Central European Science Journals
w w w . c e s j . c o m

Central European Journal of Mathematics
Central European Science Journals
CEJM 3(3) 2005 342–397

Diffusion times and stability exponents for nearly integrable analytic systems

Pierre Lochak*, Jean-Pierre Marco†

Analyse algébrique,
Université Paris VI, UMR 7586,
4 place Jussieu, 75252 Paris, cedex 05, France

Received 20 December 2004; accepted 30 March 2005

Abstract: For a positive integer n and $R > 0$, we set $B^n_R = \{x \in \mathbb{R}^n \mid \|x\|_\infty < R\}$. Given $R > 1$ and $n \geq 4$ we construct a sequence of analytic perturbations (H_j) of the completely integrable Hamiltonian $h(r) = \frac{1}{2}r_1^2 + \cdots \frac{1}{2}r_{n-1}^2 + r_n$ on $\mathbb{T}^n \times B^n_R$, with unstable orbits for which we can estimate the time of drift in the action space. These functions H_j are analytic on a fixed complex neighborhood V of $\mathbb{T}^n \times B^n_R$, and setting $\varepsilon_j := \|h - H_j\|_{C^0(V)}$ the time of drift of these orbits is smaller than $\exp(c(1/\varepsilon_j)^{1/2(n-3)})$ for a fixed constant $c > 0$. Our unstable orbits stay close to a doubly resonant surface, the result is therefore almost optimal since the stability exponent for such orbits is $1/2(n-2)$. An analogous result for Hamiltonian diffeomorphisms is also proved. Two main ingredients are used in order to deal with the analytic setting: a version of Sternberg's conjugacy theorem in a neighborhood of a normally hyperbolic manifold in a symplectic system, for which we give a complete (and seemingly new) proof; and Easton windowing method that allow us to approximately localize the wandering orbits and estimate their speed of drift.

Keywords: Perturbations, normal forms, small divisors, Arnol'd diffusion
MSC (2000): 37J40, 37B, 37D

1 Introduction and main results

The present work is devoted to the optimality of stability exponents for analytic quasi-convex near-integrable Hamiltonian systems, which amounts to the search for an example of an unstable orbit with the highest possible speed of drift.

* E-mail: lochak@math.jussieu.fr
† E-mail: marco@math.jussieu.fr

We begin with a short reminder on stability over exponentially long times (as pioneered by N.N.Nekhorhoshev) in the analytic and Gevrey categories and the optimaity problem for the stability exponents. We then state our main instability results in the framework of discrete as well as continuous systems.

1.1 The general problem

1.1.1 Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ and $\mathbb{A}^n = T^*\mathbb{T}^n = \mathbb{T}^n \times \mathbb{R}^n$ for a positive integer n. In this paper we deal with Hamiltonian systems close to an integrable one on the annulus $\mathbb{A}^n$, of the form

$$H(\theta, r) = h(r) + \varepsilon f(\theta, r),$$

which gives rise to the following vector field

$$X_H \left| \begin{array}{ll} \dot{\theta}_i = \partial_{r_i} h(r) + \varepsilon \partial_{r_i} f(\theta, r), & \\ & \\ \dot{r}_i = -\varepsilon \partial_{\theta_i} f(\theta, r), & \qquad i = 1, \ldots, n. \end{array} \right.$$

The canonical coordinates $(\theta, r) \in \mathbb{T}^n \times \mathbb{R}^n$ are angle-action coordinates for the integrable part h. When $\varepsilon = 0$, the actions r_i are first integrals of the system and the motion takes place on the corresponding invariant tori $\mathbb{T}^n \times \{r\}$, all the solutions being quasiperiodic.

For a generic real-analytic function h and for any real-analytic perturbation f, Nekhoroshev theorem [29] asserts that all solutions remain stable in action over exponentially long time intervals: there exist positive numbers a and b, depending only on h, such that for each small enough $\varepsilon > 0$ any initial condition (θ_0, r_0) gives rise to a solution $(\theta(t), r(t))$ which is defined at least for $|t| \leq \exp(\text{ const } (\frac{1}{\varepsilon})^a)$ and satisfies $\|r(t) - r(0)\| \leq$ const ε^b in that range.

When $n = 2$ and h is nondegenerate (or simply isoenergetically nondegenerate), the KAM Theorem yields more than Nekhoroshev theorem, since on each energy level the trajectories are confined on or between invariant tori. For $n \geq 3$ however, KAM tori do not *a priori* prevent the projection in action space of a solution from drifting arbitrarily far from its initial location; in this case Nekhoroshev theorem becomes fully relevant.

A main question now is to determine how large the stability exponents a and b can be taken in general. This is especially relevant for the first one: the larger a, the longer the time of stability guaranteed by the theorem. As for b, its value controls the closeness of the action variables ot their initial values.

1.1.2 The generic condition imposed by Nekhoroshev on the unperturbed Hamiltonian h is a transversality property called steepness. Here we will confine attention to quasiconvex functions h, which is a simple particular class of steep functions. Recall that a function h is quasiconvex when it has no critical points on its domain, and when there exists $m > 0$ such that, at any point r of that domain, the inequality $D^2h(r)(v, v) \geq m \|v\|^2$ holds for all vectors v orthogonal to $\nabla h(r)$.

As noticed by the Italian school ([5], [18], [4]), assuming the (quasi) convexity of the unperturbed Hamiltonian h yields refined results from the point of view of finite time stability. The introduction of simultaneous (diophantine) approximation, in conjunction with these remarks were the main ingredients in [22] which was designed to determine the best possible stability exponents a and b. As a result of this paper and minor subsequent improvements, one finds that if h is assumed to be quasiconvex, Nekhoroshev result holds with the exponents.

$$a = b = \frac{1}{2n}$$

as proved independently in [24, 25] and [30]; the latter paper actually takes up Nekhoroshev's original strategy and improves it to reach the above mentioned values. Note that the prediction of these values was part of the problem and comes from heuristic ideas of B.V.Chirikov, as formalized in [22] (see further references in that paper).

Moreover, again as predicted by B.V.Chirikov [14] and proved in [22] one can state *local* results in action space, near resonant surfaces. If $m \in \{1, \ldots, n-1\}$, a set of m independent linear relations with integer coefficients to be satisfied by the $\partial_{r_i} h(r)$ determines a *resonant surface of multiplicity* m in the action space. Given any $\varrho > 0$, for the trajectories starting at a distance less than $\varrho\, \varepsilon^{1/2}$ of such a surface one can take the larger exponents

$$a = b = \frac{1}{2(n-m)}.$$

This may be rather surprising at first sight as it shows that resonance enhances the stability of the nearby trajectories, at least over exponentially long times, whereas it is usually thought of as a cause of *in*stability.

1.1.3 The optimality question for the exponents amounts to looking for systems which are arbitrarily close to integrable, admit unstable orbits, *i.e.* orbits experiencing a drift in action independent of the size of the perturbation (we will let aside the role of the second exponent b), and such that one can prove an asymptotic *upper* bound for the time of drift which is close as possible to the lower bound $\exp(\,\mathrm{const}\ (\frac{1}{\varepsilon})^a)$ provided by the stability results.

In Arnold's famous note [1], an example of a three-degree-of-freedom system was proposed in view of exploring the complement of KAM tori in phase space, and instability was obtained from heteroclinic connections between whiskered tori.

It is by no means obvious that the diffusion time one obtains in Arnold's example (or immediate higher-dimensional generalizations) is comparable with the predictions of the stability theory. Again, the first heuristic arguments in this direction are to be found in [14], see also [22].

The first rigorous results on this problem were proved by U. Bessi. Making use of Arnold's model and a four degrees of freedom variant, he obtained in ([6, 7]) an answer for the optimality of the exponents in the cases $n = 3, 4$. He succeded in constructing orbits drifting in a time $\exp(\,\mathrm{const}\ (\frac{1}{\varepsilon})^{1/2})$ for $n = 3$, and $\exp(\,\mathrm{const}\ (\frac{1}{\varepsilon})^{1/4})$ for $n = 4$. These orbits pass close enough to a double resonance, thus the exponents cannot be improved for such trajectories; this shows that the exponent $1/2(n-2)$ for doubly-resonant surfaces is optimal when $n = 3$ or 4. It seems however difficult to generalize these ideas to higher dimensional systems, essentially due to the lack of a satisfactory higher dimensional analog of the continued fraction theory.

1.1.4 Following new ideas of Herman the framework was enlarged in [28] so as to be able to deal with Gevrey perturbations of integrable systems. Recall that given two real numbers $\alpha \geq 1$ and $L > 0$, and a positive radius R, a C^∞ function φ on $K = \mathbb{T}^n \times \overline{B}_\infty(0,R) \subset \mathbb{A}^n$ is said to be Gevrey-(α, L) on K when

$$\|\varphi\|_{\alpha,L} := \sum_{k\in\mathbb{N}^{2n}} \frac{L^{|k|\alpha}}{k!^\alpha} \left\|\partial^k\varphi\right\|_{C^0(K)} < \infty \tag{1.1}$$

with the usual notation for multi-indices of derivation: $|k| = k_1+\cdots+k_{2n}$, $k! = k_1!\cdots k_{2n}!$, $\partial^k = \partial_{x_1}^{k_1}\cdots\partial_{x_{2n}}^{k_{2n}}$. We denote by $G^{\alpha,L}(K)$ the Banach algebra formed by such functions. Real-analytic functions are recovered in the special case when $\alpha = 1$, the number L then indicates the size of a complex domain of analytic extension.

For $h \in G^{\alpha,L}(\overline{B}_R)$ which is quasi-convex and without critical point, it is proved in [28] that Nekhoroshev theorem holds with the exponents

$$a = \frac{1}{2(n-m)\alpha}, \qquad b = \frac{1}{2(n-m)},$$

for orbits passing close enough to resonant surfaces of multiplicity m.

