and $\check{\mathcal{E}}_i(N,c)$, $i=1,2$, respectively, and they do not depend of the choices of points $\zeta_1\in\mathcal{S}_1$ and $\zeta_2\in\mathcal{S}_2$. Define

$$b^-_{N,c}(\theta)=h(\zeta_1,\theta)+h(\theta,\zeta_2)-h(\zeta_1,\zeta_2)$$

and $b^+_{N,c}$ is similarly defined with ζ_1 and ζ_2 switched. The functions $b^\pm_{N,c}$ do not depend on the choice of points $\zeta_1\in\mathcal{S}_1$ and $\zeta_2\in\mathcal{S}_2$, they are non-negative, and vanish, respectively, on $\mathcal{H}_{12}\cup\mathcal{S}_1\cup\mathcal{S}_2$ and on $\mathcal{H}_{21}\cup\mathcal{S}_1\cup\mathcal{S}_2$.

Given $\underline{c}\in\Gamma_1$, we consider the compact subset $\underline{\mathcal{K}}\subset\mathbb{T}^n$ formed by points θ such that $d(\theta^s(\underline{c}),\theta^s)\geqslant\frac{1}{10}$. There exists $\sigma>0$ such that the Mañé set $\mathcal{N}(N,c)$ is disjoint from $\underline{\mathcal{K}}$ for each $c\in\Gamma_1\cap\mathcal{B}_\sigma(\underline{c})$ and $N\in\mathcal{R}(r,\varepsilon,\delta_0)$. The compact set $\mathcal{K}=\xi^{-1}(\underline{\mathcal{K}})$ (ξ is the double covering) is then disjoint from $\mathcal{A}_{N\circ\Xi}(\xi^*c)$. Moreover, for these N and c, the set $\pi^{-1}(\mathcal{K})$ intersects each orbit of $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\tilde{\mathcal{A}}_{N\circ\Xi}(\xi^*c)$.

Since the compact interval Γ_1 is the union of finitely many compact segments, each contained in a ball of the form $B_\sigma(\underline{c})$, it suffices to prove Theorem 5.1 for each segment. Therefore, we may assume without loss of generality that Γ_1 is actually contained in one of these balls. Then, there exists a compact set $\mathcal{K}$ such that

- For each $c\in\Gamma_1$ and $N\in\mathcal{R}(r,\varepsilon,\delta_0)$, $\mathcal{K}$ is disjoint from $\mathcal{A}_{N\circ\Xi}(\xi^*c)$ and $\pi^{-1}(\mathcal{K})$ intersects each orbit of $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\tilde{\mathcal{A}}_{N\circ\Xi}(\xi^*c)$.

We make this additional assumption for the sequel of the section.

LEMMA 5.2. *For each* $(N,c)\in\mathcal{R}_1\times\Gamma_1$, *the set* $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\Xi^{-1}\widetilde{\mathcal{N}}_N(c)$ *is totally disconnected if and only if the set*

$$\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}=(\mathcal{H}_{12}(N,c)\cup\mathcal{H}_{21}(N,c))\cap\mathcal{K}$$

is totally disconnected.

Proof. The set $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$ is a compact metric space, so it is totally disconnected if and only if it has topological dimension zero, see [38]. Assuming that this property holds, the set $\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)\cap\pi^{-1}(K)$ is the disjoint union of two homeomorphic copies of $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$, hence it is compact and of zero topological dimension. As a consequence, each of the sets $\phi^k(\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}), k\in\mathbb{Z}$, is compact and of zero topological dimension, where ϕ^k is the time-k Hamiltonian flow of N. The countable union

$$\widetilde{\mathcal{N}}_{N\circ\Xi}(\xi^*c)-\tilde{\mathcal{A}}_{N\circ\Xi}(\xi^*c)=\bigcup_{k\in\mathbb{Z}}\phi^k(\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K})$$

is then also of zero dimension. As a consequence, the projection $\mathcal{N}_{N\circ\Xi}(\xi^*c)-\mathcal{A}_{N\circ\Xi}(\xi^*c)$ is of zero topological dimension, and hence it is totally disconnected. □

We want to prove that a dense G_δ set of Hamiltonians $N\in\mathcal{R}_1$ have the property that $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$ is totally disconnected for each $c\in\Gamma_*(N)$. The G_δ part follows from the next lemma.

LEMMA 5.3. *Let* $J\subset\Gamma_1$ *and* $K\subset\mathbb{T}^n$ *be compact subsets, then the set of remainders* $R\in\mathcal{R}$ *such that for all* $c\in J\cap\Gamma_*(N)$ *the set of points*

$$Q(N,c,K):=\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap K \tag{20}$$

is totally disconnected, is a G_δ *set.*

Proof. Consider a Hamiltonian N satisfying the conditions of the lemma. Then, for every $c\in J\cap\Gamma_*(N)$, $Q(N,c,K)$ is compact and totally disconnected, and hence has zero topological dimension.

We call a compact subset $(1/k)$-*disconnected* if it admits a finite disjoint covering by compact subsets of diameter at most $1/k$. If N satisfies the conditions of the lemma, then $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap K$ is $(1/k)$-disconnected for each $k\in\mathbb{N}$ and each $c\in J\cap\Gamma_*(N)$. Since the Mañé set is upper semi-continuous in the Hamiltonian (in the C^2 topology), so is $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap K$ and we have, for each fixed k, the following result:

There exists an open set Γ' containing $\Gamma_*(N)\cap J$ and a neighborhood $\mathcal{U}$ of N in C^2 such that the set $\mathcal{N}_{N'\circ\Xi}(\xi^*c')\cap K$ is $(1/k)$-disconnected for all $c'\in\Gamma'$ and all $N'\in\mathcal{U}$.

We now use the observation that $\Gamma_*(N)$ is upper semi-continuous in N, hence so is $J\cap\Gamma^*(N)$ since J is compact. We deduce the existence of a smaller neighborhood $\mathcal{U}'\subset\mathcal{U}$ of N, such that $J\cap\Gamma_*(N')\subset\Gamma'$ for each $N'\in\mathcal{U}'$. We have proved: the property that $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap K$ is $(1/k)$-disconnected for each $c\in\Gamma_*(N)\cap J$ is C^2 open (and hence C^r open). The lemma follows by taking the intersection over k. □

We now adress the density part. Let us consider the product space

$$C^r(\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T})\times\mathbb{R}^n$$

with the standard norms on both spaces. Define the following subset

$$\mathcal{Q}=\{(N,c):\, N\in\mathcal{R}_1,\ c\in\Gamma_*(N)\}\subset\mathcal{R}\times\Gamma_1\subset C^r(\mathbb{T}^n\times\mathbb{R}^n\times\mathbb{T})\times\mathbb{R}^n.$$

The following proposition allows us to perturb the functions $b^{\pm}_{N,c}$ locally simultaneously for an open set of c. The proof is given in §5.2.

Proposition 5.4. *Let $(N_0,c_0)\in\mathcal{Q}$ and let $K\subset\mathbb{T}^n$ be a compact set disjoint from $\mathcal{A}_{N_0\circ\Xi}(\xi^*c_0)$. Then there is $\sigma>0$ such that for all $N\in\mathcal{R}_1\cap B_\sigma(N_0)$, $\theta_0\in K\cap\mathcal{H}_{12}(N_0,c_0)$, and $\varphi\in C^r_c(B_\sigma(\theta_0))$ with $\|\varphi\|_{C^r}<\sigma$, there exists a Hamiltonian N_φ such that the following conditions hold*:

(1) *For all $c\in B_\sigma(c_0)$, the Aubry set $\tilde{\mathcal{A}}_{N_\varphi\circ\Xi}(\xi^*c)$ coincides with $\tilde{\mathcal{A}}_{N\circ\Xi}(\xi^*c)$, with the same static classes. In particular, $B_\sigma(c_0)\cap\Gamma_*(N)=B_\sigma(c_0)\cap\Gamma_*(N_\varphi)$.*

(2) *For all $c\in B_\sigma(c_0)\cap\Gamma_*(N)$, there exists a constant $e\in\mathbb{R}$ such that*

$$b^+_{N_\varphi,c}(\theta)=b^+_{N,c}(\theta)+\varphi(\theta)+e,\quad \theta\in B_\sigma(\theta_0). \tag{21}$$

The same holds for $\theta_0\in K\cap\mathcal{H}_{21}(N_0,c_0)$, with b^+ replaced by b^- in (21). Moreover, for each $N\in\mathcal{R}_1\cap B_\sigma(N_0)$, $\|N_\varphi-N\|_{C^r}\to 0$ when $\|\varphi\|_{C^r}\to 0$.

We will use Proposition 5.4 to perturb all barrier functions near a given c_0 simultaneously. Because we are perturbing an uncountable family of functions, we need additional information on how the functions $b^{\pm}_{N,c}$ depends on c. The proof is given in §5.3.

Proposition 5.5. *For each $N\in\mathcal{R}_1$, the maps $c\mapsto b^+_{N,c}$ and $c\mapsto b^-_{N,c}$ are $\frac{1}{2}$-Hölder from $\Gamma_*(N)$ to $C^0(\mathbb{T}^n,\mathbb{R})$.*

This regularity implies that the set $\{b^{\pm}_{N,c}: c\in\Gamma^*(N)\}$ is compact and has Hausdorff dimension at most 2 in $C^0(\mathbb{T}^n,\mathbb{R})$. The following lemma will allow us to take advantage of this fact.

LEMMA 5.6. *Let $\mathcal{F}\subset C^0([-1,1]^n,\mathbb{R})$ be a compact set of finite Hausdorff dimension. The following property is satisfied on a residue set of functions $\varphi\in C^r(\mathbb{R}^n,\mathbb{R})$ (with the uniform C^r norm):*

For each $f\in\mathcal{F}$, the set of minima of the function $f+\varphi$ on $[-1,1]^n$ is totally disconnected.

As a consequence, for each open neighborhood Ω of $[-1,1]^n$ in $\mathbb{R}^n$, there exist arbitrarily C^r-small compactly supported functions $\varphi:\Omega\to\mathbb{R}$ satisfying this property.

Proof. We first consider the case $n=1$. The set $\widetilde{\mathcal{F}}=\{c-f:f\in\mathcal{F},c\in\mathbb{R}\}$ is compact and of finite Hausdorff dimension (one more than the dimension of $\mathcal{F}$). For each compact subinterval $J\subset[-1,1]$, the set $\widetilde{\mathcal{F}}_J\subset C(J,\mathbb{R})$ is also compact and finite-dimensional, since the restriction map is Lipschitz. If J is non-trivial, the complement

$$\Phi(J):=C^r(\mathbb{R},\mathbb{R})-(\widetilde{\mathcal{F}}_J\cap C^r(\mathbb{R},\mathbb{R}))$$

is open and dense in $C^r(\mathbb{R},\mathbb{R})$. To prove density, we consider a subspace $H\subset C^r(\mathbb{R},\mathbb{R})$ of finite dimension larger than the Hausdorff dimension of $\widetilde{\mathcal{F}}$. We moreover assume that all functions of H are compactly supported inside the interior of J. Given $\varphi\in C^r(\mathbb{R},\mathbb{R})$, we consider the affine space $\varphi+H$. Considering the $C^0([-1,1],\mathbb{R})$ distance, the Hausdorff dimension of $\widetilde{\mathcal{F}}_J\cap(\varphi+H)$ is not greater than the Hausdorff dimension of $\widetilde{\mathcal{F}}$, hence it is less than the dimension of H. This implies that the complement $(\varphi+H)-\widetilde{\mathcal{F}}$ is dense in $\varphi+H$ endowed with the C^0 distance. Since the C^0 and C^r norms are equivalent on the finite-dimensional space $\varphi+H$, we conclude that φ belongs to the closure of $\Phi(J)$ in $C^r(\mathbb{R},\mathbb{R})$.

Let J_k be a sequence of compact subintervals of $[-1,1]$ such that each open interval contains one of the J_k. Then if $\varphi\in\bigcap_k\Phi(J_k)$ (this intersection is a dense G_δ set), each of the functions $f+\varphi$, $f\in\mathcal{F}$, has the property that it is not constant on any open interval, hence its set of minima in $[-1,1]$ is totally disconnected.

Let us now turn to the general case. We denote by $\pi_i:[-1,1]^n\to[-1,1]$ the projections on the factors. We associate with each function $f\in C^0([-1,1]^n,\mathbb{R})$ the functions

$$f_i:[-1,1]\ni x_i\longmapsto f_i(x_i)=\min_{\pi_i(x)=x_i}f(x).$$

For each k and i, the following property holds on an open and dense subset of functions $\varphi\in C^r(\mathbb{R}^n,\mathbb{R})$: None of the functions $(f+\varphi)_i$, $f\in\mathcal{F}$, is constant on J_k.

To prove density, we consider a function $\varphi\in C^r(\mathbb{R}^n,\mathbb{R})$. The map $f\mapsto f_i$ is Lipschitz, and hence the set $\mathcal{F}_i(\varphi)=\{(f+\varphi)_i:f\in\mathcal{F}\}\subset C^0([-1,1],\mathbb{R})$ is compact and has finite Hausdorff dimension. We can apply the result for $n=1$ to this family and obtain that for generic $\varphi_1\in C^r(\mathbb{R},\mathbb{R})$, none of the functions

$$(f+\varphi)_i+\varphi_i=(f+\varphi+\varphi_i)_i$$

for $f\in\mathcal{F}$ is constant on the interval J_k.

By taking the intersection over n and k, we obtain that, for generic $\varphi\in C^r(\mathbb{R}^n,\mathbb{R})$, each of the functions $(f+\varphi)_i$ has a totally disconnected set of minima in $[-1,1]$.

Since $\pi_i(\operatorname{argmin}(f+\varphi))\subset\operatorname{argmin}(f+\varphi)_i$, this implies that $\operatorname{argmin}(f+\varphi)$ is totally disconnected. □

Proof of Theorem 5.1. Let $\mathcal{R}_2\subset\mathcal{R}_1$ be the set of Hamiltonians N which have the property that $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$ is totally disconnected for each $c\in\Gamma_*(N)$.

By Lemma 5.1, it is enough to prove that $\mathcal{R}_2$ is a dense G_δ set. By Lemma 5.3, $\mathcal{R}_2$ is a G_δ, so left to prove is density.

Let us fix $N_0\in\mathcal{R}_1$. For each $\theta_0\in\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$, we consider $\sigma>0$ small enough so that Proposition 5.4 applies. We define the cube

$$D_\sigma(\theta_0)=\left\{\theta:\max_i|\theta^i-\theta_0^i|\leqslant\frac{\sigma}{2\sqrt{n}}\right\}\subset B_\sigma(\theta_0).$$

In view of Proposition 5.5, we may apply Lemma 5.6 to the family of functions $b^\pm_{N,c}$, $c\in\Gamma_1\cap\Gamma_*(N)$ on the cube $D_\sigma(\theta_0)$ for each $N\in\mathcal{R}_1$. We find arbitrarily small functions φ compactly supported in $B_\sigma(\theta_0)$ and such that each of the functions $b^\pm_{N,c}+\varphi$, $c\in\Gamma^*(N)\cap\Gamma_1$ have a totally disconnected set of minima in $D_\sigma(\theta_0)$. If $N\in\mathcal{R}_1\cap B_\sigma(N_0)$, we can apply Proposition 5.4 to get Hamiltonians N_φ approximating N. We obtain the following results:

- The set of Hamiltonians N such that $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap D_\sigma(\theta_0)$ is totally disconnected for each $c\in\Gamma_*(N)$ is dense in $\mathcal{R}_1\cap B_\sigma(N_0)$. By Lemma 5.3, it is a G_δ set.

Since $\mathcal{K}$ is compact, there is a finite cover $\mathcal{K}\subset\bigcup_{i=1}^k D_{\sigma_i}(\theta_i)$, such that the above can be applied on each $D_{\sigma_i}(\theta_i)$ for some constant $\sigma_i>0$. For $\sigma_0=\min_i\sigma_i>0$, we obtain the following statement:

- For a residue set of $N\in B_{\sigma_0}(N_0)$, the set $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap D_{\sigma_i}(\theta_i)$ is totally disconnected for all $i=1,...,k$ and $c\in\Gamma_*(N)$.

Taking the intersection over i, we obtain the following statement:

- For a residue set of $N\in B_{\sigma_0}(N_0)$, the set $\mathcal{N}_{N\circ\Xi}(\xi^*c)\cap\mathcal{K}$ is totally disconnected for all $c\in\Gamma_*(N)$.

In particular, N_0 is in the closure of $\mathcal{R}_2$. □

5.2. Perturbing the Peierls' barrier functions

Let L be the Lagrangian for $N=H_0+\varepsilon Z+\varepsilon R$. We define the generating function

$$G_N:\mathbb{R}^n\times\mathbb{R}^n\longrightarrow\mathbb{R}$$

by

$$G_N(x,x') = \min_{\gamma} \int_0^1 L(\gamma,\dot\gamma,t)\,dt, \quad \gamma(0)=x, \ \ \gamma(1)=x'.$$

Note that $G_N(x+k,x'+k)=G_L(x,x')$ for all $x,x'\in\mathbb{R}^n$ and $k\in\mathbb{Z}^n$. If ε is sufficiently small, there is a one-to-one correspondence between the time-1 map of the Euler–Lagrange flow of L, and the generating function G_N. We will also consider the generating function of the Hamiltonian $N\circ\Xi$ (pull back of the double covering), which satisfies

$$G_{N\circ\Xi}(x,x') = G_N(\xi x, \xi x'), \tag{22}$$

where we have lifted ξ to a map $\mathbb{R}^n\to\mathbb{R}^n$. It is important to keep in mind that $G_{N\circ\Xi}$ has an additional symmetry $G_{N\circ\Xi}(x,x')=G_{N\circ\Xi}\big(x+\frac12 e_1, x'+\frac12 e_1\big)$, where $e_1=(1,0,...,0)$, corresponding to the deck transformation of ξ. We also denote

$$A^M_{N,c}(\theta_1,\theta_2) = \min_{\gamma} \int_0^M (L(\gamma,\dot\gamma,t)-c\cdot\dot\gamma+\alpha_N(c))\,dt, \quad \gamma(0)=\theta_1, \ \ \gamma(M)=\theta_2\in\mathbb{T}^n,$$

and note that $A^M_{N,c}$ and therefore $h_{N,c}$ is completely determined by G_N. We will perturb the barrier functions by perturbing G_N.

Let $U,V\subset\mathbb{R}^n$ be open sets which projects injectively to $\mathbb{T}^n$, namely $U\cap(U+k)=\varnothing$ for all $k\in\mathbb{Z}^d$. We define a *perturbation block* to be the set

$$\mathcal{B}_N(U,V) := \phi_N(U\times\mathbb{R}^n)\cap(V\times\mathbb{R}^n)\subset\mathbb{R}^n\times\mathbb{R}^n.$$

In other words, this is the set of pairs (θ,p) such that $\theta\in V$ and $\pi_\theta\Phi_N^{-1}(\theta,p)\in U$, where ϕ_N is the time-1 map of the Hamiltonian N. We can also consider $\mathcal{B}_N$ as a subset of $\mathbb{T}^n\times\mathbb{R}^n$ since V projects injectively to $\mathbb{T}^n$.

Given $U_1\subset U_2\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^n$ as before, for $\varphi\in C^r_c(V)$, we define a perturbation of the generating function (depending on φ, U_1, U_2, and V) by

$$G_\varphi(x,x') = G_N(x,x') + \varrho(x)\varphi(x'), \tag{23}$$

and extend it by periodicity:

$$G_\varphi(x+k,x'+k) = G_\varphi(x,x') \quad \text{for all } k\in\mathbb{Z}^n.$$

Here $\varrho\colon\mathbb{R}^n\to\mathbb{R}^+\cup\{0\}$ is a standard mollifier function such that

$$\varrho|_{U_1} = 1 \quad \text{and} \quad \varrho|_{U_2^c} = 0.$$

LEMMA 5.7. *When $\|\varphi\|_{C^r}$ is small enough, there exists a Tonelli Hamiltonian N_φ whose generating function is equal to G_φ. Moreover, $\|N_\varphi-N\|_{C^r}\to 0$ as $\|\varphi\|_{C^r}\to 0$.*

Proof. Let $g(x,x')=\varrho(x)\varphi(x')$, extended by periodicity, then $\|g\|_{C^r}\leqslant C\|\varphi\|_{C^r}$ for some $C>0$ depending on ϱ. Let $G_t(x,x')$ be the generating function of the time-t map of the Hamiltonian N. We consider the following functions

$$G'_t(x,x')=G_t(x,x')+s(t)g(x,x'),$$

where $s:[0,1]\to[0,1]$ is a C^∞ mollifier function with $s(t)=0$ for $t\in\left[0,\frac{1}{3}\right]$ and $s(t)=1$ for $t\in\left[\frac{2}{3},1\right]$. When $\|g\|_{C^2}$ is small enough, the functions G'_t uniquely determines exact symplectic maps $\psi_t:\mathbb{T}^n\times\mathbb{R}^n\to\mathbb{T}^n\times\mathbb{R}^n$.

It is easy to see that there exists an exact symplectic isotopy between ψ_t and ϕ_t, then there is an exact symplectic isotopy between $\phi_t^{-1}\circ\psi_t$ and id. In view of Proposition 9.19 and Corollary 9.20 of [55], we get that $\{\psi_t\}_{0\leqslant t\leqslant 1}$ is a Hamiltonian isotopy. Moreover, since $d\psi_t/dt$ is periodic in t, it must be generated by a time-periodic Hamiltonian $N'(\theta,p,t)$. The maps are C^{r-1} in (θ,p) and C^∞ in t, the vector fields are C^{r-1} and the Hamiltonians are C^r.

Moreover, it is easy to see that $\psi_t\circ\phi_t^{-1}$ converges in C^{r-1} to the identity uniformly over t as $\|g\|_{C^r}\to 0$. Since $\psi_t\circ\phi_t^{-1}$ has $-N_t\circ\phi_t+N'_t\circ\phi_t$ as Hamiltonian function (see [55, Proposition 10.2]) we conclude that $\|N_t-N'_t\|_{C^r}\to 0$ as $\|g\|_{C^r}\to 0$. □

The following lemma prepares us for the perturbation. For an orbit contained in the psudograph $\overline{\mathcal{E}_1(N,c)}$, there exists a perturbation block that the orbit of (θ,p) never returns to in backward time. Moreover, the orbit also does not return to the "copy" of the perturbation block under the deck transformation of Ξ. This is important because we would like to perturb the generating function $G_{N\circ\Xi}$ by perturbing only N.

LEMMA 5.8. *Consider $(N_0,c_0)\in\mathcal{Q}$ and $(\theta_0,p_0)\in\widetilde{\mathcal{H}}_{12}(N_0,c_0)$. Then there exist $\sigma>0$, and open sets $V\ni\theta_0$ and $U_1\subset U_2\subset\mathbb{R}^n$, such that the following conditions hold:*

- *the covering map $\xi:\mathbb{T}^n\to\mathbb{T}^n$ is injective on $\overline{U}_2$ and $\overline{V}$;*
- *$\overline{U}_2\cup\left(\overline{U}_2+\frac{1}{2}e_1\right)$ and $\overline{V}\cup\left(\overline{V}+\frac{1}{2}e_1\right)$ are disjoint from $\mathcal{A}_{N_0\circ\Xi}(\xi^*c)$.*

The following hold for each $(N,c)\in\mathcal{Q}\cap B_\sigma(N_0,c_0)$:

(1) *For $\theta\in V$, let (θ,p) be contained in the closure of the psudograph $\overline{\mathcal{E}_1(N,c)}$.*

(1.a) *$(\theta,p)\in\mathcal{B}_{N\circ\Xi}(U_1,V)$;*

(1.b) *the backward orbit $\phi^{-k}_{N\circ\Xi}(\theta,p)$ is asymptotic to $\widetilde{\mathcal{S}}_1(N,c)$;*

(1.c) *for $k\geqslant 1$, $\phi^{-k}_{N\circ\Xi}(\theta,p)$ is not contained in $\mathcal{B}_{N\circ\Xi}(U_2,V)$, nor in*

$$\mathcal{B}_{N\circ\Xi}\left(U_2+\tfrac{1}{2}e_1, V+\tfrac{1}{2}e_1\right).$$

(2) *For $\theta\in V$, let (θ,p) be contained in the closure of the psudograph $\overline{\check{\mathcal{E}}_2(N,c)}$.*

(2.a) *The forward orbit $\phi^k_N(\theta,p)$ is asymptotic to $\widetilde{\mathcal{S}}_2$;*

(2.b) *for* $k\geqslant 1$, $\phi_N^k(\Xi(\theta,p))$ *is not contained in* $\mathcal{B}_{N\circ\Xi}(U_2,V)$, *nor in*

$$\mathcal{B}_{N\circ\Xi}\big(U_2+\tfrac{1}{2}e_1, V+\tfrac{1}{2}e_1\big).$$

Moreover, an analogous statement holds for $\mathcal{H}_{21}$, *where the roles of* $\mathcal{E}_1$ *and* $\check{\mathcal{E}}_2$ *are replaced by* $\mathcal{E}_2$ *and* $\check{\mathcal{E}}_1$, *respectively.*

Proof. First we claim the following statement: for any $\iota>0$ there is $\sigma>0$ such that, if $\|\theta-\theta_0\|<\sigma$ and $(N,c)\in B_\sigma(N_0,c_0)\cap\mathcal{Q}$, then $(\theta,p)\in\widetilde{\mathcal{E}}_1(N,c)$ implies the following conditions:

(c1) $\|p-p_0\|<\iota$;

(c2) the backward orbit $\phi_{N\circ\Xi}^{-k}(\theta,p)$ is asymptotic to $\widetilde{\mathcal{S}}_1(N,c)$;

(c3) there exists $M>0$ such that $k>M$ implies $\operatorname{dist}(\phi_{N\circ\Xi}^{-k}(\theta,p),\widetilde{\mathcal{S}}_1(N,c))<\iota$.

We note that $\theta_0\in\mathcal{H}_{12}(N_0,c_0)$ implies that the weak KAM solution $h(\zeta_1,\cdot)$ is differentiable at θ_0, and therefore p_0 is the unique super-differential. Item (c1) then follows from semi-continuity of super-differentials, see Proposition C.1.

Since $\theta_0\in\mathcal{H}_{12}(N_0,c_0)$, we have, for $h=h_{N_0\circ\Xi,\xi^*c}$,

$$h(\zeta_1,\theta_0)+h(\theta_0,\zeta_2)=\min_\theta(h(\zeta_1,\cdot)+h(\cdot,\zeta_2))=h(\zeta_1,\zeta_2). \tag{24}$$

Assume, by contradiction, that for $(N_k,c_k)\to(N_0,c_0)$ in $\mathcal{Q}$, and $(\theta_k,p_k)\in\mathcal{E}_1(N_k,c_k)$ with $\theta_k\to\theta_0$, the backward orbit of (θ_k,p_k) accumulates to $\mathcal{S}_2(N_k,c_k)$. This implies that

$$h_{N_k\circ\Xi,\xi^*c_k}(\zeta_1^k,\theta_k)=h_{N_k\circ\Xi,\xi^*c_k}(\zeta_1^k,\zeta_2^k)+h_{N_k\circ\Xi,\xi^*c_k}(\zeta_2^k,\theta_k),\quad \zeta_1^k\in\mathcal{S}_1,\ \zeta_2^k\in\mathcal{S}_2.$$

Taking the limit as $k\to\infty$ (by Proposition C.1), we obtain

$$h_{N_0\circ\Xi,\xi^*c_0}(\zeta_1,\theta_0)=h_{N_0\circ\Xi,\xi^*c_0}(\zeta_1,\zeta_2)+h_{N_0\circ\Xi,\xi^*c_0}(\zeta_2,\theta_0),\quad \zeta_1\in\mathcal{S}_1,\ \zeta_2\in\mathcal{S}_2.$$

Combining this with (24), we get (omitting the subscript of h)

$$h(\zeta_1,\zeta_2)=h(\zeta_1,\theta_0)+h(\theta_0,\zeta_2)=h(\zeta_1,\zeta_2)+h(\zeta_2,\theta_0)+h(\theta_0,\zeta_2),$$

or $h(\zeta_2,\theta_0)+h(\theta_0,\zeta_2)=0$, which is in contradiction with $\theta_0\notin\mathcal{S}_2$.