In the same paper examples of unstable systems are constructed when the Gevrey exponent α is > 1. In this case the Gevrey class G^α is effectively larger than the space of real analytic functions, as it contains compactly supported functions which gives a lot of flexibility in the construction of examples. The main result of [28] goes as follows: let $n \geq 3$ and $\alpha > 1$, and set

$$a^* = \frac{1}{2(n-2)\alpha}.$$

Given $L > 0$ and $R > 1$, there exist a sequence of functions $(f_j)_{j\geq 0}$ converging to 0 in the space $G^{\alpha,L}(\mathbb{T}^n \times \overline{B}_R)$ and an increasing sequence of integers $(\tau_j)_{j\geq 0}$ such that, for each $j \geq 0$, the Hamiltonian system generated by

$$\mathcal{H}_j(\theta, r) = \frac{1}{2}(r_1^2 + \cdots + r_{n-1}^2) + r_n + f_j(\theta, r)$$

admits a solution $(\theta(t), r(t))$ defined at least for $t \in [0, \tau_j]$ and for which $r_1(0) = 0$ and $r_1(\tau_j) = 1$. Moreover, there exist positive constants $C_1 < C_2$ such that the time of drift τ_j and the norm $\varepsilon_j = \|f_j\|_{\alpha,L}$ are related by

$$\frac{C_1}{\varepsilon_j^2} \exp\Big(C_1\big(\frac{1}{\varepsilon_j}\big)^{a^*}\Big) \leq \tau_j \leq \frac{C_2}{\varepsilon_j^2} \exp\Big(C_2\big(\frac{1}{\varepsilon_j}\big)^{a^*}\Big),$$

for $j \geq 0$. Moreover, one proves that our solution passes through doubly-resonant domains, so the corresponding stability exponent is $a = 1/2(n-2)\alpha$. Therefore our result proves the optimality of that local exponent for the Gevrey classes of exponent $\alpha > 1$.

1.2 Main results of the paper

In this paper we are concerned with the same optimality problem in the analytic category, for which we have to introduce new ideas to construct examples.

Let d_∞ denote the product distance (supnorm) in $\mathbb{C}^n$. We adopt the following notation for complex domains: for $\rho > 0$, we write $V_\rho(\mathbb{T}^n)$ (or simply V_ρ) for the closed neighborhood of width ρ of the real torus $\mathbb{T}^n$ in $\mathbb{C}^n/\mathbb{Z}^n$, that is $V_\rho = \{z \in \mathbb{C}^n \mid d_\infty(z, \mathbb{T}^n) \leq \rho\}$, and for a domain D in $\mathbb{R}^n$, we set $W_\rho(D) = \{z \in \mathbb{C}^n/\zeta^n \mid d_\infty(z, D) \leq \rho\}$. We write $U_\rho(D) = V_\rho \times W_\rho(D)$. We endow the spaces of bounded analytic functions on these domains with its usual C^0 norm.

We first state our instability result in the framework of exact symplectic diffeomorphisms. Given a point $z \in \mathbb{A}^n$, we denote by $r_i(z)$ the component of rank i of its action variable r. Given a Hamiltonian H, we denote by Φ^H the corresponding time-one map (provided it exists).

Theorem A (Instability example in the discrete case)**.** Let n be an integer ≥ 3, and set

$$a_d^* = \frac{1}{2(n-2)}.$$

Let $h(r) = \frac{1}{2}(r_1^2 + \cdots + r_n^2)$. Then there exist $\rho > 0$ and a sequence $(\Psi_j)_{j \geq 0}$ of real-analytic exact symplectic diffeomorphisms of $\mathbb{A}^n$, with analytic continuation to $U_\rho = U_\rho(\mathbb{R}^n)$, verifying

$$\varepsilon_j := \left\| \Psi_j - \Phi^h \right\|_{C^0(U_\rho)} \to 0 \quad \text{when} \quad j \to \infty,$$

such that each Ψ_j admits a wandering point $z^{(j)}$. Moreover there exists a sequence $(\kappa_j)_{j \geq 0}$ of positive integers and a constant $C > 0$ satisfying

$$\kappa_j \leq \frac{C}{\varepsilon_j^2} \exp\Big(C\big(\frac{1}{\varepsilon_j}\big)^{a^*}\Big)$$

such that

$$r_2(\Psi_j^{\kappa_j}(z^{(j)})) - r_2(z^{(j)}) \geq 1$$

for $j \geq 0$. The constant C depends only on n and R.

The main part of this paper is devoted to the proof of Theorem A. We then easily deduce the following result from the analytic suspension technique of [21].

Theorem B (Instability example in the continuous case)**.** Let $n \geq 4$ and set

$$a_c^* = \frac{1}{2(n-3)}.$$

Let $R > 1$. Then there exist $\rho > 0$, a sequence of analytic functions $(f_j)_{j\geq 0}$ with analytic continuation to the domain $U_\rho = U_\rho(B_R)$, and an increasing sequence of integers $(\tau_j)_{j\geq 0}$ such that, for each $j \geq 0$, the Hamiltonian system generated by

$$\mathcal{H}_j(\theta, r) = \frac{1}{2}(r_1^2 + \cdots + r_{n-1}^2) + r_n + f_j(\theta, r)$$

admits a solution $(\theta(t), r(t))$ defined at least for $t \in [0, \tau_j]$ and for which $r_2(0) = 0$ and $r_2(\tau_j) = 1$. Moreover, there exists a positive constant C which depends only on n and R, such that the time of drift τ_j and the norm $\varepsilon_j = \|f_j\|_{C^0(U_\rho)}$ are related by

$$\tau_j \leq \frac{C}{\varepsilon_j^2} \exp\Big(C\big(\frac{1}{\varepsilon_j}\big)^{a_c^*}\Big), \qquad j \geq 0.$$

As in the Gevrey case the orbits we construct pass very close to double resonant surfaces, so the optimal value for the exponent a_c^* would be $1/2(n-2)$. We could not reach this value due to some technical difficulties in the construction of our analytic example, but this result is almost optimal, and becomes more and more so when the number of degrees of freedom tends to infinity, which was our original goal. Nevertheless we think that an improved construction could yield the correct exponent, as well as unstable orbits which are close to simple resonances. But these methods will contain technical refinements which can obscure the underlying ideas, so we see the present work as a first significant step in the direction of optimality in the analytic category, as well as a basis for further work.

1.3 Description of the method and general comments

The proof splits into two main parts. The first one (Sections 2 and 3) gathers the dynamical constructions: roughly speaking the whole method relies on an embedding of a two-dimensional normally hyperbolic annulus with homoclinic connections into a higher dimensional near-integrable system, which enables us to obtain the drifting orbits by means of a semi-local analysis combining the dynamics near the annulus with heteroclinic excursions. In order to perform an accurate enough analysis of the local dynamics in the neighborhood of the annulus it is necessary to conjugate our system to a direct product. This kind of result, close in spirit to Sternberg's conjugacy theorem, is in large part classical but we could not find in the literature a version that would suit our needs. The second part of the proof (Section 4) is devoted to the extension of Sternberg's theorem to our normally hyperbolic and symplectic framework.

For the convenience of the reader we describe our constructions a little bit more in the present section. We take the opportunity to point out the similarities between our method and that introduced by J. Bourgain and V. Kaloshin in [8, 9], which nevertheless yields qualitatively very different results. We conclude the section with a discussion of the respective scopes of the two approaches.

1.3.1 The dynamical constructions

The construction of our examples is reminiscent of that of [28], with which it presents some similarities: the major part of the work consists in producing discrete systems with wandering points (first theorem), and we then recover the continuous setting (second theorem) thanks to an analytic suspension process. Also, we "embed" here low dimensional diffeomorphisms with controlled dynamics − essentially the existence of wandering points with estimates on the speed of drift − into high-dimensional near integrable ones, and deduce the existence of instability in these systems from that of the wandering points in the low dimensional ones.

But here, due to analytic rigidity, we are led to vary the geometry of our diffeomorphisms. While in [28] our low dimensional systems were particular standard maps on the two-dimensional annulus $\mathbb{A}$, here we have to make use of suitable discrete systems defined on the annulus $\mathbb{A}^2$. The construction of these diffeomorphisms on $\mathbb{A}^2$ is indeed the main part of the present work.

The main point is that in [28] the "embeddings" of the standard map were made possible by the existence of compactly supported functions in the Gevrey category. Here, we can only obtain *approximate* embeddings and it is necessary to introduce perturbative techniques in order to keep control of the orbits of the wandering points.

This is precisely the reason why we first construct intermediate systems on the annulus $\mathbb{A}^2$, in the the same way as in [27], into which we are able to embed approximate standard maps defined on $\mathbb{A}$. These systems on $\mathbb{A}^2$ are analytic perturbations of the time-one map of the (hyperbolic) Hamiltonian

$$K(\theta, r) = \tfrac{1}{2}(r_1^2 + r_2^2) + \cos 2\pi\theta_1$$

on $\mathbb{A}^2$, *i.e.* the product of a pendulum and an oscillator. These perturbations still admit the annulus $\mathcal{A} = (0,0) \times \mathbb{A}$ (that is the product of the hyperbolic point of the pendulum with the second factor) as a normally hyperbolic invariant manifold. In [27] we proved that $\mathcal{A}$ admits a homoclinic two-dimensional annulus, and (even if we will not make use of such an elaborate construction) it is possible to prove the existence of a family of two-dimensional annuli which are invariant under the q^{th}-iterates of the system, for each q large enough. The perturbation is chosen in such a way that the induced dynamics on these invariant annuli uniformly approximate suitable standard maps with wandering points, which we see this way as approximately embedded in our system.

Indeed, it will not be necessary to perform such a refined dynamical analysis. Our construction may also be viewed as a discrete version of Arnold's example in which every quantity is (almost) explicitly computable. The invariant annulus $\mathcal{A}$ is foliated by invariant circles, and the effect of the perturbation is to create heteroclinic connections between their invariant manifolds. Our strategy will be to use Easton's windowing technique to detect and localize the drifting points located in the neighborhood of these heteroclinic intersections. One could check a posteriori that the drifting points coincide with those of the embedded standard map.

Let us describe more precisely our method. One main remark is that the wandering points have to stay most of the time in a small neighborhood of the invariant annulus $\mathcal{A}$. Therefore we can expect to control a large number of their iterates as soon as a precise knowledge of the dynamics near the hyperbolic manifold is possible. To this end we derived a new version of the Sternberg conjugacy theorem, adapted to the case of noncompact normally hyperbolic manifolds in *analytic* systems, based on Moser's deformation method as described in [2]. This way our system appears to be locally conjugate to the product of a neighborhood of the hyperbolic fixed point of the pendulum map with the harmonic oscillator. This brings us back to the case of compactly supported functions much as in the Gevrey category (see Lemma 2.5).