To prove (c3) we again argue by contradiction. Let N_k, c_k, θ_k, and p_k be as before. We assume that there exists $M_k\to\infty$ such that $\operatorname{dist}(\phi_{N\circ\Xi}^{-M_k}(\theta_k,p_k),\widetilde{\mathcal{S}}_1(N,c))\geqslant\varepsilon$. Let $m_k=\pi\phi_{N\circ\Xi}^{-M_k}(\theta_k,p_k)$, using the fact that backward orbit of (θ_k,p_k) is calibrated, we have

$$h_{N_k\circ\Xi,\xi^*c_k}(\zeta_1,\theta_k)=h_{N_k\circ\Xi,\xi^*c_k}(\zeta_1,m_k)+A^{M_k}_{N_k\circ\Xi,\xi^*c_k}(m_k,\theta_k).$$

Up to taking a subsequence, assume $m_k\to m_0$ and take the limit as $k\to\infty$ to get

$$h(\zeta_1,\theta_0)\geqslant h(\zeta_1,m_0)+h(m_0,\theta_0)=h(\zeta_1,m_0)+\min_{i=1,2}(h(m_0,\zeta_i)+h(\zeta_i,\theta_0)),$$

where h is evaluated at $N_0 \circ \Xi, \xi^* c_0$. Since $h(\zeta_1, m_0) + h(m_0, \zeta_1) > 0$, the above minimum is not reached at ζ_1. Therefore

$$h(\zeta_1, \theta_0) \geqslant h(\zeta_1, m_0) + h(m_0, \zeta_2) + h(\zeta_2, \theta_0) \geqslant h(\zeta_1, \zeta_2) + h(\zeta_2, \theta_0),$$

but we showed (in the proof of (c2)) that this is also impossible.

We now define the sets U and V. Since $\phi^{-k}_{N_0 \circ \Xi}(\theta_0, p_0)$ is asymptotic to $\mathcal{S}_1(N_0, c_0)$, projection via Ξ implies that $\phi^{-k}_{N_0}(\Xi(\theta_0, p_0))$ is asymptotic to $\Xi(\mathcal{S}_1) = \mathcal{A}_{N_0}(c_0)$. There exists $\iota_1 > 0$ such that

$$\phi^{-k}_N(\Xi(\theta_0, p_0)) \cap \Xi(B_\varepsilon(\theta_0, p_0)) = \varnothing,$$

and $\xi(B_{\iota_1}(\theta_0)) \cap \mathcal{A}_N = \varnothing$ for all $N \in B_{\iota_1}(N_0) \cap \mathcal{R}_1$.

Applying claim (c1)–(c3) to $\iota = \frac{1}{2}\iota_1$, we obtain the parameters σ and M. Since the orbit of (θ_0, p_0) is wandering, there exists $0 < \sigma_1 < \sigma$ such that $(\theta, p) \in B_{\sigma_1}(\theta_0, p_0)$ and $N \in B_{\sigma_1}(N_0)$ imply that

$$\phi^{-k}_N(\Xi(B_{\sigma_1}(\theta_0, p_0))) \cap \Xi(B_{\sigma_1}(\theta_0, p_0)) = \varnothing, \quad 1 \leqslant k \leqslant M.$$

Applying the relation $\Xi \circ \phi_{N \circ \Xi} = \phi_N \circ \Xi$, we get

$$\phi^{-k}_{N \circ \Xi}(\Xi(B_{\sigma_1}(\theta_0, p_0))) \cap \Xi^{-1}\Xi(B_{\sigma_1}(\theta_0, p_0)) = \varnothing, \quad 1 \leqslant k \leqslant M. \tag{25}$$

For a pair $\sigma_2 < \sigma_1$, which is determined later, choose $\sigma_3 < \sigma_2$ using claim (c1) again to ensure that any $(\theta, p) \in \mathcal{E}_1(N, c)$ with $\|\theta - \theta_0\| < \sigma_3$ implies that $\|p - p_0\| < \sigma_2$. Define $V = B_{\sigma_3}(\theta_0)$,

$$U_1 = \bigcup_{N \in B_{\sigma_3}(N_0)} \pi \phi^{-1}_{N \circ \Xi}(B_{\sigma_3}(\theta_0) \times B_{\sigma_2}(p_0)), \tag{26}$$

and $U_2 = B_{\sigma_2}(U_1)$. Since $U_1 \to \pi \phi^{-1}_{N_0 \circ \Xi}(\theta_0, p_0)$, as $\sigma_2, \sigma_3 \to 0$, we can choose σ_2, σ_3 small enough such that

$$\mathcal{B}_{N \circ \Xi}(\overline{U}_2, \overline{V}) \subset B_{\sigma_1}(\theta_0, p_0), \quad \text{for all } N \in B_{\sigma_3}(N_0).$$

We now verify that for $\theta \in V$ and $(\theta, p) \in \mathcal{E}_1(N, c)$, one has $\phi^{-1}_{N \circ \Xi}(\theta, p) \in U_1$, due to (26). Moreover, since

$$\mathcal{B}_{N \circ \Xi}(\overline{U}_2, \overline{V}) \cup \mathcal{B}_{N \circ \Xi}\big(\overline{U}_2 + \tfrac{1}{2}e_1, \overline{V} + \tfrac{1}{2}e_1\big) \subset \Xi^{-1}\Xi B_{\sigma_1}(\theta_0, p_0).$$

Formula (25) implies (1.c) for $1 \leqslant k \leqslant M$. On the other hand, (c3) ensures the same for $k > M$ as well.

The proof of (2.a) and (2.b) and the remaining part is analogous, so we omit it. □

Proof of Proposition 5.4. Given $\theta_0 \in K \cap \mathcal{H}_{12}(N_0, c_0)$, let (θ_0, p_0) be the corresponding point in $\widetilde{\mathcal{H}}_{12}(N_0, c_0)$. Choose $\sigma > 0$, U_1, U_2, and V as in Lemma 5.8. For $\varphi \in C_c^r(\xi V)$, consider the perturbation N_φ via (23) using the neighborhoods ξU_1, ξU_2 and ξV. Note that for $W = U_i$ or $W = V$ we have $\xi^{-1}\xi W = W \cup \big(W + \frac{1}{2}e_1\big)$, and we will use this notation throughout the proof. First, notice that according to Lemma 5.7, $\|N_\varphi - N\|_{C^r} \to 0$ as $\|\varphi\|_{C^r} \to 0$.

Item 1. We first show that the perturbation N_φ does not affect the Aubry set and the static classes. Lemma 5.8 asserts that $\xi^{-1}\xi \overline{U}_2$ and $\xi^{-1}\xi \overline{V}$ are disjoint from $\mathcal{A}_{N_0 \circ \Xi}(\xi^* c_0)$. For $(N, c) \in B_\sigma(N_0, c_0)$ and σ small enough, using semi-continuity, $\xi^{-1}\xi \overline{U}_2$ and $\xi^{-1}\xi \overline{V}$ are disjoint from $\mathcal{A}_{N \circ \Xi}(\xi^* c)$ and $\mathcal{A}_{N_\varphi \circ \Xi}(\xi^* c)$. Then (23) and (22) implies that the $L_{N \circ \Xi}$ action and $L_{N_\varphi \circ \Xi}$ action coincide on orbits of $\tilde{\mathcal{A}}_{N \circ \Xi}(\xi^* c)$ and $\tilde{\mathcal{A}}_{N_\varphi \circ \Xi}(\xi^* c)$. As a result, $\tilde{\mathcal{A}}_{N \circ \Xi}(\xi^* c)$ and $\tilde{\mathcal{A}}_{N_\varphi \circ \Xi}(\xi^* c)$ must coincide with the same static classes.

Item 2. We proceed to prove (21). Let $(\theta, p) \in \overline{\mathcal{E}_1(N, c)}$, then $\gamma(t) := \pi_\theta \circ \phi^t(\theta, p)$ is a calibrated orbit (on $(-\infty, 0]$) for the weak KAM solution $h_{N_\varphi \circ \Xi, \xi^* c}(\zeta_1, \cdot)$, with $\zeta_1 \in \mathcal{S}_1$. Write $\gamma_t = \gamma(t)$. Since $\gamma(t)$ is backward asymptotic to $\mathcal{S}_1$, there is $i_k \to \infty$ such that

$$
\begin{aligned}
h_{N_\varphi \circ \Xi, \xi^* c}(\zeta_1, \theta) &= \lim_{k \to \infty} A^{i_k}_{N_\varphi \circ \Xi, \xi^* c}(\gamma_{-i_k}, \gamma_0) \\
&= \lim_{k \to \infty} \sum_{j = -i_k}^{-1} \big(G_{N_\varphi \circ \Xi}(\gamma_j, \gamma_{j+1}) - \xi^* c \cdot (\gamma_{j+1} - \gamma_j) + \alpha_{N_\varphi \circ \Xi}(\xi^* c)\big),
\end{aligned}
\tag{27}
$$

where in the last line γ is lifted to $\mathbb{R}^n$. In view of (1.c) and (23), for any $j \leqslant -2$ we have

$$
G_{N_\varphi \circ \Xi}(\gamma_j, \gamma_{j+1}) = G_{N_\varphi}(\xi\gamma_j, \xi\gamma_{j+1}) = G_N(\xi\gamma_j, \xi\gamma_{j+1}) = G_{N \circ \Xi}(\gamma_j, \gamma_{j+1}).
$$

By the same reasoning, we have

$$
G_{N_\varphi \circ \Xi}(\gamma_{-1}, \gamma_0) = G_{N \circ \Xi}(\gamma_{-1}, \gamma_0) + \varrho(\gamma_{-1})\varphi(\gamma_0) = G_{N \circ \Xi}(\gamma_{-1}, \gamma_0).
$$

Using (27), we get

$$
h_{N_\varphi \circ \Xi, \xi^* c}(\zeta_1, \theta) = \lim_{k \to \infty} A^{i_k}_{N \circ \Xi, \xi^* c}(\gamma_{-i_k}, \gamma_0) \leqslant h_{N \circ \Xi, \xi^* c}(\zeta_1, \theta).
$$

Observe that the previous arguments hold when N_φ and N are switched, and the last displayed formula becomes an equality. By the same reasoning, using Lemma 5.8, (2.a) and (2.b), we obtain

$$
h_{N_\varphi \circ \Xi,\, \xi^* c}(\theta, \zeta_2) = h_{N \circ \Xi, \xi^* c}(\theta, \zeta_2), \quad \zeta_2 \in \mathcal{S}_2.
$$

Thus (21) follows. The proof for b^- is identical with two static classes switched. □

5.3. Hölder continuity of the barrier functions

We prove Proposition 5.5 by relating the barriers to the stable and unstable manifolds of the Aubry sets.

Recall that the system N admits a weakly invariant cylinder $\mathcal{C}$ which contains the Aubry set $\tilde{\mathcal{A}}_N(c)$ for $c\in\Gamma_1$. Using the covering map Ξ, we obtain $\Xi^{-1}\mathcal{C}=\mathcal{C}_1\cup\mathcal{C}_2$ and denote $\widetilde{\mathcal{S}}_i(N,c)=\mathcal{C}_i\cap\Xi^{-1}(\tilde{\mathcal{A}}(c))$, $i=1,2$, for all $c\in\Gamma_*(N)$.

Recall that $\Gamma_*(N)$ is the set of $c\in\Gamma_1$ such that $\mathcal{A}_N(c)$ is an invariant curve contained in $\mathcal{C}$. Let $c^\pm$ be the $c\in\Gamma_*(N)$ with the smallest and largest p^f component. Then the component of $\mathcal{C}$ bounded by $\mathcal{A}_N(c^\pm)$ is an invariant set for ϕ_N, we denote it Λ_*. Let Λ_1 and Λ_2 be the lifts under Ξ, then $\Lambda_i\subset\mathcal{C}_i$ are normally hyperbolic invariant manifolds for $\phi_{N\circ\Xi}$.

They admit C^2 center stable and center unstable manifolds $W^{cs/cu}$, which locally are graphs above (θ,p^f). These manifolds are foliated by the strong stable and unstable manifolds $W^{s,u}(z)$ of the points of Λ_i, see Appendix A. The leaves $W^{s,u}(z)$ of this foliation are C^2, they are locally graphs above θ^s. The foliation itself is C^1.

Consider $c\in\Gamma_*(N)$, then for $i=1,2$, $\widetilde{\mathcal{S}}_i(N,c)$ is a Lipschitz invariant curve. Define the sets

$$W_i^{u/s}(N,c)=\bigcup_{z\in\widetilde{\mathcal{S}}_i(N,c)}W^{u/s}(z).$$

Since $\widetilde{\mathcal{S}}_i(N,c)$ are Lipschitz graphs over θ^f, and since $W^{u,s}$ are C^1 foliations whose leaves are graphs over θ^s, $W_i^{u/s}(N,c)$ are Lipschitz graphs over θ in a neighborhood of $\widetilde{\mathcal{S}}_i$. We will show that they coincide with the pseudographs $\mathcal{E}_i(N,c)$ in a neighborhood of $\mathcal{S}_i(N,c)$.

Lemma 5.9. *For $i,j=1,2$, if $(\theta,p)\in\overline{\mathcal{E}_i(N,c)}$ is backward asymptotic to $\mathcal{S}_j(N,c)$, then there exists $M>0$ such that $\phi_{N\circ\Xi}^{-k}\in W_j^u(N,c)$ for each $k>M$.*

Suppose an orbit is backward asymptotic to $\mathcal{S}_1(N,c)$, then it is asymptotic to the normally hyperbolic set Λ_1. This orbit is contained in the strong manifold of a point $z'\in\Lambda_1$ which is asymptotic to $\mathcal{S}_1(N,c)$, but which in principle may not belong to $\mathcal{S}_1(N,c)$. To prove that $z'\in\mathcal{S}_1(N,c)$, we need an argument similar to Theorem 1.4.

We need the following version of Proposition 4.11.

Proposition 5.10. *Suppose $k\geqslant 1/\sqrt{\varepsilon}$. Then for each semi-concave function u_0, the function $u_k=T_c^k u_0$ is $6D\sqrt{\varepsilon}-$semi-concave and $6D\sqrt{n\varepsilon}-$Lipschitz. A similar statement holds for $\breve{T}_c^k u$. As a result, for any weak* KAM *solution u and $k\geqslant 1/\sqrt{\varepsilon}$, the set*

$$\phi_N^{-k}(\bar{\mathcal{G}}_{c,u})$$

is a $6D\sqrt{\varepsilon}-$Lipschitz graph over the θ component.

Proof. We observe that the proof of Proposition 4.3 applies as long as we replace $u(\theta)$ by u_k and $u(\Theta(0))$ by $u_0(\Theta(0))$. The assumption $k\geqslant 1/\sqrt{\varepsilon}$ ensures that we can choose $T\in[1/2\sqrt{\varepsilon}, 1/\sqrt{\varepsilon}\,]$ in that proof.

For the second part, observe that

$$\phi_N^{-k}(\bar{\mathcal{G}}_{c,u})\subset\mathcal{G}_{c,u}\tilde{\wedge}\check{\mathcal{G}}_{c,\overline{T}_c^k u}$$

and the proof is similar to that of Theorem 4.1. □

For the rest of this section, ϕ denotes $\phi_{N\circ\Xi}$.

Proof of Lemma 5.9. We only prove the lemma for the case $i=j=1$ as the other cases are similar. Since $z:=(\theta,p)$ is backward asymptotic to $\mathcal{S}_1(N,c)\subset\Lambda_1$, then there exists $z_1\in\Lambda_1$ such that $(\theta,p)\in W^u(z_1)$. Necessarily $\phi^{-k}(z_1)$ converges to $\mathcal{S}_1(N,c)$. We will show that $z_1\in\mathcal{S}_1(N,c)$.

Arguing by contradiction, suppose $z_1\notin\mathcal{S}_1(N,c)$, then using the fact that $T\mathcal{C}_1$ is the central direction, $\operatorname{dist}(\phi^{-k}z_1,\mathcal{S}_1(N,c))$ converges at a maximal rate of ϱ^n.

Set $z_1^k=\phi^{-k}(z_1)$. Since $\mathcal{S}_1(N,c)$ projects onto the θ^f component, for each $k\in\mathbb{N}$ there exists $z_2^k\in\mathcal{S}_1(N,c)$ such that $\theta^f(z_1^k)=\theta^f(z_2^k)$. According to Theorem 3.1, there exists $D_1>1$ such that $\mathcal{C}$ is a $(D_1/\sqrt{\varepsilon}\,)$-graph over (θ^f,p^f), which implies that

$$\|p^f(z_1^k)-p^f(z_2^k)\|\geqslant\frac{\sqrt{\varepsilon}\|z_1^k-z_2^k\|}{D_1}\geqslant\frac{\sqrt{\varepsilon}\varrho^k}{D_2}\tag{28}$$

for some $D_2>1$. Let $z^k=\phi^{-k}(z)$, then we have $\|z^k-z_1^k\|<C\lambda^k$. Suppose that k is large enough so that $C\lambda^k<\frac{1}{2}D_2^{-1}\sqrt{\varepsilon}\varrho^k$. Then

$$\|p^f(z^k)-p^f(z_2^k)\|\geqslant\|p^f(z_1^k)-p^f(z_2^k)\|-\|p^f(z^k)-p^f(z_1^k)\|\geqslant\tfrac{1}{2}\|p^f(z_1^k)-p^f(z_2^k)\|.\tag{29}$$

Assume that $k\geqslant 1/\sqrt{\varepsilon}$. We now use Proposition 5.10 to get, for some $D_3>1$,

$$\begin{aligned}\|p(z^k)-p(z_2^k)\|&\leqslant D_3\sqrt{\varepsilon}(\|\theta^s(z^k)-\theta^s(z_3^k)\|+\|\theta^f(z^k)-\theta^f(z_3^k)\|)\\&\leqslant D_3\sqrt{\varepsilon}(\|\theta^s(z^k)-\theta^s(z_3^k)\|+\|\theta^f(z^k)-\theta^f(z_1^k)\|)\\&\leqslant D_2\sqrt{\varepsilon}(\|\theta^s(z^k)-\theta^s(z_2^k)\|)+D_2D_3\sqrt{\varepsilon}\lambda^k\end{aligned}\tag{30}$$

keeping in mind that $\theta^f(z_1^k)=\theta^f(z_2^k)$. Since $z_1^k,z_2^k\in\mathcal{C}_1$, using Theorem 1.4, we get, for small ε,

$$\begin{aligned}\|\theta^s(z^k)-\theta^s(z_2^k)\|&\leqslant\|\theta^s(z_1^k)-\theta^s(z_2^k)\|+C\lambda^k\\&\leqslant\frac{1+\sqrt{\delta/\varepsilon}}{\varkappa}(\|\theta^f(z_1^k)-\theta^f(z_2^k)\|+\|p^f(z_1^k)-p^f(z_2^k)\|)+C\lambda^k\\&\leqslant 4\varkappa^{-1}\delta^{1/2}\varepsilon^{-1/2}\|p^f(z^k)-p^f(z_2^k)\|+C\lambda^k.\end{aligned}$$

Combining with (30), we get

$$\|p(z^k)-p(z_2^k)\| \leqslant 4C\varkappa^{-1}\delta^{1/2}\|p(z^k)-p(z_2^k)\|+2D_2D_3\sqrt{\varepsilon}\lambda^k.$$

When $\varkappa^{-1}\delta^{1/2}<\frac{1}{2}$, we get $\|p(z^k)-p(z_2^k)\|\leqslant 4D_2D_3\sqrt{\varepsilon}\lambda^k$, but this is in contradiction with (28) and (29). □

LEMMA 5.11. *Let $(N,c_0)\in\mathcal{Q}$. There exist σ_1, σ_2, and $M>0$ such that, for all $c\in B_{\sigma_1}(c_0)\cap\Gamma_*(N)$, we have, for $i=1,2$,*

(1)

$$\overline{\mathcal{E}_i(N,c)}\cap\pi^{-1}(B_{\sigma_2}(\mathcal{S}_i(N,c_0)))\subset W_i^u(N,c).$$

This also implies that $\mathcal{E}_i(N,c)=\overline{\mathcal{E}_i(N,c)}$ and is C^1 over $B_{\sigma_2}(\mathcal{S}_i(N,c_0))$.

(2) *For each $(\theta,p)\in\overline{\mathcal{E}_i(N,c)}$, there exists $k\leqslant M$ such that*

$$\phi^{-k}(\theta,p)\in B_{\sigma_2}(\mathcal{S}_1(N,c)\cup\mathcal{S}_2(N,c)).$$

Proof. We prove statement (1) for $i=1$, the proof for $i=2$ being identical. We first prove the statement for $c=c_0$, and then extend to a neighborhood by continuity. First of all, we refer to [8, Lemma 4.4], to get the existence of $\sigma_3>0$ such that every $(\theta,p)\in\mathcal{E}_1(N,c)$ with $\theta\in B_{\sigma_3}(\mathcal{S}_1(N,c_0))$ is backward asymptotic to $\mathcal{S}_1$. By Lemma 5.9, there exists k such that $\phi^{-k}(\theta,p)\in W_1^u(N,c)$. We now show that k can be chosen uniformly for all $\theta\in\overline{B_{\sigma_3/2}(\mathcal{S}_1(N,c_0))}$. Arguing by contradiction, if there is $k_i\to\infty$ and $\phi^{-j}(\theta_i,p_i)\notin W_1^u(N,c)$ for all $0\leqslant j\leqslant k_i$, after taking a convergent subsequence, we get $(\theta_i,p_i)\to(\theta_*,p_*)\in\overline{\mathcal{E}_1(N,c)}$, whose backward orbit does not intersect $W_1^u(N,c)$. This is a contradiction. Using a similar compactness argument over c, we obtain the following statement:

There exists $\sigma_4,\sigma_5>0$ and $M>0$, such that for all $c\in B_{\sigma_4}(c_0)\cap\Gamma_*(N)$ and $(\theta,p)\in B_{\sigma_5}(\mathcal{S}(N,c_0))$ we have $\phi^{-k}(\theta,p)\in W_1^u(N,c)$ for all $k\geqslant M$.

Finally, we choose σ_6 small enough so that $B_{\sigma_6}(\mathcal{S}_1(N,c_0))\subset\phi^{-M}(B_{\sigma_5}(\mathcal{S}_1(n,c_0)))$. Since $\mathcal{S}_1(N,c)$ is semi-continuous in c, this property extends to a small neighborhood of $c\in\Gamma_*(N)$.

We now prove statement (2), for $i=1$. Assume that there exist $\sigma_7>0$, $k_i\to\infty$, and $(\theta_i,p_i)\in\overline{\mathcal{E}_1(N,c_i)}$ with $c_i\to c_0$, such that $\phi^{-j}(\theta_i,p_i)\notin B_{\sigma_7}(\mathcal{S}_1\cup\mathcal{S}_2)$ for all $0\leqslant j\leqslant k_i$. Taking the limit up to a subsequence, we obtain an orbit $(\theta_*,p_*)\in\overline{\mathcal{E}_1(N,c_0)}$ *not* backward asymptotic to $\mathcal{S}_1\cup\mathcal{S}_2$, a contradiction. □

For each $c\in\Gamma_*(N)$, the set $\widetilde{\mathcal{S}}_1(N,c)$ is a graph over θ^f, and hence there exists a map $\eta_c:\mathbb{T}\longrightarrow\mathbb{T}^n\times\mathbb{R}^n$ such that $\mathcal{S}_1(N,C)$ is the image of η_c and $\pi_{\theta^f}\circ\eta_c(s)=s$.

LEMMA 5.12. *There exists $C_1>0$ such that*

$$\sup_s \|\eta_c(s)-\eta_{c'}(s)\| \leqslant C_1\|c-c'\|^{1/2}$$

for each c and c' in $\Gamma_*(N)$.

Proof. We denote by D_i different positive constants that may depend on ε and δ. Since $\mathcal{C}_1$ is a Lipschitz graph over (θ^f, p^f),

$$\sup_s \|\eta_c(s)-\eta_{c'}(s)\| \leqslant D_1 \sup_s \|\pi_{p^f}\eta_c(s)-\pi_{p^f}\eta_{c'}(s)\|. \tag{31}$$

Each weak KAM solution u_c is differentiable on $\mathcal{S}_1(N,c)$, and we have

$$\pi_p\circ\eta_c = c+du_c(\pi_\theta\circ\eta_c).$$

We have

$$\int_\eta p\,d\theta = \int_\eta c\,d\theta + \int_\eta du_c(\pi_\theta\circ\eta_c)\,d\theta = \pi_{p^f}(c),$$

hence the symplectic area $A(\eta_c,\eta_{c'})$ of the domain of $\mathcal{C}_1$ delimited by the curves η_c and $\eta_{c'}$ is

$$A(\eta_c,\eta_{c'}) = \left(\int_\eta - \int_{\eta_{c'}}\right) p\,d\theta = \pi_{p^f}(c)-\pi_{p^f}(c').$$

Recall that the cylinder $\mathcal{C}_1$ is given by a graph $(\theta^s,p^s)=(\Theta^s,P^s)(\theta^f,p^f)$. The estimates (4) imply that, if v,v' are two vectors tangent to $\mathcal{C}_1$, then

$$|(d\Theta^s\wedge dP^s)(v,v')| \leqslant C\sqrt{\delta}|\,d\theta^f\wedge dp^f(v,v')|,$$

and hence, if δ is small enough,

$$|(d\Theta\wedge dP)(v,v')| \geqslant \tfrac{1}{2}|(d\theta^f\wedge dp^f)(v,v')|.$$

Note that, given two C-Lipschitz functions $\gamma_1,\gamma_2\colon\mathbb{T}\to\mathbb{R}$ with $\gamma_1(s)>\gamma_2(s)$, one has

$$\int(\gamma_1-\gamma_2)\,ds \geqslant \frac{1}{4C}\sup_s\|\gamma_1(s)-\gamma_2(s)\|^2.$$

Let Ω denote the region on $\mathcal{C}_1$ between η_c and $\eta_{c'}$. For $c,c'\in\Gamma_*$, there are $D_3,D_4>1$ such that

$$\begin{aligned} D_3\|c-c'\| \geqslant \|\pi_{p^f}(c)-\pi_{p^f}(c')\| = |A(\eta_c,\eta_{c'})| &\geqslant \frac{1}{2}\left|\int_\Omega d\theta^f\wedge dp^f\right| \\ = \frac{1}{2}\left|\int(\pi_{p^f}\circ\eta_c(s)-\pi_{p^f}\circ\eta_{c'}(s))\,ds\right| &\geqslant \frac{1}{D_4}\sup_s\|\pi_{p^f}\circ\eta_c(s)-\pi_{p^f}\circ\eta_{c'}(s)\|^2. \end{aligned} \tag{32}$$

Combining this with (31), we get our conclusion. □

LEMMA 5.13. *In the context of Lemma 5.11, consider for* $c, c' \in B_{\sigma_1}(c_0) \cap \Gamma_*(N)$, $\zeta_1 \in \mathcal{S}_1(N, c)$, *and* $\zeta_1' \in \mathcal{S}_1(N, c')$, *set*

$$u_c(\cdot) = h_{\xi^* c}(\zeta_1, \cdot) = h_{N \circ \Xi, \xi^* c}(\zeta_1, \cdot) \quad \textit{and} \quad u_{c'}(\cdot) = h_{\xi^* c'}(\zeta_2, \cdot) = h_{N \circ \Xi, \xi^* c'}(\zeta_1', \cdot).$$

Then, for $\theta \in B_{\sigma_2}(\mathcal{S}_1(N, c_0))$,

(1) $|\nabla u_c(\theta) - \nabla u_{c'}(\theta)| \leqslant C_2 \|c - c'\|^{1/2}$;

(2) $|u_c(\theta) - u_{c'}(\theta) - C_3| \leqslant C_2 \|c - c'\|^{1/2}$.