Next, in order to evaluate the drift along one action axis, we introduce as in [26] a method based on the shadowing lemma of Easton, which consists in constructing small boxes localized very near the heteroclinic points and enjoying suitable intersections properties under the effect of the diffeomorphism. Here the use of that method is facilitated by the almost product structure of our system and the final situation is very similar to that of [28]. Moreover, windowing is robust enough so that we can include the remainders originating from Sternberg's conjugacy (see Lemma 2.9), which enables us to "shadow the boxes" in the final system. Since these boxes may be chosen regularly spaced along one of the action coordinate axes, we finally easily obtain our drifting orbits and control their drifting time.

In conclusion, we wish to point out that our system is also very close to an anti-integrable limit, and can be seen as an example of the methods developed by D. Treschev in [35], which could probably apply in our context to simplify the windowing control. Another remark is that more general examples could certainly be obtained using the preparation method developed in [17]. We hope to get back to that question in a subsequent paper.

1.3.2 The conjugacy theorem

The conjugacy result we alluded to above deals with analytic diffeomorphisms, in a symplectic setting and along a normally hyperbolic non compact invariant submanifold. This prompted us to develop a tailor-made version of the theorem we are interested in and in so doing we were led to some observations which may be of independent interest. We hope to return to these points elsewhere, in more detail and in a more general setting.

Let us be more precise. Let f_0, f_1 be two symplectic diffeomorphisms of some symplectic manifold V, which preserve the submanifold $M \subset V$ and are normally hyperbolic along M. All these data, namely V, M, f_0, f_1 are assumed to be analytic. We wish to show that if f_0 and f_1 have a contact of large enough order along M, they are C^ℓ

conjugate in a neighborhood of M, for an integer $\ell \geq 1$ which we will compute.

Such results originate in [32] for (germs of) diffeomorphisms of the Euclidean space near the origin, which is assumed to be a hyperbolic fixed point. In short such germs are conjugate if and only if they are formally conjugate, a phenomenon which elaborates on considerations first made by Poincaré (see [31, 32]). For all the metamorphoses, the modern proofs are still quite close to the original one by S.Sternberg. In particular, in [10] such results are proved in a more modern and precise fashion, namely using –by now classical– fixed point theorems in Banach spaces. The symplectic setting is only briefly mentioned at the end of the book. In [13], [11, 12] (see additional references in these papers) it is shown how such conjugacy results can in principle be reduced to general invariant submanifold theorems although that reduction may not be concretely so easy or effective; it is advocated that this more abstract framework should make it possible to derive more general results.

Here we will follow the strategy developed in [2], which connects this circle of problems with two classical and well-established techniques, namely the deformation method and the various theories of "normal forms," the germs of which can be found (as usual) in Poincaré. We refer to the clear and thorough discussion in [2] for more detail. This in particular enables one to easily incorporate the various types of geometry in the discussion; in [2] four kinds of geometry are discussed, namely general (no invariant), symplectic, volume preserving and contact diffeomorphisms (see also [3] for this last type). One also easily incorporates the continuous setting, *i.e.* replaces diffeomorphisms with flows.

We insist that we start here from analytic data, with an invariant submanifold M which is not reduced to a point, as is the case in all papers we have mentioned so far. The output, namely the local conjugacy, is only finitely differentiable but the analyticity of the data will help to simplify the proof and it leads to interesting specific and perhaps surprising phenomena in terms of regularity properties along the invariant submanifold M (see §§4.3, 4.4).

1.3.3 Asymptotics and high dimensional diffusion

To begin with let us remark that our problem is to find *asymptotic* estimates when the size of the perturbation tends to 0. Our construction here may be viewed as lying between the method developed in [28] and the original Arnold mechanism. Indeed, as is proved in [27], the hyperbolic annulus described aboved admits a continuous foliation by invariant circles, such that two nearby circles possess heteroclinic intersections. It is therefore possible to extract a "transition chain" from that family, and usual results ([26]) prove the existence of drifting points along such a chain. Our main difficulty was to explicitely determine the time of drift from the data, which necessitates a very precise control of all the parameters and makes it necessary to use a Sternberg type conjugacy result as explained above.

The approach in [9] involves similar constructions, while the purpose is not the same: the authors produce examples of perturbations of a given completely integrable system which admit unstable orbits whose drifting time is *linear* with respect to the inverse of the perturbation. Due to the classical exponential normal form theory and stability estimates this cannot be an asymptotic mechanism. Instead, the perturbation is assumed to be *not too small*, typically $\varepsilon \geq \exp -d$, where d is the dimension of the phase space. It is then only when $d \to \infty$ that the system can be considered as a genuine perturbation of an integrable one.

While the dynamical constructions are similar, there are two main differences between these two approaches, actually aiming at different goals. The first one is that here we limit ourselves to a simple example of standard map in order to produce unstable orbits, whereas the use of Mather's theory in [9] makes it possible to use more general examples, and as a consequence to extend the validity of the method to broader classes of unperturbed systems. The second one is that thanks to the short time needed for the orbits to drift one can first construct the perturbations in the C^∞ category and then use smoothing results in order to restore analyticity. Due to the much longer drifting times involved in the present paper (which are again unavoidable because of the stability theorems in the perturbative framework) we cannot use this more direct path, and this is precisely the reason why we had to develop the conjugacy results presented and used in this paper. We believe that a slight modification of our systems would make it possible to exhibit analytic examples of high dimensional diffusion which belong to the class constructed in [9], but we have not pursued the matter further.

Finally we remark that we also could have chosen a non-convex unperturbed Hamiltonian h, of the form

$$h(r) = \tfrac{1}{2}(\ell_1 r_1^2 + \cdots + \ell_{n-1} r_{n-1}^2) + \ell_n r_n$$

with $(\ell_1, \cdots, \ell_n) \in \{-1, 1\}^n$. The necessary modifications are almost obvious, we refer to [27] for details. One should however beware of the fact that when the quadratic form is not definite (i.e. the ℓ_i's are not all equal), one can very easily construct a perturbation of h of size ε for which the action variables experience a drift with average speed ε along the isotropic planes of h (the stability theorems do not apply there).

Acknowledgments: It is a pleasure to thank H.S.Dumas for a careful reading of a first version of the manuscript.

2 Speed of drift for diffeomorphisms on $\mathbb{A}^2$

The family $(\mathcal{F}_q)_{q\in\mathbb{N}}$ of diffeomorphisms we consider in this section was introduced in [27], to which we refer for a detailed study. When $q \to \infty$ the maps $\mathcal{F}_q$ are analytic perturbations of the time-one map

$$\mathcal{F}_* = \Phi^{\frac{1}{2}(r_1^2+r_2^2)+\cos 2\pi\theta_1} = \Phi^{\frac{1}{2}r_1^2+\cos 2\pi\theta_1} \times \Phi^{\frac{1}{2}r_2^2} \tag{2.2}$$

so we call it an initially hyperbolic (or *a priori* unstable) family. Here we first briefly recall the main properties of the maps $\mathcal{F}_q$, namely the existence of a normally hyperbolic manifold that admits an invariant foliation by invariant circles, from which one can extract a transition chain (that is a discrete subfamily of heteroclinically connected circles with minimal rotation). We then make use of that chain to construct drifting points by a windowing method due to Easton [16]. For each q large enough, we construct a countable family of small balls $(\mathcal{B}^{(q,k)})_{k\in\mathbb{Z}}$ (the images of the windows) located very near the heteroclinic points, such that $(\mathcal{F}_q)^q(\mathcal{B}^{(q,k)})$ intersects $\mathcal{B}^{(q,k+1)}$ in a convenient way, which will be described below. By Easton's shadowing lemma, this yields the existence of a point $\zeta^{(q)}$ such that the iterate $(\mathcal{F}_q)^{kq}(\zeta^{(q)})$ belongs to $\mathcal{B}^{(q,k)}$ for each integer $k\in\mathbb{Z}$. The distance between two consecutive balls $B^{(q,k)}$ and $B^{(q,k+1)}$ is very close to $1/q$, and as a consequence the number of iterates needed to make the point $\zeta^{(q)}$ drift over an interval of length 1 is approximately q^2; this will enable us to estimate the time of instability as a function of the size of the perturbation in the next section.

2.1 The diffeomorphisms $\mathcal{F}_q$

This paragraph is devoted to a brief description of the form and geometric structure of the maps $\mathcal{F}_q$.

In the following we fix a positive real number σ, and we measure the C^0-norms of our various functions over the domain $U_\sigma(\mathbb{R}^2)$ (see the definition at the beginning of Section 1.2). The width σ will have to be chosen small enough below, in order to simplify some technical estimates.

2.1.1 We obtain the diffeomorphisms $\mathcal{F}_q$ by composing $\mathcal{F}_*$ with the time-one map of a small Hamiltonian function. For $q\geq 1$, we set

$$\mathcal{F}_q=\Phi^{\frac{1}{q}f^{(q)}}\circ\mathcal{F}_* \tag{2.3}$$

where the function $f^{(q)}$ depends only on the angles θ_1 and θ_2 and has the product form $f^{(q)}(\theta_1,\theta_2)=f_1^{(q)}(\theta_1)f_2(\theta_2)$, with

$$f_1^{(q)}(\theta_1)=(\sin\pi\theta_1)^{\nu(q;\sigma)},\qquad f_2(\theta_2)=-\tfrac{1}{\pi}\Big(2+\sin 2\pi\big(\theta_2+\tfrac{1}{6}\big)\Big). \tag{2.4}$$

The exponent $\nu(q;\sigma)$ in the function $f_1^{(q)}$ plays a crucial role in the construction. We set

$$\nu(q;\sigma)=2\Big[\frac{\operatorname{Log} q}{4\pi\sigma}+1\Big],\qquad q\geq q_\sigma, \tag{2.5}$$

where $[x]$ denotes the integer part of the real number x, and where q_σ is the smaller positive integer such that $\nu(q_\sigma;\sigma)=2$, so $\nu(q;\sigma)\geq 2$ for $q\geq q_\sigma$. Note that since $\nu(q;\sigma)$ is even $f^{(q)}$ is a well-defined 1-periodic function.

Remark that the function $f_1^{(q)}$ has a contact of order $\nu(q;\sigma)$ with 0 at the point $\theta_1=0$, and that the perturbative diffeomorphism $\Phi^{\frac{1}{q}f^{(q)}}$ admits the following explicit expression :

$$\Phi^{\frac{1}{q}f^{(q)}}\big((\theta_1,r_1),(\theta_2,r_2)\big)=\Big(\theta_1,\ r_1-\tfrac{1}{q}f_2(\theta_2)(f_1^{(q)})'(\theta_1),\ \theta_2,\ r_2-\tfrac{1}{q}f_1^{(q)}(\theta_1)f_2'(\theta_2)\Big) \tag{2.6}$$

from which one immediately deduces that the diffeomorphisms $\mathcal{F}_*$ and $\mathcal{F}_q$ have a contact of order $\nu(q;\sigma)-1\geq 1$ along the submanifold of equation $\theta_1=0$.

2.1.2 Let us briefly depict the main invariant hyperbolic objects of the maps $\mathcal{F}_q$. First consider the system $\mathcal{F}_*$, the hyperbolic properties of which come from those of the pendulum map Φ^P. We denote by $O=(0,0)$ the hyperbolic fixed point of Φ^P, and focus on the *upper* part of its homoclinic loop, that is the curve of equation $r_1=2\,|\sin\pi\theta_1|$. With a slight abuse of notation, we write $W^+(O,\Phi^P)=W^-(O,\Phi^P)$ for that upper separatrix.

For the product map $\mathcal{F}_*$, the annulus $\mathcal{A}=\{O\}\times\mathbb{A}$ is a normally hyperbolic invariant manifold, which is obviously symplectic for the canonical structure of $\mathbb{A}^2$, being identified with the one-dimensional annulus $\mathbb{A}$ by means of the coordinates (θ_2,r_2). In the following we are interested only in the part of its invariant manifolds corresponding to the upper separatrix of O, and we write

$$W^{\pm}(\mathcal{A},\mathcal{F}_*)=W^{\pm}(O,\Phi^P)\times\mathbb{A}$$

for these stable and unstable manifolds, which obviously coincide.

On the invariant annulus $\mathcal{A}$ itself, with the previous identification, the restriction of the map $\mathcal{F}_*$ is the integrable twist map $(\theta_2,r_2)\mapsto(\theta_2+r_2,r_2)$. The circles

$$\mathcal{C}_{r_2^0}=\{O\}\times(\mathbb{T}\times\{r_2^0\}),\quad r_2^0\in\mathbb{R},$$

are therefore invariant and partially hyperbolic for $\mathcal{F}_*$. As above, we consider only the part of their invariant manifolds corresponding to the upper separatrix of the pendulum map, and we set

$$W^{\pm}(\mathcal{C}_{r_2^0},\mathcal{F}_*)=W^{\pm}(O,\Phi^P)\times(\mathbb{T}\times\{r_2^0\}).$$

Again, they obviously coincide.

As for the perturbed diffeomorphisms $\mathcal{F}_q$, the contact of $\mathcal{F}_q$ with $\mathcal{F}_*$ along $\{\theta_1=0\}$ shows that the annulus $\mathcal{A}=\{O\}\times\mathbb{A}$ is still invariant and normally hyperbolic, and that the restriction of $\mathcal{F}_q$ to $\mathcal{A}$ coincides with that of $\mathcal{F}_*$. Therefore, the circles $\mathcal{C}_{r_2^0}$ are invariant and partially hyperbolic for $\mathcal{F}_q$. The stable and unstable manifolds of $\mathcal{A}$ for $\mathcal{F}_q$ are tangent along $\mathcal{A}$ to those obtained for $\mathcal{F}_*$; we denote by $W^{\pm}(\mathcal{A},\mathcal{F}_q)$ the parts of these manifolds which are tangent to the manifolds $W^{\pm}(\mathcal{A},\mathcal{F}_*)$ defined above, and we define the invariant manifolds $W^{\pm}(\mathcal{C}_{r_2^0},\mathcal{F}_q)$ in the same way.

2.1.3 We now add some comments on the functions $f_1^{(q)}$ and f_2. First note that the function $f_1^{(q)}$ satisfies the inequality $\left|f_1^{(q)}(\theta_1)\right|\leq(\pi\delta)^{\nu(q;\sigma)}$ for $\left|\theta_1\right|\leq\delta$. We will apply this estimate in small neighborhoods of 0. To be more precise, given $a\in]0,1[$ and a positive integer p, one easily checks that

$$a^{\nu(q;\sigma)}=o\,(1/q^p)\qquad\text{when}\qquad q\to\infty \tag{2.7}$$

provided that the width satisfies the inequality

$$\sigma<\sigma_p=\frac{|\mathrm{Log}\,a|}{4\,\pi\,p}. \tag{2.8}$$

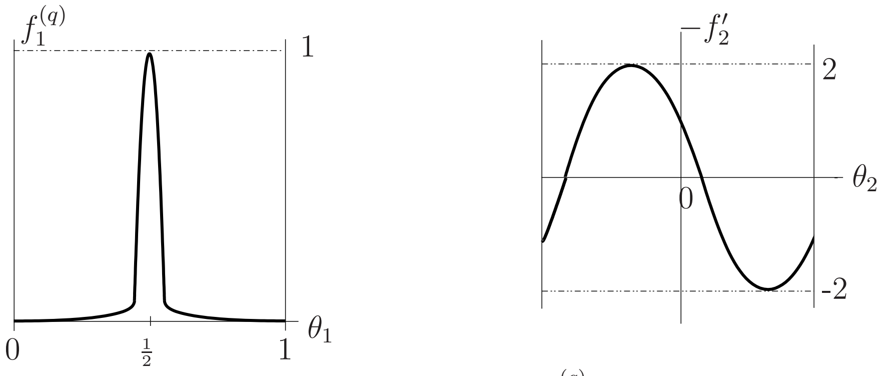

Fig. 1 Graphs of the functions $f_1^{(q)}$ and $-f_2'$

It is therefore possible to think of the term $a^{\nu(q;\sigma)}$ as "exponentially decreasing to 0" when $q \to \infty$, implicitly reducing the width σ as much as necessary to obtain the correct decreasing rate. This proves very useful in the following constructions.

The other important feature of $f_1^{(q)}$ is the constant value $f_1^{(q)}(\frac{1}{2}) = 1$, for all $q \geq q_\sigma$. Moreover, one sees that the function $f_1^{(q)}$ converges to the constant 1 uniformly when $q \to \infty$ on intervals of the form $\left|\theta_1 - \frac{1}{2}\right| < 1/\nu(q;\sigma)$. The derivatives of $f_1^{(q)}$ can also be uniformly estimated on such intervals.

Roughly speaking, the behaviour of the function $f_1^{(q)}$ near the origin enables us to control the local invariant manifolds in the perturbed system $\mathcal{F}_q$ and keep them very close to those of $\mathcal{F}_*$, while the behaviour near the point $\theta_1 = \frac{1}{2}$ is the main ingredient for creating a transverse intersection of $W^+(\mathcal{A}, \mathcal{F}_q)$ and $W^-(\mathcal{A}, \mathcal{F}_q)$ in the neighborhood of $\{\theta_1 = \frac{1}{2}\}$. We proved indeed in [27] that the intersection $W^+(\mathcal{A}, \mathcal{F}_q) \cap W^-(\mathcal{A}, \mathcal{F}_q)$ contains a two-dimensional annulus $\mathcal{I}_\mu$, which itself contains all the interesting homoclinic and heteroclinic objects of our system.

As for the function f_2, apart from the obvious inequality $|f_2| \geq \frac{1}{\pi}$ (in particular f_2 does not vanish on $\mathbb{T}$), we will be mainly interested in the properties of the derivative f_2', the zeroes of which correspond to homoclinic points, and of the second derivative f_2'', which provides us with lower estimates for the splitting in the θ_2-direction. The additional property $f_2'(0) = -1$ allows us to produce and localize heteroclinic points in a very simple way.

Finally, observe that $\mathcal{F}_q$ is indeed an analytic perturbation of $\mathcal{F}_*$, with the following inequality

$$\left\|\tfrac{1}{q} f^{(q)}\right\|_{C^0(V_\sigma)} \leq \|f_2\|_{C^0(V_\sigma)} \frac{1}{\sqrt{q}}, \tag{2.9}$$

which shows that $\mathcal{F}_q \to \mathcal{F}_*$ when $q \to \infty$ with respect to the C^0 analytic topology on $U_\sigma(\mathbb{R}^2)$.

2.1.4 For each integer $q \geq q_\sigma$ we will focus on the sequence of invariant circles $(\mathcal{C}_{k/q})_{k\in\mathbb{Z}}$. We proved in [27] that for all $k \in \mathbb{Z}$, there exists a heteroclinic point $\zeta^{(q,k)}$ which satisfies

$$\zeta^{(q,k)} \in W^-(\mathcal{C}_{k/q}, \mathcal{F}_q) \cap W^+(\mathcal{C}_{(k+1)/q}, \mathcal{F}_q)$$

and which is moreover located very near the point $\varpi^{(q,k)}$ defined by the coordinates

$$\varpi^{(q,k)} : (\theta_1 = \tfrac{1}{2}, r_1 = 2, \theta_2 = 0, r_2 = (k+1)/q).$$

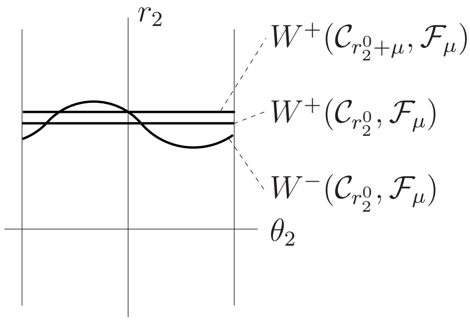

Fig. 2 Homoclinic and heteroclinic intersections in the annulus $\mathcal{I}_\mu$

To be more precise, there exists a constant $d \in]0, 1[$ such that $\left\|\zeta^{(q,k)} - \varpi^{(q,k)}\right\| \leq d^{\nu(q;\sigma)}$, for q large enough and $k \in \mathbb{Z}$.