Moreover, the same holds with $\mathcal{S}_1$ *replaced with* $\mathcal{S}_2$.

Proof. For $\theta \in B_{\sigma_2}(\mathcal{S}_1(N, c_0))$, let $y = (\theta, \nabla u_c(\theta))$, and let $z \in \mathcal{S}_1(N, c)$ be such that $y \in W^s(z)$. We then define $z' \in \mathcal{S}_1(N, c')$ to be the unique such point with $\theta^f(z') = \theta^f(z)$. Finally, define $y' \in W^u(z')$ such that $\theta^s(y') = \theta^s(y)$, which is possible since $W^u(z')$ is locally a graph over θ^s.

We note that within the center unstable manifold $W^u(\Lambda)$, the NHIC Λ on one hand, and $\theta^s = \theta^s(y)$ on the other hand, serves as two transversals to the strong unstable foliation $\{W^u(\cdot)\}$. Since the foliation is C^1, there exists $D_1 > 0$ such that

$$\|y - y'\| \leqslant D_1 \|z - z'\| \leqslant C_1 D_1 \|c - c'\|^{1/2},$$

where C_1 is the constant from Lemma 5.12. Let $w = (\theta, \nabla u_{c'}(\theta))$, and noting that $y' \in W_1^u(N, c') = \{(x, \nabla u_{c'}(x))\}$ which is locally a C^1 graph, we get, for $D_2 > 0$,

$$\|w - y'\| \leqslant D_2 \|\pi_\theta(w) - \pi_\theta(y')\| = D_2 \|\pi_\theta(y) - \pi_\theta(y')\| \leqslant D_2 \|y - y'\|,$$

therefore

$$\|\nabla u_c(\theta) - \nabla u_{c'}(\theta)\| \leqslant \|w - y\| \leqslant \|w - y'\| + \|y - y'\| \leqslant D_3 \|y - y'\| \leqslant D_4 \|c - c'\|^{1/2}.$$

Item (1) follows. For item (2), we consider $\theta, \theta_0 \in B_{\sigma_2}(\mathcal{S}_1(N, c_0))$, then integrating item (1) leads to

$$|u_c(\theta) - u_{c'}(\theta) - (u_c(\theta_0) - u_{c'}(\theta_0))| \leqslant D_5 \|c - c'\|^{1/2}. \tag{33}$$

Item (2) follows by taking $C_3 = u_c(\theta_0) - u_{c'}(\theta_0)$. □

Proof of Proposition 5.5. Fix $(N, c_0) \in \mathcal{Q}$ and consider $c \in B_{\sigma_2}(c_0) \cap \Gamma_*(N)$ in the context of Lemma 5.11. From item (2) of that lemma, for every $\theta \in \mathbb{T}^n$, there exists a calibrated orbit $\gamma: (-\infty, 0] \to \mathbb{T}^n$, with $\gamma(0) = \theta$, such that $\gamma(t) \in B_{\sigma_2}(\mathcal{S}_1(N, c) \cup \mathcal{S}_2(N, c))$ whenever $t < -M$. Then (omitting the subscript $N \circ \Xi$)

$$h_{\xi^* c}(\zeta_1, \theta) = \min_{i=1,2} \min_{k \leqslant M} \min_{\theta' \in B_{\sigma_2}} \{h_{\xi^* c}(\zeta_1, \zeta_i) + h_{\xi^* c}(\zeta_i, \theta') + A^k_{\xi^* c}(\theta', \theta)\}.$$

As $h_{\xi^* c}(\zeta_i, \theta')$ is uniformly $\frac{1}{2}$-Holder in c for $\theta' \in B_{\sigma_2}(\mathcal{S}_i(N, c))$ and $c \in B_{\sigma_2}(c_0) \cap \Gamma_*(N)$, each $A^k_{\xi^* c}$ are uniformly Lipschitz in c, the family $h_{\xi^* c}(\zeta_1, \theta)$ is $\frac{1}{2}$-Hölder in c. □

Appendix A. Normally hyperbolic manifold

Let $F:\mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a C^1 vector field. We give sufficient conditions for the existence of a Normally hyperbolic invariant graph of F. We split the space $\mathbb{R}^n$ as $\mathbb{R}^{n_u}\times\mathbb{R}^{n_s}\times\mathbb{R}^{n_c}$, and denote by $x=(u,s,c)$ the points of $\mathbb{R}^n$. We denote by (F_u,F_s,F_c) the components of F:

$$F(x)=(F_u(x),F_s(x),F_c(x)).$$

We study the flow of F in the domain

$$\Omega=B^u\times B^s\times\Omega^c$$

where B^u and B^s are the open Euclidean balls of radius r_u and r_s in $\mathbb{R}^{n_u}$ and $\mathbb{R}^{n_s}$, and Ω^c is a convex open subset of $\mathbb{R}^{n_c}$. We denote by

$$L(x)=dF(x)=\begin{bmatrix} L_{uu}(x) & L_{us}(x) & L_{uc}(x)\\ L_{su}(x) & L_{ss}(x) & L_{sc}(x)\\ L_{cu}(x) & L_{cs}(x) & L_{cc}(x)\end{bmatrix}$$

the linearized vector field at the point x. We assume that $\|L(x)\|$ is bounded on Ω, which implies that each trajectory of F is defined until it leaves Ω. We denote by W^c the union of full orbits contained in Ω. In other words, this is the set of initial conditions $\underline{x}\in\Omega$ such that there exists a solution $x:\mathbb{R}\to\Omega$ of the equation $\dot{x}=F(x)$ satisfying $x(0)=\underline{x}$. We denote by W^{sc} the set of points whose positive orbit remains inside Ω. In other words, this is the set of initial conditions $\underline{x}\in\Omega$ such that there exists a solution $x:[0,\infty)\to\Omega$ of the equation $\dot{x}=F(x)$ satisfying $x(0)=\underline{x}$. Finally, we denote by W^{uc} the set of points whose negative orbit remains inside Ω. In other words, this is the set of initial conditions $\underline{x}\in\Omega$ such that there exists a solution $x:(\infty,0]\to\Omega$ of the equation $\dot{x}=F(x)$ satisfying $x(0)=\underline{x}$. These sets have specific features under the following assumptions.

Hypothesis 3. (Isolating block)

- $F_c=0$ on $B^u\times B^s\times\partial\Omega^c$;
- $F_u(u,s,c)\cdot u>0$ on $\partial B^u\times\overline{B}^s\times\overline{\Omega}^c$;
- $F_s(u,s,c)\cdot s<0$ on $\overline{B}^u\times\partial B^s\times\overline{\Omega}^c$.

Hypothesis 4. There exist positive constants α and m such that

- $L_{uu}(x)\geqslant\alpha I$ and $L_{ss}(x)\leqslant-\alpha I$ for each $x\in\Omega$ in the sense of quadratic forms;
- for each $x\in\Omega$,

$$\|L_{us}(x)\|+\|L_{uc}(x)\|+\|L_{su}(x)\|+\|L_{sc}(x)\|+\|L_{cu}(x)\|+\|L_{cs}(x)\|+\|L_{cc}(x)\|\leqslant m.$$

Theorem A.1. *Assume that Hypotheses* 3 *and* 4 *hold, and that*

$$0 \leqslant K := \frac{m}{\alpha - 2m} \leqslant \frac{1}{\sqrt{2}}.$$

Then the set W^{sc} *is the graph of a* C^1 *function*

$$w^{sc}: B^s \times \Omega^c \longrightarrow B^u,$$

the set W^{uc} *is the graph of a* C^1 *function*

$$w^{uc}: B^u \times \Omega^c \longrightarrow B^s,$$

and the set W^c *is the graph of a* C^1 *function*

$$w^c = (w^c_u, w^c_s): \Omega^c \longrightarrow B^u \times B^s.$$

Moreover, we have the estimates

$$\|dw^{sc}\| \leqslant K, \quad \|dw^{uc}\| \leqslant K, \quad \textit{and} \quad \|dw^c\| \leqslant 2K.$$

Proof. These results could be reduced to several already existing ones, (see [33], [37], [56], and [19]) or proved directly by well-known methods. We shall use [64, Theorem 1.1], which is the closest to our needs because it is expressed in terms of vector fields. We first derive some conclusions from the isolating block conditions. We denote by π^{sc} the projection $(u, s, c) \mapsto (s, c)$, and so on.

Lemma A.2. *If Hypothesis* 3 *holds, then*

$$\pi^{sc}(W^{sc}) = B^s \times \Omega^c \quad \textit{and} \quad \pi^{uc}(W^{uc}) = B^u \times \Omega^c.$$

Moreover, the closures of W^{sc} *and* W^{uc} *satisfy*

$$\overline{W}^{sc} \subset B^u \times \overline{B}^s \times \overline{\Omega}^c \quad \textit{and} \quad \overline{W}^{uc} \subset \overline{B}^u \times B^s \times \overline{\Omega}^c.$$

Proof. Let us define $T^+(x) \in [0, \infty]$ as the first positive time where the orbit of x hits the boundary $\partial\Omega$. Let us denote by $\varphi(t, x)$ the flow of F. If $T^+(x) < \infty$ (which is equivalent to $x \notin W^{sc}$), we have $\varphi(T^+(x), x) \in \partial B^u \times B^s \times \Omega^c$, as follows from Hypothesis 3. Then, it is easy to check that the function T^+ is continuous, and even C^1, at x.

We prove the first equality of the lemma by contradiction, and assume that there exists a point $(s, c) \in B^s \times \Omega^c$ such that W^{sc} does not intersect the disc $B^u \times \{s\} \times \{c\}$. Then, the first exit map

$$B^u \ni u \longmapsto \pi^u \circ \varphi(T^+(u, s, c), (u, s, c)) \in \partial B^u$$

extends by continuity to a continuous retraction from $\bar{B}^u$ to its boundary ∂B^u. Such a retraction does not exist. The proof of the other equality is similar.

Finally, we have

$$\overline{W}^{sc} \subset \bar{B}^u \times \bar{B}^s \times \bar{\Omega}^c = (B^u \times \bar{B}^s \times \bar{\Omega}^c) \cup (\partial B^u \times \bar{B}^s \times \bar{\Omega}^c).$$

Hypothesis 3 implies that each point of $\partial B^u \times \bar{B}^s \times \bar{\Omega}^c$ has a neighborhood formed of points which leave Ω after a small time. As a consequence, the set $\partial B^u \times \bar{B}^s \times \bar{\Omega}^c$ cannot intersect $\overline{W}^{uc}$, and we have proved that $\overline{W}^{sc} \subset B^u \times \bar{B}^s \times \bar{\Omega}^c$. The other inclusion can be proved in a similar way. □

In order to prove the statement of the Theorem concerning W^{sc}, we apply [64, Theorem 1.1]. More precisely, using the notation of that paper, we set

$$a = \frac{u}{K}, \quad z = (s, c), \quad f(a, z) = \frac{F_u(Ka, z)}{K}, \quad \text{and} \quad g(a, z) = (F_s(Ka, z), F_c(Ka, z)).$$

We have the estimates

$$\partial_a f = L_{uu} \geqslant \alpha \quad \text{and} \quad \partial_z g = \begin{bmatrix} L_{ss} & L_{sc} \\ L_{cs} & L_{cc} \end{bmatrix} \leqslant m,$$

in the sense of quadratic forms. Moreover, we have the estimates

$$\|\partial_z f\| \leqslant \frac{m}{K} \quad \text{and} \quad \|\partial_a g\| \leqslant Km.$$

Since

$$m + \frac{m}{K} + Km < 2m + \frac{m}{K} = \alpha,$$

we conclude that Hypothesis 2 of [64] is satisfied. Hypothesis 1 of [64] is verified by the domain Ω, and Hypothesis 3 is precisely the conclusion of Lemma A.2. As a consequence, we can apply Theorem 1.1 of [64], and conclude that the set W^{sc} is the graph of a C^1 and 1-Lipschitz map above $B^s \times \Omega^c$ in (a, z) coordinates, and therefore the graph of a K-Lipschitz C^1 map $w^{sc}: B^s \times \Omega^c \longrightarrow B^u$ in (u, s, c) coordinates.

In order to prove the statement concerning W^{uc}, we apply [64, Theorem 1.1] with

$$a = \frac{s}{K}, \quad z = (u, c), \quad f(a, z) = -\frac{F_s(Ka, z)}{K}, \quad \text{and} \quad g(a, z) = -(F_u(Ka, z), F_c(Ka, z)).$$

It is easy to check as above that all hypotheses are satisfied.