Our drifting points $\zeta^{(q)}$ for $\mathcal{F}_q$ will be constructed in such a way that their orbits will pass successively extremely close to each of the heteroclinic points. The main result of this section is the following proposition.

Proposition 2.1. *There exists a width $\overline{\sigma}$, an integer $\overline{q}$ and a constant $\overline{d} \in]0, 1[$ such that for each integer $q \geq \overline{q}$ the diffeomorphism $\mathcal{F}_q$ admits a wandering point $\zeta^{(q)}$ which satisfies*

$$\left\|\mathcal{F}_q^{kq}(\zeta^{(q)}) - \varpi^{(q,k)}\right\| \leq \overline{d}^{\,\nu(q;\sigma)}, \qquad \forall k \in \mathbb{Z}. \tag{2.10}$$

The remainder of this section is devoted to the proof of Proposition 2.1, which will rely on several technical lemmas.

2.2 Windows

We now recall the definition and main properties of windows, following Easton [16]. Let M be a C^1 manifold of dimension $d \geq 2$, and let d_h, d_v be two positive integers such that $d_h + d_v = d$. A (d_h, d_v)-window with values in M is a C^1 diffeomorphism of $[-1, 1]^d$ into M. If $\mathcal{D}$ is such a window, its horizontals are the partial maps $\mathcal{D}(., y_v)$ for $y_v \in [-1, 1]^{d_v}$, and its verticals are the partial maps $\mathcal{D}(y_h, .)$ for $y_h \in [-1, 1]^{d_h}$. We denote by $\widetilde{\mathcal{C}}$ the image of the window $\mathcal{C}$.

Let $\mathcal{C}$ and $\mathcal{D}$ be two (d_h, d_v)-windows with values in M. One says that $\mathcal{C}$ is aligned with $\mathcal{D}$ when for each $y_h \in [-1, 1]^{d_h}$ and $y_v \in [-1, 1]^{d_v}$ the vertical $\mathcal{C}(y_h, .)$ and the horizontal $\mathcal{D}(., y_v)$ are transverse, and their images intersect at a unique point $a = \mathcal{C}(y_h, x_v) = \mathcal{D}(x_h, y_v)$ which satisfies $x_h \in]-1, 1[^{d_h}$ and $x_v \in]-1, 1[^{d_v}$.

Let us examine the simple example of affine windows, which will be of interest later. Consider the two (d_h, d_v)-windows with values in $\mathbb{R}^d$ defined by

$$\mathcal{C}(x) = c + C\,x \qquad \text{and} \qquad \mathcal{D}(x) = d + D\,x$$

where c, d are two points of $\mathbb{R}^d$ and C, D are two linear maps of $\mathbb{R}^d$, that we identify with their matrices in the canonical basis. These admit the following block decomposition:

$$C = \begin{bmatrix} C_1 & C_3 \\ C_2 & C_4 \end{bmatrix}, \qquad D = \begin{bmatrix} D_1 & D_3 \\ D_2 & D_4 \end{bmatrix}.$$

Define the intermediate matrices associated with the pair $(\mathcal{C}, \mathcal{D})$ by

$$M[C,D] := \begin{bmatrix} -D_1 & C_3 \\ -D_2 & C_4 \end{bmatrix}, \qquad N[C,D] := \begin{bmatrix} -C_1 & D_3 \\ -C_2 & D_4 \end{bmatrix}.$$

We denote by $\| \ \|_\infty$ the product norm in $\mathbb{R}^d$, and we equip the various spaces of linear maps with the induced norm. Then one easily checks that a necessary and sufficient condition for the window $\mathcal{C}$ to be aligned with the window $\mathcal{D}$ is that the matrix M is invertible, and that moreover the following inequality

$$\chi[\mathcal{C}, \mathcal{D}] := \operatorname{Sup}_{y \in [-1,1]^d} \big\| M^{-1}(d-c) + M^{-1}Ny \big\|_\infty < 1 \tag{2.11}$$

is satisfied. The previous intersection points then all satisfy $\|a\|_\infty \leq \chi[\mathcal{C}, \mathcal{D}]$. We call $\mu(\mathcal{C}, \mathcal{D}) = \big\| (M[C,D])^{-1} \big\|$ and $\chi[\mathcal{C}, \mathcal{D}]$ the alignment parameters of the pair $(\mathcal{C}, \mathcal{D})$ of affine windows.

For the sake of completeness we state the following easy lemma.

Lemma 2.2. *Let d_1 and d_2 be two integers ≥ 2 and consider the affine windows*

$$\mathcal{C}_i(x) = c_i + C_i x, \qquad \mathcal{D}_i = d_i = D_i x$$

with values in $\mathbb{R}^{d_i}$ for $i \in \{1,2\}$. Assume that $\mathcal{C}_i$ is aligned with $\mathcal{D}_i$, with parameters μ_i and χ_i. Then the product window $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$ is aligned with the product window $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$, with parameters

$$\mu[\mathcal{C}, \mathcal{D}] = \operatorname{Max}(\mu_1, \mu_2) \qquad \textit{and} \qquad \chi[\mathcal{C}, \mathcal{D}] = \operatorname{Max}(\chi_1, \chi_2).$$

The following shadowing lemma was proved by Easton in [16], it will be a main ingredient for the construction of our drifting points.

Lemma 2.3. *Let Φ be a C^1-diffeomorphism of a manifold M. Assume that there exists a sequence $(\mathcal{D}_k)_{k \in \mathbb{Z}}$ of (d_h, d_v)-windows with values in M, such that for each $k \in \mathbb{Z}$ the window $\mathcal{D}_k$ is aligned with the window $\mathcal{D}_{k+1}$. Then there exists a point z_0 such that $\Phi^k(z_0)$ is contained in the image $\widetilde{\mathcal{D}}_k$ of the window $\mathcal{D}_k$, for each $k \in \mathbb{Z}$.*

Observe that if the sequence $\mathcal{D}_k$ satisfies the assumptions of Easton's lemma for the diffeomorphism Φ, then it is also the case for a small enough C^1-perturbation of Φ. One has indeed to consider only the alignment problem for any two consecutive windows of the sequence, which can be done using the next lemma, introduced in [26]. When $\mathcal{C}$ is a C^1 function from an open set $V \subset \mathbb{R}^d$ to $\mathbb{R}^d$, we set

$$\left\|\mathcal{C}\right\|_{C^1(V)} = \operatorname*{Sup}_{x\in V} \|\mathcal{C}(x)\|_\infty + \operatorname*{Sup}_{x\in V} \|D\mathcal{C}(x)\|.$$

Lemma 2.4. *Let $\mathcal{C}$ and $\mathcal{D}$ be two (d_h, d_v)-windows with values in $\mathbb{R}^d$, of the following form*

$$\mathcal{C}(x) = c + C\,x + \widehat{\mathcal{C}}(x), \qquad \mathcal{D}(x) = d + D\,x + \widehat{\mathcal{D}}(x),$$

where c, d are two points of $\mathbb{R}^d$, C, D two linear maps of $\mathbb{R}^d$ and $\widehat{\mathcal{C}}$, $\widehat{\mathcal{D}}$ two maps of class C^2 from a neighborhood V of $[-1,1]^d$ to $\mathbb{R}^d$. Let $\mathcal{C}^a$ and $\mathcal{D}^a$ be the affine windows defined by $\mathcal{C}^a(x) = c + C\,x$ and $\mathcal{D}^a(x) = d + D\,x$.

Assume that the window $\mathcal{C}^a$ is aligned with $\mathcal{D}^a$, with alignment parameters μ and χ, and set $\chi' = \operatorname{Max}\left(\|\widehat{\mathcal{C}}\|_{C^1(V)}, \|\widehat{\mathcal{D}}\|_{C^1(V)}\right)$. Assume moreover that

$$4\mu\chi' < 1 \qquad \textit{and} \qquad \chi + \frac{4\mu\chi'}{1-4\mu\chi'} < 1. \tag{2.12}$$

Then the window $\mathcal{C}$ is aligned with the window $\mathcal{D}$.

Proof. We first prove that the verticals of $\mathcal{C}$ intersect the horizontals of $\mathcal{D}$. Given (y_h, y_v) in $[-1,1]^d$ we denote by (x_h^a, x_v^a) the unique point satisfying $\mathcal{C}^a(y_h, x_v^a) = \mathcal{D}^a(x_h^a, y_v)$, and we search for solutions (x_h, x_v) of the full system

$$\mathcal{C}(y_h, x_v) = \mathcal{D}(x_h, y_v)$$

of the form $(x_h, x_v) = (x_h^a, x_v^a) + (z_h, z_v)$. One easily checks that a necessary and sufficient condition for (x_h, x_v) to be a solution is that $z = (z_h, z_v)$ be a solution of the equation $z = F(z)$, where

$$F(z) = M^{-1}\Big(\widehat{\mathcal{D}}(x_h^a + z_h, y_v) - \widehat{\mathcal{C}}(y_h, x_v^a + z_v)\Big).$$

Moreover one checks that $\|F\|_{C^1(V)} \le 4\mu\chi'$, so F is a contracting map by condition (2.12), and F sends the ball $\overline{B}_\infty(0,r)$ into the ball $\overline{B}_\infty(0, 4\mu\chi'(1+r))$. So Banach fixed point theorem applies in the ball $\overline{B}_\infty(0,r)$ if one sets

$$r = \frac{4\mu\chi'}{1-4\mu\chi'}$$

and proves the existence of a unique solution of $z = F(z)$ in the ball of radius r. In order to ensure that the final point $(x_h^a, x_v^a) + (z_h, z_v)$ belongs to the open ball $]-1,1[^d$ one only needs to assume that $\chi + r < 1$, which is exactly the second part of condition (2.12).

As for transversality, for $x \in [-1,1]^d$ consider the matrix $\widehat{M}(x) = M[D\widehat{\mathcal{C}}(x), D\widehat{\mathcal{D}}(x)]$. We have to check that the matrix $M + \widehat{M}(x)$ is invertible for each solution x of the previous intersection equation, which is plain since $M + \widehat{M}(x) = M(\mathrm{Id} + M^{-1}\widehat{M}(x))$ with $\left\|M^{-1}\widehat{M}(x)\right\| \le 2\mu\chi' < 1$. $\square$

2.2.1 Proof of Proposition 2.1. It will be an easy corollary of a lemma on the existence of windows which we are now in a position to state.

Lemma 2.5. *There exists a width $\overline{\sigma}$ and an integer $\overline{q}$ such that for each $q \geq \overline{q}$ there exists a sequence $(\mathcal{D}^{(q,k)})_{k\in\mathbb{Z}}$ of $(2,2)$-windows with values in $\mathbb{A}^2$, such for each $k \in \mathbb{Z}$ the composed window $\mathcal{C}^{(q,k)} = \mathcal{F}_q^q \circ \mathcal{D}^{(q,k)}$ is aligned with $\mathcal{D}^{(q,k)}$. Moreover there exists $\overline{d} \in]0,1[$ such that for each $q \geq \overline{q}$ and $k \in \mathbb{Z}$ the image $\widetilde{\mathcal{D}}^{(q,k)}$ is contained in the ball centered at $\varpi^{(q,k)}$ of radius $\overline{d}^{\,\nu(q;\sigma)}$.*

Proposition 2.1 is an immediate consequence of Lemma 2.5 and Lemma 2.3. Indeed, the former applied to the sequence $(\mathcal{D}^{(q,k)})_{k\in\mathbb{Z}}$ and the diffeomorphism $\mathcal{F}_q^q$ yields the existence of a point $\zeta^{(q)}$ satisfying $\mathcal{F}_q^q(\zeta^{(q)}) \in \widetilde{\mathcal{D}}^{(q,k)}$ for each $k \in \mathbb{Z}$, which by the last assertion of Lemma 2.5 in turn implies that $\mathcal{F}_q^q(\zeta^{(q)})$ is in the ball centered at ϖ_k of radius $\overline{d}^{\,\nu(q;\sigma)}$. □

The rest of the section is devoted to the proof of Lemma 2.5. We will first introduce a sequence $(\overline{\mathcal{F}}_q)$ of approximations of $\mathcal{F}_q$, for which one easily constructs windows and check their alignment, and we will then deduce the lemma from the closeness of $\overline{\mathcal{F}}_q$ and $\mathcal{F}_q$.

2.3 The approximate maps $\overline{\mathcal{F}}_q$

To introduce the maps $\overline{\mathcal{F}}_q$ we first take advantage of the form of $f_1^{(q)}$ and define a new function $\overline{f}_1^{(q)} : \mathbb{T} \to \mathbb{R}$, which satisfies

$$\overline{f}_1^{(q)}(\theta_1) = f_1^{(q)}(\theta_1) \quad \text{for} \quad \left|\theta_1 - \tfrac{1}{2}\right| \leq \tfrac{1}{8}, \qquad \overline{f}_1^{(q)}(\theta_1) = 0 \quad \text{for} \quad |\theta_1| \leq \tfrac{1}{8}, \tag{2.13}$$

and which is continued in the complement so as to be of class C^∞ on $\mathbb{T}$. To fix ideas one can even assume that $0 \leq \overline{f}_1^{(q)} \leq f_1^{(q)}$, although it is not necessary. We then set

$$\overline{\mathcal{F}}_q = \Phi^{\frac{1}{q}\overline{f}^{(q)}} \circ \mathcal{F}_* \qquad \text{with} \qquad \overline{f}^{(q)} = \overline{f}_1^{(q)} \otimes f_2. \tag{2.14}$$

We will show in the rest of this section that it is possible to define domains in which the iterate $\overline{\mathcal{F}}_q^q$ is explicitly determined, along with estimates of the C^1-norm of the difference $\mathcal{F}_q^q - \overline{\mathcal{F}}_q^q$.

2.3.1 We first need to introduce suitable flow-box coordinates for the pendulum map. We write $P(\theta_1, r_1) = \frac{1}{2}r_1^2 + \cos 2\pi\theta_1$ and $\mathcal{P} = \Phi^P$. We will work in the open domain $\mathcal{E}$ located above the upper separatrix defined by

$$\mathcal{E} = \{(\theta_1, r_1) \in \mathbb{A} \mid 0 < \theta_1 < 1,\ r_1 > 2\,|\sin \pi\theta_1|\}.$$

Using $\{\theta_1 = \frac{1}{2}\}$ as a reference section, we define the time-energy coordinates (τ, h) of a point $(\theta_1, r_1) \in \mathcal{E}$ as

$$h(\theta_1, r_1) = \tfrac{1}{2}r_1^2 + (\cos 2\pi\theta_1 - 1), \qquad \tau(\theta_1, r_1) = \int_{\frac{1}{2}}^{\theta_1} \frac{d\theta}{\sqrt{2(h(\theta_1, r_1) - V(\theta)}},$$

with $V(\theta) = \cos 2\pi\theta - 1$. It is well-known that these coordinates are symplectic. The period of motion is given as a (decreasing) function of energy by the formula

$$T(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{d\theta}{\sqrt{2(h - V(\theta))}}, \tag{2.15}$$

and the range of the coordinate change is the domain

$$\mathcal{E}_* = \{(\tau, h) \in \mathbb{R}^2 \mid h > 0,\ |\tau| < \tfrac{1}{2}T(h)\}.$$

In the coordinates (τ, h) the flow of P is straightened out, *i.e.* $\Phi^{tP} : \ (\tau, h) \mapsto (\tau + t, h)$ for $(\tau, h) \in \mathcal{E}_*$ and $|t|$ small enough. We write $H = T^{-1}$ for the inverse function of T. Given an integer $q \geq 1$, we define the strip

$$\mathfrak{S}(q) = \{(\tau, h) \in \mathcal{E}_* \mid |\tau| < 1,\ H(q + \tfrac{1}{2}) < h < H(q - \tfrac{1}{2})\}.$$

Assume now $q \geq 4$. Then the mapping $\mathcal{P}^q$ is well-defined in $\mathfrak{S}(q)$ with values in $\mathcal{E}_*$, with the following explicit expression

$$\mathcal{P}^q : \ (\tau, h) \mapsto \Big(\tau + q - T(h),\ h\Big). \tag{2.16}$$

(as usual we do no introduce a new notation for the diffeomorphisms expressed in new coordinates).

2.3.2 We can now define the domains we were looking for. For each integer $q \geq 3$ we will first be interested in a neighborhood $N^{(q)}$ of the point $a^{(q)}$ with (τ, h) coordinates $(0, H(q))$, that is the center of the strip $\mathfrak{S}(q)$ (in the initial coordinates, the point $a^{(q)}$ is the intersection of $\{\theta_1 = \frac{1}{2}\}$ with the unique orbit of the pendulum which has period q and is located above the upper separatrix).

We want to define the neighborhood $N^{(q)} \subset \mathfrak{S}(q)$ so as to satisfy the conditions

$$\mathcal{P}^k(N^{(q)}) \subset \{|\theta_1| < \tfrac{1}{8}\}, \quad \forall k \in \{1, \ldots, q-1\}. \tag{2.17}$$

Let b be the point of the upper separatrix with $\theta_1 = \frac{1}{8}$ in the initial coordinates, therefore the h-coordinate of b is zero, and one easily checks that $\tau(b) < 1$. Let $\varrho = 1 - \tau(b)$, and set

$$N^{(q)} = \{(\tau, h) \in \mathfrak{S}(q) \mid |\tau| < \varrho/4,\ H(q + \varrho/4) < h < H(q - \varrho/4)\}.$$

It is not difficult to see that $N^{(q)}$ satisfies our requirements (see Figure 3).

Turning back to the diffeomorphism $\overline{\mathcal{F}}_q$, we introduce the domain $\mathcal{N}^{(q)} = N^{(q)} \times \mathbb{A}$, in which the q^{th}-iterate $\overline{\mathcal{F}}_q^q$ has the following simple expression

$$\overline{\mathcal{F}}_q^q = \Phi^{\frac{1}{q}\overline{f}^{(q)}} \circ \mathcal{F}_*^q = \Phi^{\frac{1}{q}f^{(q)}} \circ \mathcal{F}_*^q. \tag{2.18}$$

To see this observe that the function $\overline{f}_1^{(q)}$ vanishes on the strip $\{|\theta_1| < \frac{1}{8}\}$ and so the diffeomorphism $\Phi^{\frac{1}{q}\overline{f}^{(q)}}$ reduces to the identity on that domain. Therefore the conditions

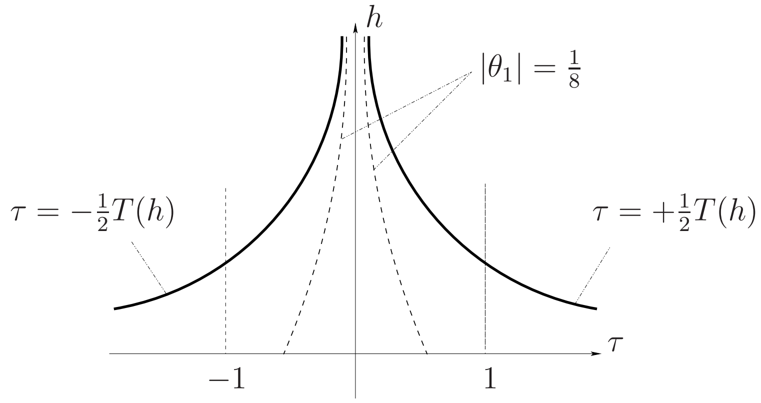

Fig. 3 The domain $\mathcal{E}_*$ and the limit curves $|\theta_1| = \frac{1}{8}$.

(2.17) immediately yield the first equality, thanks to the product form of $\mathcal{F}_*$. The second equality comes from the form of $\overline{f}_1^{(q)}$ and the choice of the neighborhood $N^{(q)}$

The simple expression (2.18) now enables us to make use of Sternberg's estimates of Section 4. This is the only (but crucial) place where we use this conjugacy result, which brings us back (up to a controlled remainder) to the approximate system $\overline{\mathcal{F}}_q$, which is dynamically easier to handle. More precisely we use Theorem E, which yields the following lemma.