Let us now study the set $W^c = W^{sc} \cap W^{uc}$. First, let us prove that W^c is a C^1 graph above Ω^c. We know that W^{sc} is the graph of a K-Lipschitz C^1 function $w^{sc}(s, c)$ and that W^{uc} is the graph of a K-Lipschitz C^1 function $w^{uc}(u,c)$. The point (u,s,c) belongs to W^c if and only if

$$u=w^{sc}(s,c) \quad \text{and} \quad s=w^{uc}(u,c),$$

or in other words if and only if (u,s) is a fixed point of the K-Lipschitz C^1 map

$$(u,s)\longmapsto (w^{sc}(s,c), w^{uc}(u,c)).$$

For each c, this contracting map has a unique fixed point in $\overline{B}^u\times\overline{B}^s$, which corresponds to a point of $\overline{W}^{sc}\cap\overline{W}^{uc}$. It follows from Lemma A.2 that this point is contained in $B^u\times B^s$. Then, it depends in a C^1 way on the parameter c. We have proved that W^c is the graph of a C^1 function w^c. In order to estimate the Lipschitz constant of this graph, we consider two points (u_i,s_i,c_i), $i=0,1$, in W^c. We have

$$\|u_1-u_0\|^2\leqslant K^2(\|s_1-s_0\|^2+\|c_1-c_0\|^2)$$

and

$$\|s_1-s_0\|^2\leqslant K^2(\|u_1-u_0\|^2+\|c_1-c_0\|^2).$$

Taking the sum gives

$$(1-K^2)(\|u_1-u_0\|^2+\|s_1-s_0\|^2)\leqslant 2K^2\|c_1-c_0\|^2$$

and

$$\|(u_1,s_1)-(u_0,s_0)\|\leqslant\sqrt{\frac{2K^2}{1-K^2}}\|c_1-c_0\|\leqslant 2K\|c_1-c_0\|,$$

since $K\leqslant 1/\sqrt{2}$. We conclude that w^c is $2K$-Lipschitz. □

It is useful to go a bit further in the study of the invariant manifold

$$W^c=\{(w_u^c(c), w_c^s(c), c)\}.$$

This manifold is a partially hyperbolic invariant set, hence by the usual theory, to each point $x\in W^c$ is attached a strong stable manifold $W^s(x)$ and a strong unstable manifold $W^u(x)$, which are C^1 (and even C^r if F is C^r). The manifolds $W^u(x), x\in W^c$, partition W^{uc}, although this partition is not usually a C^1 foliation. For each $x\in W^{uc}$, we denote by $E^u(x)$ the strong unstable space, which is the tangent space at x of the only unstable manifold $W^u(x_0)$ which contains x_0. We define the exponents

$$e_u:=-\sup_{\substack{x\in W^c\\ v\in E^u(x)}}\limsup_{t\to\infty}\frac{\log\|v(-t)\|}{t}=-\sup_{\substack{x\in W^{uc}\\ v\in E^u(x)}}\limsup_{t\to\infty}\frac{\log\|v(-t)\|}{t},$$

$$e_c^+:=\sup_{\substack{x\in W_c\\ v\in T_xW_c}}\limsup_{t\to\infty}\frac{\log\|v(t)\|}{t},$$

$$e_c^-:=\inf_{\substack{x\in W_c\\ v\in T_xW_c}}\liminf_{t\to\infty}\frac{\log\|v(t)\|}{t},$$

where $v(t)$ is the solution of the linearized equation $\dot{v}(t)=dF_{x(t)}\cdot v(t)$ with initial condition $v(0)=v$, and $x(t)$ is the solution of $\dot{x}(t)=F\circ x(t)$ starting from $x(0)=x$.

LEMMA A.3. *We have*

$$-m-2mK\leqslant e_c^-\leqslant e_c^+\leqslant m+2mK.$$

Proof. We consider an orbit $x(t)\in W^c$, and a variational orbit $v(t)=(u'(t),s'(t),c'(t))$ tangent to W^c. Observe that $\|(u',s')\|\leqslant 2K\|c'\|$ for each t, which implies that

$$\left|\frac{d}{dt}\|c'\|^2\right|=2|\langle c',L_{cu}u'+L_{cs}s'+L_{cc}c'\rangle|\leqslant 2(m+2Km)\|c'\|^2. \qquad \square$$

The next lemma implies that the manifolds $W^{s,c}(x)$ are the graphs of C^1 and K-Lipschitz maps $w_x^s\colon B^s\to B^u\times\Omega^c$ and $w_x^u\colon B^u\to B^s\times\Omega^c$.

LEMMA A.4. *If $x\colon]T^-,T^+[\to\Omega$ is an orbit of F, then the linearized equation $\dot{v}(t)=dF_{x(t)}\cdot v(t)$ preserves the cone $C^u=\{\|(s',c')\|\leqslant K\|u'\|\}$ in forward time, and the cone $C^s=\{\|(u',c')\|\leqslant K\|s'\|\}$ in backward time.*

We have $E^u(x)\subset C^u$ for each $x\in W^{uc}$, and $E^s(x)\subset C^s$ for each $x\in W^{sc}$.

Finally we have the estimate

$$e_u\geqslant\alpha-2mK>\tfrac{1}{2}\alpha.$$

Proof. Let $v(t)=(u'(t),s'(t),c'(t))$ be a solution of the linearized equation along $x(t)$. Then

$$\frac{d}{dt}\|u'\|^2=\langle u',L_{uu}u'+L_{us}s'+L_{uc}c'\rangle\geqslant\alpha\|u'\|^2-m\|(s',c')\|\|u'\|\geqslant(\alpha-mK)\|u'\|^2$$

(this estimate will also provide the desired growth rate in the unstable direction) and

$$\begin{aligned}\frac{d}{dt}\|(s',c')\|^2&=\langle s',L_{su}u'+L_{ss}s'+L_{sc}c'\rangle+\langle c',L_{cu}u'+L_{cs}s'+L_{cc}c'\rangle\\&\leqslant m\|(s',c')\|(\|u'\|+\|(s',c')\|)\leqslant mK(1+K)\|u'\|^2.\end{aligned}$$

This implies that

$$\frac{d}{dt}(K^2\|u'\|^2-\|(s',c')\|^2)\geqslant K^2\left(\alpha-mK-m-\frac{m}{K}\right)\|u'\|^2\geqslant 0,$$

recalling that $m+m/K+mK<\alpha$. The estimates concerning C^s are similar. $\square$

In general, the maps w_x^s and w_x^u are not better than Hölder continuous in x, but we can obtain better regularity under stronger hypotheses.

Theorem A.5. *In the context of Theorem A.1, let us assume the additional assumptions that F is C^2 and $K<\frac{1}{8}$ (or equivalently, $m<\frac{1}{6}\alpha$). Then each of the manifolds W^c, W^{uc}, W^{sc} is C^2, and the manifolds $W^u(x)$, $x\in W^c$, form a C^1 foliation of W^{uc} (similarly for W^s in W^{us}). The foliations are C^1 in the strongest possible sense, namely the map $x\mapsto E^u(x)$ is C^1 on E^{cu}, which imply that the foliation admits C^1 charts, and that the local holonomies are C^1.*

Proof. An easy computation shows that $m+2mK<\frac{1}{4}\alpha$, and hence we obtain

$$e_u>\tfrac{1}{2}\alpha,\quad e_c^+<\tfrac{1}{4}\alpha,\quad \text{and}\quad e_c^->-\tfrac{1}{4}\alpha.$$

This implies that $e_u>2e_c^+$, and so W^c is 2-normally hyperbolic, and hence it is C^2, as well as W^{uc} and W^{sc}; see [33] and [37].

Moreover, we have the bunching condition $e_u>e_c^+-e_c^-$, which implies the C^1 regularity of the unstable foliation; see [34], [58], and [29]. □

We need the following easy addendum.

Proposition A.6. *Assume in addition that there exists a translation g of $\mathbb{R}^{n_c}$ such that*

$$g(\Omega^c)=\Omega^c\quad \textit{and}\quad F\circ(\mathrm{id}\otimes\mathrm{id}\otimes g)=F.$$

Then we have

$$w^{sc}\circ(\mathrm{id}\otimes g)=w^{sc},\quad w^{uc}\circ(\mathrm{id}\otimes g)=w^{uc},\quad \textit{and}\quad w^c\circ g=w^c.$$

Proof. It follows immediately from the definition of the sets W^{sc}, W^{uc}, and W^c that $g(W^{sc})=W^{sc}$, $g(W^{uc})=W^{uc}$, and $g(W^c)=W^c$. □

In applications the first condition of Hypothesis 3 is usually not satisfied, except in the case where $\Omega^c=\mathbb{R}^{n_c}$. In view of the applications we have in mind, it is useful to split the central variables into two groups and consider

$$\Omega^c=\mathbb{R}^{n_c^1}\times\Omega^{c_2},$$

where Ω^{c_2} is a convex open set in $\mathbb{R}^{n_c^2}$, $n_c^1+n_c^2=n_c$. Given a positive parameter σ, let $\Omega^{c_2}_\sigma$ be the set of points $c_2\in\mathbb{R}^{n_c^2}$ such that $d(c,\Omega^{c_2})<\sigma$. This is a convex open subset of $\mathbb{R}^{n_c^2}$ containing Ω^{c_2}. We denote the product $\mathbb{R}^{n_c^1}\times\Omega^{c_2}_\sigma$ by Ω^c_σ and $B^u\times B^s\times\Omega^c_\sigma$ by Ω_σ. With the notation $F_c=(F_{c_1},F_{c_2})$, and denoting by $W^{sc}(F,\Omega), W^{uc}(F,\Omega), W^c(F,\Omega)$ the set of positive half orbits (resp. negative half orbits, full orbits) of F contained in Ω, we have the following result.

PROPOSITION A.7. *Let* $F:\mathbb{R}^{n_u}\times\mathbb{R}^{n_s}\times\Omega^c_\sigma\to\mathbb{R}^{n_u}\times\mathbb{R}^{n_s}\times\mathbb{R}^{n_c}$ *be a* C^2 *vector field. Assume that there exist* $\lambda, m, \sigma>0$ *such that*

- $F_u(u,s,c)\cdot u>0$ *on* $\partial B^u\times\overline{B}^s\times\overline{\Omega}^c_\sigma$;
- $F_s(u,s,c)\cdot s<0$ *on* $\overline{B}^u\times\partial B^s\times\overline{\Omega}^c_\sigma$;
- $L_{uu}(x)\geqslant\alpha I$ *and* $L_{ss}(x)\leqslant-\alpha I$ *for each* $x\in\Omega_\sigma$ *in the sense of quadratic forms*;
- *for each* $x\in\Omega_\sigma$,

$$\|L_{us}(x)\|+\|L_{uc}(x)\|+\|L_{ss}(x)\|+\|L_{sc}(x)\|+\|L_{cu}(x)\|+\|L_{cs}(x)\|+\|L_{cc}(x)\|\leqslant m;$$

- *for each* $x\in\Omega_\sigma\setminus\Omega$,

$$\begin{aligned}&\|L_{us}(x)\|+\|L_{uc}(x)\|+\|L_{ss}(x)\|+\|L_{sc}(x)\|\\&\qquad+\|L_{cu}(x)\|+\|L_{cs}(x)\|+\|L_{cc}(x)\|+\frac{2\|F_{c_2}(x)\|}{\sigma}\leqslant m.\end{aligned}$$

Assume furthermore that

$$K:=\frac{m}{\alpha-2m}\leqslant\frac{1}{8},$$

then there exist C^2 *maps*

$$w^{sc}: B^s\times\Omega^c_\sigma\longrightarrow B^u,\quad w^{uc}: B^u\times\Omega^c_\sigma\longrightarrow B^s,\quad \textit{and}\quad w^c:\Omega^c_\sigma\longrightarrow B^u\times B^s$$

satisfying the estimates

$$\|dw^{sc}\|\leqslant K,\quad \|dw^{uc}\|\leqslant K,\quad \textit{and}\quad \|dw^c\|\leqslant 2K,$$

the graphs of which respectively contain $W^{sc}(F,\Omega)$, $W^{uc}(F,\Omega)$, *and* $W^c(F,\Omega)$. *Moreover, the graphs of the restrictions of* w^{sc}, w^{uc}, *and* w^c *to, respectively,* $B^s\times\Omega^c$, $B^u\times\Omega^c$, *and* Ω^c, *are tangent to the flow.*

There exists an invariant C^1 *foliation of the graph of* w^{uc} *whose leaves are graphs of* K*-Lipschitz maps above* B^u. *The set* $W^{uc}(F,\Omega)$ *is a union of leaves*: *it has the structure of an invariant* C^1 *lamination. Two points* x *and* x' *belong to the same leaf of this lamination if and only if* $d(x(t),x'(t))e^{t\alpha/4}$ *is bounded on* $\mathbb{R}^-$.

If in addition there exists a group G *of translations of* $\mathbb{R}^{n_{c_1}}$ *such that*

$$F\circ(\mathrm{id}\otimes\mathrm{id}\otimes g\otimes\mathrm{id})=F$$

for each $g\in G$, *then the maps* w^* *can be chosen such that*

$$w^{sc}\circ(\mathrm{id}\otimes g\otimes\mathrm{id})=w^{sc},\quad w^{uc}\circ(\mathrm{id}\otimes g\otimes\mathrm{id})=w^{uc},\quad \textit{and}\quad w^c\circ(g\otimes\mathrm{id})=w^c \tag{34}$$

for each $g\in G$. *The lamination is also translation invariant.*

In contrast to the earlier results of this section, the map w^{sc} is not uniquely defined, and neither is its restriction to $B^s\times\Omega^c$. Moreover, the intersection with Ω of the graph of w^{sc} is not necessarily positively invariant. It can contain strictly the set $W^{sc}(F,\Omega)$. Similar remarks apply to w^{uc} and w^c.

Proof. We take a function $\varrho:\Omega^{c_2}_\sigma\longrightarrow[0,1]$ such that

- $\varrho=0$ near the boundary of $\Omega^{c_2}_\sigma$;
- $\varrho=1$ on Ω^{c_2};
- $\|d\varrho\|\leqslant 2/\sigma$ uniformly.

We claim that the vector field

$$\widetilde{F}(u,s,c):=(F_u(u,s,c_1,c_2),F_s(u,s,c_1,c_2),F_{c_1}(u,s,c_1,c_2),\varrho(c_2)F_{c_2}(u,s,c_1,c_2))$$

satisfies all the hypotheses of Theorem A.1 on Ω_σ. Note also that $\widetilde{F}=F$ on Ω. Denoting by $\tilde{L}_{**}$ the variational matrix associated with $\widetilde{F}$, we see that

$$\begin{aligned}\tilde{L}_{cu}(u,s,c)&=\varrho(c_2)L_{cu}(u,s,c), & \tilde{L}_{cs}(u,s,c)&=\varrho(c_2)L_{cs}(u,s,c),\\ \tilde{L}_{c_1c_1}(u,s,c)&=\varrho(c_2)L_{c_1c_1}(u,s,c), & \tilde{L}_{c_1c_2}(u,s,c)&=\varrho(c_2)L_{c_1c_2}(u,s,c),\end{aligned}$$

and

$$\tilde{L}_{c_2c_2}(u,s,c)=\varrho(c_2)L_{c_2c_2}(u,s,c)+d\varrho(c_2)\otimes F_{c_2}(u,s,c).$$

As a consequence, we have

$$\begin{aligned}&\|\tilde{L}_{us}(x)\|+\|\tilde{L}_{uc}(x)\|+\|\tilde{L}_{ss}(x)\|+\|\tilde{L}_{sc}(x)\|+\|\tilde{L}_{cu}(x)\|+\|\tilde{L}_{cs}(x)\|+\|\tilde{L}_{cc}(x)\|\\ &\quad=\varrho(c_2)(\|L_{us}(x)\|+\|L_{uc}(x)\|+\|L_{ss}(x)\|+\|L_{sc}(x)\|+\|L_{cu}(x)\|+\|L_{cs}(x)\|+\|L_{cc}(x)\|)\\ &\qquad\quad+\|F_{c_2}(x)\|\,\|d\varrho(c_2)\|\\ &\quad\leqslant m.\end{aligned}$$

The claim is proved. We define w^{sc}, w^{uc}, and w^c as the maps given by Theorem A.1 applied to $\widetilde{F}$ on Ω_σ. Since $\widetilde{F}=F$ on Ω, we have $W^*(F,\Omega)\subset W^*(\widetilde{F},\Omega_\sigma)$ for $*\in\{sc,uc,c\}$. These maps may depend on the choice of the function ϱ but, once the function ϱ is chosen, they are uniquely defined. In the case where a group G of translations exists as in the statement, then we have $\widetilde{F}\circ(\mathrm{id}\otimes\mathrm{id}\otimes g\otimes\mathrm{id})=\widetilde{F}$ for each $g\in G$. The uniqueness then implies (34). By definition, $W^*(\widetilde{F},\Omega_\sigma)$ is the graph of w^*, the statement follows from this observation. □

Appendix B. Disconnectedness of heteroclinic orbits

We consider a Tonelli Hamiltonian H, a cohomology c, and the associated Aubry and Mañé sets $\tilde{\mathcal{A}}$ and $\widetilde{\mathcal{N}}$. We assume that the Aubry set is the union of two static classes $\widetilde{\mathcal{S}}_i, i=1,2$. The Mañé set can then be written as the disjoint union

$$\widetilde{\mathcal{N}} = \widetilde{\mathcal{S}}_1 \cup \widetilde{\mathcal{S}}_2 \cup \widetilde{\mathcal{H}}_{12} \cup \widetilde{\mathcal{H}}_{21},$$

where $\widetilde{\mathcal{H}}_{12}$ is a set of heteroclinic orbits from $\widetilde{\mathcal{S}}_1$ to $\widetilde{\mathcal{S}}_2$, and $\widetilde{\mathcal{H}}_{21}$ is a set of heteroclinic orbits from $\widetilde{\mathcal{S}}_2$ to $\widetilde{\mathcal{S}}_1$. Morever, the sets

$$\widetilde{\mathcal{I}}_{12} := \widetilde{\mathcal{S}}_1 \cup \widetilde{\mathcal{S}}_2 \cup \widetilde{\mathcal{H}}_{12} \quad \text{and} \quad \widetilde{\mathcal{I}}_{21} := \widetilde{\mathcal{S}}_1 \cup \widetilde{\mathcal{S}}_2 \cup \widetilde{\mathcal{H}}_{21}$$

are invariant compact Lipschitz graphs. In the notations of [8], we have

$$\widetilde{\mathcal{I}}_{12} = \widetilde{\mathcal{I}}(E_{\mathcal{S}_1}) = E_{\mathcal{S}_1} \wedge \breve{E}_{\mathcal{S}_2} \quad \text{and} \quad \widetilde{\mathcal{I}}_{21} = \widetilde{\mathcal{I}}(E_{\mathcal{S}_2}) = E_{\mathcal{S}_2} \wedge \breve{E}_{\mathcal{S}_1}.$$

In [8, §9] it is proved that the cohomology c is in the interior of its forcing class provided each of the sets $\widetilde{\mathcal{H}}_{12}$ and $\widetilde{\mathcal{H}}_{21}$ is neat in the following sense:

The set $\widetilde{\mathcal{H}}_{12}$ is neat if there exists a compact subset $\tilde{\mathcal{K}}_{12}$ which contains one and only one point in each orbit of $\varphi_{|\widetilde{\mathcal{H}}_{12}}$ and which is acyclic, which means that there exists an open neighborhood U of $\mathcal{K}_{12}$ in TM such that the inclusion of U into TM generates the null map in homology.

In §1.4 of the present paper, we apply this result under the assumption that the sets $\widetilde{\mathcal{H}}_{12}$ and $\widetilde{\mathcal{H}}_{21}$ are totally disconnected. We can do so in view of the following.

PROPOSITION B.1. *The set $\widetilde{\mathcal{H}}_{12}$ (or $\widetilde{\mathcal{H}}_{21}$) is neat if it is totally disconnected.*

Proof. We first recall that a compact metric space is totally disconnected if and only if it has dimension zero, which means that each of its points has a basis of neighborhood made of open and closed sets, see [38, §II.4].

By removing small open neighborhoods of $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$ in $\widetilde{\mathcal{I}}_{12}$, we form a compact subset of $\widetilde{\mathcal{H}}_{12}$ which contains at least one point in each orbit. This compact subset is totally disconnected (it is a subset of $\widetilde{\mathcal{H}}_{12}$) hence each of its points is contained in an open and closed set which is disjoint from both $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$. We cover our compact by finitely many of these neighborhoods. Their union is a compact and open subset $\widetilde{\mathcal{Q}}$ of $\widetilde{\mathcal{H}}_{12}$ which contains at least one point in each orbit. The set $\tilde{\mathcal{K}}_{12} := \widetilde{\mathcal{Q}} - \varphi(\widetilde{\mathcal{Q}})$ is then compact and open, and it contains exactly one point of each φ-orbit. It is totally disconnected, and therefore acyclic, in view of the following lemma. □

LEMMA B.2. *Let M be a manifold and let $K \subset M$ be a totally disconnected compact subset of M. Then K is acyclic.*

Proof. The subset K has dimension zero; see [38]. As a consequence, each point of K is contained in an open, closed, and acyclic neighborhood (small open sets are contained in discs hence are acyclic). We cover K by finitely many of these subsets $U_1, ..., U_k$ and set $V_1 = U_1$, $V_2 = U_2 \setminus V_1$ and $V_i = U_i \setminus V_{i-1}$. We obtain k open acyclic subsets V_i which are pairwise disjoint and cover K. This implies that K is acyclic. □

Appendix C. Continuity property of the Peierls' barrier function

We consider here a general Tonelli Lagrangian L. We recall, see [7, §4], that the difference of two weak KAM solutions is constant on each static class.

Proposition C.1. *Let $L_k \to L$ be a sequence of Tonelli Lagrangians $\mathbb{T}^n \times \mathbb{R}^n \times \mathbb{T}$ converging in the C^2 compact open topology, and $c_k \to c \in \mathbb{R}^n \simeq H^1(\mathbb{T}^n, \mathbb{R})$. Assume that $\mathcal{A}_L(c)$ has finitely many static classes. Let $\zeta_k \in \mathcal{A}_{L_k}(c_k)$ be such that $\zeta_k \to \zeta_0 \in \mathcal{A}_L(c)$, then for any $\theta \in \mathbb{T}^n$, one has*

$$\lim_{k\to\infty} h_{L_k, c_k}(\zeta_k, \theta) = h_c(\zeta_0, \theta).$$

Proof. First, since each $A^M_{L,c}(\theta_1, \theta_2)$ is continuous in L and c, we obtain

$$\varlimsup_{k\to\infty} \Phi_{L_k, c_k}(\theta_1, \theta_2) \leqslant \lim_{k\to\infty} (A^M_{L_k, c_k}(\theta_1, \theta_2)) = A^M_{L,c}(\theta_1, \theta_2)$$

taking infimum over N, we get $\varlimsup_{k\to\infty} \Phi_{L_k, c_k}(\theta_1, \theta_2) \leqslant \Phi_{L,c}(\theta_1, \theta_2)$. Since $h_{L,c}(\theta_1, \theta_2) = \Phi_{L,c}(\theta_1, \theta_2)$ if either θ_1 or θ_2 is in $\mathcal{A}_L(c)$, we obtain

$$\varlimsup_{k\to\infty} h_{c_k}(\zeta_k, \theta) \leqslant h_c(\zeta_0, \theta).$$

Given $\varepsilon_k \to 0$, let $\gamma_k\colon [-Q_k, 0] \to \mathbb{T}^n$ be a sequence of extremal curves such that

$$\gamma_k(-Q_k) = \zeta_k, \quad \gamma_k(0) = \theta$$

and

$$A^{Q_k}_{L_k, c_k}(\zeta_k, \theta) \leqslant h_{L_k, c_k}(\zeta_k, \theta) + \varepsilon_k.$$

We note that on each interval $[i, j] \subset [-Q_k, 0]$ we have

$$\begin{aligned} &A^{j-i}_{L_k, c_k}(\gamma_k(i), \gamma_k(j)) \\ &\quad = A^{Q_k}(\gamma_k(-Q_k), \gamma_k(0)) - A^{i+Q_k}(\gamma_k(-Q_k), \gamma_k(i)) - A^{-j}(\gamma_k(j), \gamma_k(0)) \\ &\quad \leqslant h(\zeta_k, \theta) - \varepsilon_k - (h(\zeta_k, \gamma_k(i)) - h(\zeta_k, \gamma_k(-Q_k))) - (h(\zeta_k, \gamma_k(0)) - h(\zeta_k, \gamma_k(j))) \\ &\quad \leqslant h_{L_k, c_k}(\zeta_k, \gamma_k(j)) - h_{L_k, c_k}(\zeta_k, \gamma_k(i)) + \varepsilon_k, \end{aligned} \tag{35}$$

since $h(\zeta_k,\gamma_k(-Q_k))=h(\zeta_k,\zeta_k)=0$ and $\gamma_k(0)=\theta$. Note that we omit the subscript L_k and c_k in the intermediate calculations.

Let i_k and i'_k be two consecutive visits of $\gamma_k(i)$ to $U=B_\delta(\mathcal{A}_L(c))$, we first show that i'_k-i_k must be bounded as $k\to\infty$. Assume otherwise, then the curves

$$\gamma_k(t+i_k+1)|_{[0,i'_k-i_k-2]}$$

converge uniformly over compact sets to $\gamma_*:[0,\infty)\to\mathbb{T}^n$. Assume that the weak KAM solutions $h_{L_k,c_k}(\zeta_k,\cdot)$ converge uniformly to a weak KAM solution u of L at the cohomology c; taking limit in (35) implies that γ_* must be calibrated by u. Therefore γ_* must accumulate to $\mathcal{A}_L(c)$, which is a contradiction.

Let $\mathcal{S}_1,...,\mathcal{S}_r$ be the static classes of $\mathcal{A}(L)$. Let $U_q=B_\delta(\mathcal{S}_q)$ and assume that δ is small enough so that U_q are all disjoint. Let us note that each γ_k determines sequences $q_s\in\{1,...,r\}$, $s=1,...,r$, and $0=i_0\leqslant j_0\leqslant...\leqslant i_r\leqslant j_r\leqslant Q_k$ as follows:

- set $i_0=j_0=0$;
- let i_1 be the first visit of $\gamma(-i)$ to $\bigcup_q U_q$ and U_{q_1} is the set that $\gamma(-i_1)$ visits; let j_1 be the last visit to U_{q_1}, namely $j_1=\max\{i:\gamma(-i)\in U_{q_1)}\}$;
- the process stops if $j_{s-1}=-Q_k$; we then set set

$$i_s=j_s=...=i_r=j_r=Q_k \quad \text{and} \quad q_s=...=q_r=q_{s-1}.$$

Otherwise, let i_s be the first visits to $\bigcup_q U_q$ for $i>j_{s-1}$, and U_{q_s} the set it visits. Define j_s to be the last visit to U_{q_s} and continue.