Lemma 2.6. *There exist* $q_0 \in \mathbb{N}$ *and a constant* $\delta_0 \in]0, 1[$ *such that the inequality*

$$\left\| \mathcal{F}_q^q - \overline{\mathcal{F}}_q^q \right\|_{C^2(\mathcal{N}^{(q)})} \leq \delta_0^{\nu(q;\sigma)}$$

holds for $q \geq q_0$.

Proof. By Theorem E there exists an integer q_0, a constant $c \in]0, 1[$, and for each $q \geq q_0$ two diffeomorphisms χ_q and ψ_q of class C^2 satisfying

$$\left\| \chi_q - \mathrm{Id} \right\|_{C^2(\mathcal{N}^{(q)})} \leq c^{\nu}, \qquad \left\| \psi_q - \mathrm{Id} \right\|_{C^2(\mathcal{N}^{(q)})} \leq c^{\nu},$$

such that the intertwining relation

$$\mathcal{F}_* \circ \mathcal{F}_q^{q-1} \circ \psi_q = \chi_q \circ \mathcal{F}_*^q$$

holds true over the domain $\mathcal{N}^{(q)}$. Now over the same domain

$$\mathcal{F}_q^q = \Phi^{\frac{1}{q} f^{(q)}} \circ (\mathcal{F}_* \circ \mathcal{F}_q^{q-1}) = \Phi^{\frac{1}{q} f^{(q)}} \circ (\chi_q \circ \mathcal{F}_*^q \circ \psi_q^{-1}) = [\Phi^{\frac{1}{q} f^{(q)}} \circ \chi_q \circ (\Phi^{\frac{1}{q} f^{(q)}})^{-1}] \circ \overline{\mathcal{F}}_q^q \circ \psi_q^{-1}$$

from which one easily deduces the desired estimate, with a constant δ_0 slightly larger than c, increasing q_0 if necessary. □

2.4 Construction of the windows and proof of Lemma 2.5

We now introduce the point

$$\omega^{(q,k)} = a^{(q)} \times b^{(q,k)} \in \mathbb{A}^2$$

with $b^{(q,k)} = (0, k/q) \in \mathbb{A}$. We will go one step further in the simplification and replace the approximate map $\overline{\mathcal{F}}_q^q$ by its first order jet at the point $\omega^{(q,k)}$, which will allow us to easily construct the windows.

2.4.1 The first order jet $J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q)$ and affine windows.

Using the explicit expression (2.6) of the perturbative diffeomorphism $\Phi^{\frac{1}{q}f^{(q)}}$ one immediately checks that its first order jet at the point $\omega^{(q,k)}$ has the following product form

$$J^1_{\omega^{(q,k)}}(\Phi^{\frac{1}{q}f^{(q)}}) = J^1_{a^{(q)}}(\Phi^{\frac{1}{q}f_2(0)f_1^{(q)}}) \times J^1_{b^{(q,k)}}(\Phi^{\frac{1}{q}f_2^{(q)}})$$

Using now Equation (2.18), one sees that the first order jet of $\overline{\mathcal{F}}_q^q$ at $\omega^{(q,k)}$ has in turn the product form

$$J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q) = \Big(J^1_{a^{(q)}}\big(\Phi^{\frac{1}{q}f_2(0)f_1^{(q)}} \circ \mathcal{P}^q\big)\Big) \times \Big(J^1_{b^{(q,k)}}\big(S_q\big)\Big)$$

where

$$S_q = \Phi^{\frac{1}{q}f_2} \circ (\Phi^{\frac{1}{2}r_2^2})^q$$

has the usual form of a standard map.

Taking advantage of the product form of $J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q)$ we will first construct $(1,1)$-windows for the two factors of the annulus $\mathbb{A}$, and then make use of Lemma 2.2 to get the windows we need on $\mathbb{A}^2$.

1. Aligned affine windows for the standard map. Here we write $S_* = \Phi^{f_2} \circ \Phi^{\frac{1}{2}r_2^2}$ for the normalized standard map. The rescaled map S_q is related to the normalized one by means of the conjucacy relation $S_q = \sigma_q^{-1} \circ S_* \circ \sigma_q$, where $\sigma_q(\theta_2, r_2) = (\theta_2, q\, r_2)$, which is also clearly valid at the linearized level

$$J^1_{b^{(q,k)}}(S_q) = \sigma_q^{-1} \circ J^1_{b^{(k)}}(S_*) \circ \sigma_q \tag{2.19}$$

with $b^{(k)} = \sigma_q(b^{(q,k)}) = (0, k)$.

We first construct a sequence of $(1,1)$-windows adapted to the first order jets of S_*. Note that $S_*(0,k) = (0, k+1)$ for $k \in \mathbb{Z}$. For each $k \in \mathbb{Z}$ we define an affine $(1,1)$-window $\mathcal{D}_2^{(k)} : [-1,1]^2 \to \mathbb{A}$, satisfying $\mathcal{D}_2^{(k)}(0,0) = (0,k)$, such that the composed window $(J^1_{b^{(k)}}(S_*)) \circ \mathcal{D}_2^{(k)}$ is aligned with the window $\mathcal{D}_2^{(k+1)}$ for each $k \in \mathbb{Z}$.

One gets by easy computation

$$D_{b^{(k)}} S_* = \begin{pmatrix} 1 & 1 \\ -2\pi\sqrt{3} & 1 - 2\pi\sqrt{3} \end{pmatrix}.$$

The matrix $D_{b^{(k)}} S_*(0,k)$ is hyperbolic, with eigenvalues $\lambda_\pm = (1 - \pi\sqrt{3}) \pm \sqrt{\pi(3\pi - 2\sqrt{3})}$ (*i.e.* approximately $\lambda_+ \approx -8.78$ and $\lambda_- \approx -0.11$), and associated eigenvectors

$$u_h = (1, -(1 - \lambda_-)) \approx (1,\, 1.11), \qquad u_v = (1, -(1 - \lambda_+)) \approx (1,\, 9.78).$$

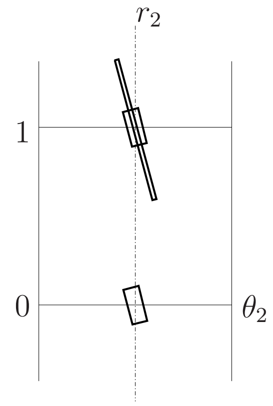

Fig. 4 Windows for the standard map S_*

Given a positive ϱ, we define for each $k \in \mathbb{Z}$ the affine window

$$\mathcal{D}_2^{(k)}(x) = b^{(k)} + \varrho\, D_2 x := (0, k) + \varrho(x^{(h)} u_h + x^{(v)} u_v), \qquad x = (x^{(h)}, x^{(v)}) \in [-1, 1]^2.$$

We then consider the affine window $\mathcal{C}_2^{(k)} := J^1_{b^{(k)}} S_* \circ \mathcal{D}_2^{(k)}$ and write

$$\mathcal{C}_2^{(k)}(x) = b^{(k+1)} + \varrho\, C_2 x$$

(note that the linear part is independent of k). Thanks to the choice of the horizontal and vertical directions one immediately sees that the window $\mathcal{C}_2^{(k)}$ is aligned with $\mathcal{D}_2^{(k+1)}$ for each $k \in \mathbb{Z}$. Moreover, the parameter $\mu_2(\mathcal{C}_2^{(k)}, \mathcal{D}_2^{(k+1)}) = \varrho^{-1} \big\| (M[C_2, D_2])^{-1} \big\|$ is independent of k and one checks that the parameter $\chi(\mathcal{C}_2^{(k)}, \mathcal{D}_2^{(k+1)}) := \chi_2$ is independent of k and ϱ.

Now let $q \geq 1$ be fixed. Coming back to the rescaled map S_q, we set $u_h^{(q)} = \sigma_q^{-1}(u_h)$ and $u_v^{(h)} = \sigma_q^{-1}(u_v)$. We fix arbitrarily a constant d_0 in the interval $]0, 1[$ and for each $k \in \mathbb{Z}$ we define the window

$$\mathcal{D}_2^{(q,k)}(x) = (0, \tfrac{k}{q}) + d_0^{\nu}(x^{(h)} u_h^{(q)} + x^{(v)} u_v^{(q)}) = b^{(q,k)} + d_0^{\nu}(\sigma_q^{-1}\, D_2)\, x, \tag{2.20}$$

for $x = (x^{(h)}, x^{(v)}) \in [-1, 1]^2$. We will prove the following lemma.

Lemma 2.7. *For each $k \in \mathbb{Z}$ the window $\mathcal{C}_2^{(q,k)} = J^1_{b(q,k)}(S_q) \circ \mathcal{D}_2^{(q,k)}$ is aligned with $\mathcal{D}_2^{(q,k+1)}$, with parameters*

$$\mu(\mathcal{C}_2^{(q,k)}, \mathcal{D}_2^{(q,k)}) \leq \overline{\mu}\, d_0^{-\nu}\, q \qquad \textit{and} \qquad \chi(\mathcal{C}_2^{(q,k)}, \mathcal{D}_2^{(q,k)}) = \chi_2$$

with $\overline{\mu} = \big\| (M[C_2, D_2])^{-1} \big\|$ and $\chi_2 = \chi(\mathcal{C}_2, \mathcal{D}_2)$.