Then,

$$\begin{aligned} h_{L_k,c_k}(\zeta_k,\theta)+\varepsilon_k & \\ \geqslant A^{Q_k}_{L_k,c_k}(\gamma_k(-Q_k),\gamma_k(0)) & \\ =\sum_{s=1}^r A^{i_s-j_{s-1}}(\gamma_k(-i_s),\gamma_k(-j_{s-1}))+\sum_{s=1}^r A^{j_s-i_s}(\gamma_k(-j_s),\gamma_k(-i_s)) & \\ \geqslant\sum_{s=1}^r A^{i_s-j_{s-1}}(\gamma_k(-i_s),\gamma_k(-j_{s-1}))+\sum_{s=1}^r(h(\zeta_k,\gamma_k(-i_s))-h(\zeta_k,\gamma_k(-j_s)))-r\varepsilon_k, & \end{aligned} \tag{36}$$

where the subscript L_k,c_k was omitted in the last two lines. By restricting to a subsequence, we may assume that, for all γ_k, the ordering of $q_1,...\,q_r$ is identical. Our previous observation implies, for $s=1,...,r$, that i_s-j_{s-1} are bounded as $k\to\infty$. By restricting to another subsequence, we may assume i_s-j_{s-1} is constant for all k, and $\gamma_k(-i_s)\to\theta_s$ and $\gamma_k(-j_s)\to\theta'_s$ as $k\to\infty$. Note that, for $s=1,...,r$, $\theta_s,\theta'_s\in\overline{B_\delta(\mathcal{S}_{q_s})}$, therefore, there exist $\eta_s,\eta'_s\in\mathcal{S}_{q_s}$ such that $\|\theta_s-\eta_s\|,\|\theta'_s-\eta'_s\|\leqslant\delta$. Let us also note, by definition, that $\theta_0=\theta'_0=\theta$

and $\theta_r=\theta'_r=\zeta_0$. Define $\eta_0=\eta'_0=\theta$ and $\eta_r=\eta'_r=\zeta_0$. Up to taking a subsequence, assume that the weak KAM solutions $h_{L_k,c_k}(\zeta_k,\cdot)$ converge to $u(\cdot)$ uniformly. Taking the limit as $k\to\infty$ in (36), we obtain

$$\begin{aligned}\lim_{k\to\infty} h_{L_k,c_k}(\zeta_k,\theta) &\geqslant \sum_{s=1}^{r}(A_{L,c}^{i_s-j_{s-1}}(\theta_s,\theta'_{s-1})+u(\theta_s)-u(\theta'_s))\\ &\geqslant \sum_{s=1}^{r}(A_{L,c}^{i_s-j_{s-1}}(\eta_s,\eta'_{s-1})+u(\eta_s)-u(\eta'_s)-4C\delta)\\ &= \sum_{s=1}^{r}(A_{L,c}^{i_s-j_{s-1}}(\eta_s,\eta'_{s-1})+h_{L,c}(\zeta_0,\eta_s)-h_{L,c}(\zeta_0,\eta'_s)-4C\delta)\\ &\geqslant \sum_{s=1}^{r}(h_{L,c}(\zeta_0,\eta'_{s-1})-h_{L,c}(\zeta_0,\eta'_s)-4C\delta)\\ &= h_{L,c}(\zeta_0,\eta'_0)-h_{L,c}(\zeta_0,\eta'_s)-4rC\delta\\ &= h_{L,c}(\zeta_0,\theta)-4rC\delta.\end{aligned}$$

Since δ is arbitrary, we obtain $\lim_{k\to\infty} h_{L_k,c_k}(\zeta_k,\theta)\geqslant h_{L,c}(\zeta_0,\theta)$. □

Acknowledgements. P. B. was supported by the European Research Council under the European Union's 7th Framework Programme (FP/2007-2013)/ERC, Grant Agreement 307062. V. K. was partially supported by the NSF grant DMS-1402164. K. Z. was supported by the NSERC Discovery grant, reference number 436169-2013.

References

[1] Arnol′d, V. I., Instability of dynamical systems with many degrees of freedom. *Dokl. Akad. Nauk SSSR*, 156 (1964), 9–12 (Russian); English translation in *Soviet Math. Dokl.*, 5 (1964), 581–585.
[2] — *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics, 60. Springer, New York, 1989.
[3] Arnol′d, V. I., Kozlov, V. V. & Neishtadt, A. I., *Mathematical Aspects of Classical and Celestial Mechanics*. Springer, Berlin–Heidelberg, 1997.
[4] Bangert, V., Minimal measures and minimizing closed normal one-currents. *Geom. Funct. Anal.*, 9 (1999), 413–427.
[5] Bernard, P., Perturbation d'un hamiltonien partiellement hyperbolique. *C. R. Acad. Sci. Paris Sér. I Math.*, 323 (1996), 189–194.
[6] — Connecting orbits of time dependent Lagrangian systems. *Ann. Inst. Fourier* (*Grenoble*), 52 (2002), 1533–1568.
[7] — Symplectic aspects of Mather theory. *Duke Math. J.*, 136 (2007), 401–420.
[8] — The dynamics of pseudographs in convex Hamiltonian systems. *J. Amer. Math. Soc.*, 21 (2008), 615–669.
[9] — Arnold's diffusion: from the *a priori* unstable to the *a priori* stable case, in *International Congress of Mathematicians*. Vol. III, pp. 1680–1700. Hindustan Book Agency, New Delhi, 2010.

[10] — Large normally hyperbolic cylinders in a priori stable Hamiltonian systems. *Ann. Henri Poincaré*, 11 (2010), 929–942.

[11] — On the number of Mather measures of Lagrangian systems. *Arch. Ration. Mech. Anal.*, 197 (2010), 1011–1031.

[12] Bernard, P. & Contreras, G., A generic property of families of Lagrangian systems. *Ann. of Math.*, 167 (2008), 1099–1108.

[13] Berti, M., Biasco, L. & Bolle, P., Drift in phase space: a new variational mechanism with optimal diffusion time. *J. Math. Pures Appl.*, 82 (2003), 613–664.

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

[15] Bessi, U., An approach to Arnol′d's diffusion through the calculus of variations. *Nonlinear Anal.*, 26 (1996), 1115–1135.

[16] — Arnold's diffusion with two resonances. *J. Differential Equations*, 137 (1997), 211–239.

[17] Bounemoura, A., An example of instability in high-dimensional Hamiltonian systems. *Int. Math. Res. Not. IMRN*, 2012 (2012), 685–716.

[18] Bourgain, J. & Kaloshin, V., On diffusion in high-dimensional Hamiltonian systems. *J. Funct. Anal.*, 229 (2005), 1–61.

[19] Chaperon, M., The Lipschitzian core of some invariant manifold theorems. *Ergodic Theory Dynam. Systems*, 28 (2008), 1419–1441.

[20] Cheng, C.-Q., Arnold diffusion in nearly integrable hamiltonian systems. Preprint, 2012. arXiv:1207.4016 [math.DS].

[21] Cheng, C.-Q. & Li, X., Connecting orbits of autonomous Lagrangian systems. *Nonlinearity*, 23 (2010), 119–141.

[22] — Variational construction of unbounded orbits in Lagrangian systems. *Sci. China Math.*, 53 (2010), 617–624.

[23] Cheng, C.-Q. & Yan, J., Existence of diffusion orbits in a priori unstable Hamiltonian systems. *J. Differential Geom.*, 67 (2004), 457–517.

[24] — Arnold diffusion in Hamiltonian systems: a priori unstable case. *J. Differential Geom.*, 82 (2009), 229–277.

[25] Contreras, G. & Paternain, G. P., Connecting orbits between static classes for generic Lagrangian systems. *Topology*, 41 (2002), 645–666.

[26] Cresson, J., Symbolic dynamics and Arnold diffusion. *J. Differential Equations*, 187 (2003), 269–292.

[27] Delshams, A., Gidea, M., de la Llave, R. & Seara, T. M., Geometric approaches to the problem of instability in Hamiltonian systems. An informal presentation, in *Hamiltonian Dynamical Systems and Applications*, NATO Sci. Peace Secur. Ser. B Phys. Biophys., pp. 285–336. Springer, Dordrecht, 2008.

[28] Delshams, A. & Huguet, G., Geography of resonances and Arnold diffusion in a priori unstable Hamiltonian systems. *Nonlinearity*, 22 (2009), 1997–2077.

[29] Delshams, A., de la Llave, R. & Seara, T. M., 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 (2006).

[30] Fathi, A., Orbites hétéroclines et ensemble de Peierls. *C. R. Acad. Sci. Paris Sér. I Math.*, 326 (1998), 1213–1216.

[31] — Weak KAM Theorem in Lagrangian Dynamics. Draft book, 2008.

[32] Fathi, A. & Siconolfi, A., Existence of C^1 critical subsolutions of the Hamilton–Jacobi equation. *Invent. Math.*, 155 (2004), 363–388.

[33] Fenichel, N., Persistence and smoothness of invariant manifolds for flows. *Indiana Univ. Math. J.*, 21 (1971/1972), 193–226.

[34] — Asymptotic stability with rate conditions. II. *Indiana Univ. Math. J.*, 26 (1977), 81–93.
[35] Gidea, M. & Robinson, C., Shadowing orbits for transition chains of invariant tori alternating with Birkhoff zones of instability. *Nonlinearity*, 20 (2007), 1115–1143.
[36] — Obstruction argument for transition chains of tori interspersed with gaps. *Discrete Contin. Dyn. Syst. Ser. S*, 2 (2009), 393–416.
[37] Hirsch, M. W., Pugh, C. C. & Shub, M., *Invariant Manifolds*. Lecture Notes in Mathematics, 583. Springer, Berlin–New York, 1977.
[38] Hurewicz, W. & Wallman, H., *Dimension Theory*. Princeton Mathematical Series, 4. Princeton University Press, Princeton, NJ, 1941.
[39] Kaloshin, K. V. Z, Normally hyperbolic invariant manifolds near strong double resonance. Preprint, 2012. arXiv:1202.1032 [math.DS].
[40] — A strong form of Arnold diffusion for two and a half degrees of freedom. Preprint, 2012. arXiv:1212.1150 [math.DS].
[41] Kaloshin, V. & Levi, M., An example of Arnold diffusion for near-integrable Hamiltonians. *Bull. Amer. Math. Soc.*, 45 (2008), 409–427.
[42] — Geometry of Arnold diffusion. *SIAM Rev.*, 50 (2008), 702–720.
[43] Kaloshin, V., Levi, M. & Saprykina, M., Arnol′d diffusion in a pendulum lattice. *Comm. Pure Appl. Math.*, 67 (2014), 748–775.
[44] Kaloshin, V. & Saprykina, M., An example of a nearly integrable Hamiltonian system with a trajectory dense in a set of maximal Hausdorff dimension. *Comm. Math. Phys.*, 315 (2012), 643–697.
[45] Kaloshin, V., Zhang, K. & Zheng, Y., Almost dense orbit on energy surface, in *XVI International Congress on Mathematical Physics*, pp. 314–322. World Sci., Hackensack, NJ, 2010.
[46] Lochak, P. & Marco, J.-P., Diffusion times and stability exponents for nearly integrable analytic systems. *Cent. Eur. J. Math.*, 3 (2005), 342–397.
[47] Mañé, R., Generic properties and problems of minimizing measures of Lagrangian systems. *Nonlinearity*, 9 (1996), 273–310.
[48] Marco, J.-P., Generic hyperbolic properties of nearly integrable systems on a^3. Preprint, 2012.
[49] Marco, J.-P. & Sauzin, D., Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems. *Publ. Math. Inst. Hautes Études Sci.*, 96 (2002), 199–275.
[50] Mather, J. N., Variational construction of connecting orbits. *Ann. Inst. Fourier* (*Grenoble*), 43 (1993), 1349–1386.
[51] — *Hamiltonian Systems*. Graduate class. Princeton University, 1998–2000.
[52] — Arnold diffusion. I. Announcement of results. *Sovrem. Mat. Fundam. Napravl.*, 2 (2003), 116–130 (Russian); English translation in *J. Math. Sci.* (N. Y.), 124 (2004), 5275–5289.
[53] — Arnold diffusion II. Preprint, 2008.
[54] — Lecture course on Arnold diffusion. Lecture notes. University of Maryland, 2010.
[55] McDuff, D. & Salamon, D., *Introduction to Symplectic Topology*. Oxford Mathematical Monographs. Oxford University Press, New York, 1995.
[56] McGehee, R., The stable manifold theorem via an isolating block, in *Symposium on Ordinary Differential Equations* (Minneapolis, MN, 1972), Lecture Notes in Math., 312, pp. 135–144. Springer, Berlin–Heidelberg, 1973.
[57] Moeckel, R., Transition tori in the five-body problem. *J. Differential Equations*, 129 (1996), 290–314.
[58] Pugh, C., Shub, M. & Wilkinson, A., Hölder foliations. *Duke Math. J.*, 86 (1997), 517–546.

[59] Rifford, L. & Ruggiero, R. O., Generic properties of closed orbits of Hamiltonian flows from Mañé's viewpoint. *Int. Math. Res. Not. IMRN*, 2012 (2012), 5246–5265.

[60] Salamon, D. & Zehnder, E., KAM theory in configuration space. *Comment. Math. Helv.*, 64 (1989), 84–132.

[61] Treschev, D., Multidimensional symplectic separatrix maps. *J. Nonlinear Sci.*, 12 (2002), 27–58.

[62] — Evolution of slow variables in a priori unstable Hamiltonian systems. *Nonlinearity*, 17 (2004), 1803–1841.

[63] Xia, Z., Arnold diffusion: a variational construction, in *Proceedings of the International Congress of Mathematicians* (Berlin, 1998), Vol. II, pp. 867–877. Documenta Mathematica, Bielefeld, 1998.

[64] Yang, D. G., An invariant manifold for ODEs and its applications. Preprint, 2009. arXiv:0909.1103 [math.DS].

[65] Zhang, K., Speed of Arnold diffusion for analytic Hamiltonian systems. *Invent. Math.*, 186 (2011), 255–290.

[66] Zheng, Y., Arnold diffusion for a-priori unstable systems and a five-body problem. Preprint, 2009.

Patrick Bernard
PSL Research University
École Normale Supérieure
Département de Mathématiques
et Applications UMR CNRS 8553
45 rue d'Ulm
FR-75230 Paris cedex 05
France
patrick.bernard@ens.fr

Ke Zhang
University of Toronto
Department of Mathematics
40 St. George St.
Toronto, ON, M5S 2E4
Canada
kzhang@math.toronto.edu

Vadim Kaloshin
University of Maryland at College Park
Department of Mathematics
3111 Math. Building
College Park, MD 20740
U.S.A.
vadim.kaloshin@gmail.com

Received April 4, 2013
Received in revised form September 28, 2016