Proof. For the window $\mathcal{C}_2^{(q,k)}$ we write

$$\mathcal{C}_2^{(k)}(x_2) = b^{(q,k+1)} + d_0^{\nu}(\sigma_q^{-1} \cdot C_2)\, x$$

so

$$M_2^{(q)} := M[\mathcal{C}_2^{(q,k)}, \mathcal{D}_2^{(q,k)}] = d_0^{\nu} M[(\sigma_q^{-1} \cdot C_2), (\sigma_q^{-1} \cdot D_2)] = d_0^{\nu} \sigma_q^{-1} \cdot M[C_2, D_2]$$

which is invertible. Moreover one gets for the first parameter

$$\mu_2(\mathcal{C}_2^{(q,k)}, \mathcal{D}_2^{(q,k)}) = \left\|(M[C_2, D_2])^{-1}\right\| d_0^{-\nu}\, q$$

The second parameter is easily seen to be constant and equal to χ_2, which shows the alignment and concludes the proof. □

2. Aligned affine windows for the perturbed pendulum. All the maps we consider here will be expressed in the (τ, h) coordinates. For $q \geq 1$ we set

$$h^{(q)} = H(q) = T^{-1}(q), \qquad T'_q = T'(h^{(q)}).$$

Note that $T'_q < 0$. For $x = (x^{(h)}, x^{(v)}) \in [-1,1]^2$ we define a first affine window

$$\mathcal{D}_1^{(q)}(x) := \Big(x^{(h)}\, d_0^{\nu},\ h^{(q)} + (x^{(h)} - x^{(v)})\frac{d_0^{\nu}}{T'_q}\Big) = a^{(q)} + D_1^{(q)} x$$

with

$$D_1^{(q)} = d_0^{\nu} \begin{pmatrix} 1 & 0 \\ -\dfrac{1}{T'_q} & \dfrac{1}{T'_q} \end{pmatrix}, \tag{2.21}$$

where d_0 is the constant introduced in (2.20) and $\nu = \nu(q; \sigma)$. In the (τ, h) coordinates the image of $\mathcal{D}_1^{(q)}$ is the convex hull of the four points

$$\begin{aligned} A_1 &= \Big(-d_0^{\nu},\, h^{(q)} - 2d_0^{\nu}/T'(h^{(q)})\Big), & A_2 &= \Big(d_0^{\nu},\, h^{(q)}\Big), \\ A_3 &= \Big(d_0^{\nu},\, h^{(q)} + 2d_0^{\nu}/T'(h^{(q)})\Big), & A_4 &= \Big(-d_0^{\nu}, h^{(q)}\Big). \end{aligned}$$

(see figure 5). Observe that the domain $\widetilde{\mathcal{D}}_1^{(q)}$ is extremely thin and nearly "horizontal". Indeed one has the well-known estimates

$$T(h) \sim_0 -\tfrac{1}{2\pi} \operatorname{Ln} h, \qquad h^{(q)} = H(q) \sim_\infty e^{-2\pi q}, \tag{2.22}$$

from which one easily deduces

$$T'(h^{(q)}) \sim_\infty -\tfrac{1}{2\pi}\, e^{2\pi q}, \qquad T''(h^{(q)}) \sim_\infty 2\pi (T'(h^{(q)}))^2. \tag{2.23}$$

So the images of the horizontals of $\mathcal{D}_1^{(q)}$ are line segments with slope very close to $-\frac{1}{2\pi}e^{-2\pi q}$, and the thickness of the image $\widetilde{\mathcal{D}}_1^{(q)}$ is about $\frac{d_0^{\nu}}{\pi} e^{-2\pi q}$.

The previous window has been chosen in order to facilitate the geometric description of the effect of the various diffeomorphisms at the linearized level. The following lemma decribes the intersection properties we need.

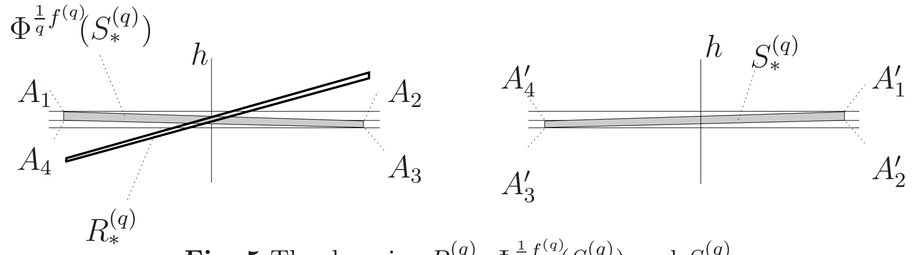

Fig. 5 The domains $R_*^{(q)}$, $\Phi^{\frac{1}{q}f^{(q)}}(S_*^{(q)})$ and $S_*^{(q)}$

Lemma 2.8. *Set $\kappa = f_2(0) = -5/2\pi$ and define the composed affine window*

$$\mathcal{C}_1^{(q)}(x) = a^{(q)} + C_1^{(q)}x := (J^1_{a^{(q)}}(\Phi^{\frac{\kappa}{q}f_1^{(q)}} \circ \mathcal{P}^q)) \circ \mathcal{D}_1^{(q)}(x), \qquad x \in [-1,1]^2.$$

Then the window $\mathcal{C}_1^{(q)}$ is aligned with $\mathcal{D}_1^{(q)}$ and the parameters $\mu_1^{(q)} = \left\| M([\mathcal{C}_1, \mathcal{D}_1])^{-1} \right\|$ and $\chi_1^{(q)} = \chi(\mathcal{C}_1, \mathcal{D}_1)$ satisfy the inequalities

$$\mu_1^{(q)} \leq 2\frac{qd_0^{-\nu}}{\bar{\kappa}\nu}, \qquad \chi_1^{(q)} \leq \frac{4q}{\bar{\kappa}\nu T'_q}. \tag{2.24}$$

Proof. First notice that $J^1_{a^{(q)}}(\Phi^{\frac{\kappa}{q}f_1^{(q)}} \circ \mathcal{P}^q) = J^1_{a^{(q)}}(\Phi^{\frac{\kappa}{q}f_1^{(q)}}) \circ (J^1_{a^{(q)}}(\mathcal{P}^q))$ with

$$J^1_{a^{(q)}}(\mathcal{P}^q)(\tau, h^{(q)} + h) = \Big(\tau - T'_q h,\ h^{(q)} + h\Big).$$

The composed window $(J^1_{a^{(q)}}(\mathcal{P}^q)) \circ \mathcal{D}_1^{(q)}$ has the explicit expression

$$(J^1_{a^{(q)}}(\mathcal{P}^q)) \circ \mathcal{D}_1^{(q)}(x) = a^{(q)} + Q_1^{(q)}x,$$

with

$$Q_1^{(q)} = d_0^{\nu} \begin{pmatrix} 0 & 1 \\ \frac{1}{T'_q} & -\frac{1}{T'_q} \end{pmatrix}$$

for $x = (x^{(h)}, x^{(v)}) \in [-1,1]^2$. Its image is the convex hull of the following four points (See figure 5).

$$A'_1 = \Big(d_0^{\nu},\ h^{(q)} - 2d_0^{\nu}/T'_q\Big), \qquad A'_2 = \Big(d_0^{\nu},\ h^{(q)}\Big),$$
$$A'_3 = \Big(-d_0^{\nu},\ h^{(q)} + 2d_0^{\nu}/T'_q\Big), \qquad A'_4 = \Big(-d_0^{\nu},\ h^{(q)}\Big).$$

Now let us examine the effect of the linearized perturbative map $J^1_{a^{(q)}}(\Phi^{\frac{\kappa}{q}f_1^{(q)}})$. In the initial (θ_1, r_1) coordinates,

$$\Phi^{\frac{\kappa}{q}f_1^{(q)}}(\theta_1, r_1) = \Big(\theta_1,\ r_1 - \tfrac{\kappa}{q}(f_1^{(q)})'(\theta_1)\Big)$$

so $\Phi^{\frac{\kappa}{q}f_1^{(q)}}(a^{(q)}) = a^{(q)}$ and

$$D_{a^{(q)}}\Phi^{\frac{\kappa}{q}f_1^{(q)}} = \begin{pmatrix} 1 & 0 \\ \kappa\pi^2\frac{\nu}{q} & 1 \end{pmatrix}.$$

Let φ be the coordinate change $\varphi(\tau,h)=(\theta_1,r_1)$, for which an easy computation yields

$$D_{a^{(q)}}\varphi=\begin{pmatrix}\sqrt{2(h^{(q)}+2)} & 0\\ 0 & \left(\sqrt{2(h^{(q)}+2)}\right)^{-1}\end{pmatrix}.$$

Therefore in the (τ,h) coordinates

$$D_{a^{(q)}}\Phi^{\frac{\kappa}{q}f_1^{(q)}}=\begin{pmatrix}1 & 0\\ \dfrac{2\kappa\pi^2\nu(h^{(q)}+2)}{q} & 1\end{pmatrix}.$$

For $x\in[-1,1]^2$, we finally get $(J^1_{a^{(q)}}(\Phi^{\frac{\kappa}{q}f_1^{(q)}}\circ\mathcal{P}^{(q)}))\circ\mathcal{D}_1^{(q)}(x)=a^{(q)}+C_1^{(q)}x$ with

$$C_1^{(q)}=d_0^\nu\begin{pmatrix}0 & 1\\ \dfrac{\bar\kappa\nu}{qT'_q} & \dfrac{\bar\kappa\nu}{q}-\dfrac{1}{T'_q}\end{pmatrix}, \tag{2.25}$$

with $\bar\kappa=2\kappa\pi^2(h^{(q)}+2)$. We therefore obtain the intermediate matrix

$$M_1^{(q)}=M[\mathcal{C}_1^{(q)},\mathcal{D}_1^{(q)}]=d_0^\nu\begin{pmatrix}-1 & 1\\ \dfrac{1}{T'_q} & \dfrac{\bar\kappa\nu}{q}-\dfrac{1}{T'_q}\end{pmatrix},\quad (M_1^{(q)})^{-1}=\frac{qd_0^{-\nu}}{\bar\kappa\nu}\begin{pmatrix}\dfrac{\bar\kappa\nu}{q}-\dfrac{1}{T'_q} & -1\\ -\dfrac{1}{T'_q} & -1\end{pmatrix},$$

from which one immediately gets the first part of (2.24). As for the parameter $\chi_1^{(q)}$, we first obtain

$$N_1^{(q)}=d_0^\nu\begin{pmatrix}0 & 0\\ -\dfrac{\bar\kappa\nu}{qT'_q} & \dfrac{1}{T'_q}\end{pmatrix},$$

therefore

$$\chi_1^{(q)}\leq\left\|(M_1^{(q)})^{-1}\right\|\left\|N_1^{(q)}\right\|\leq\frac{4q}{\bar\kappa\nu T'_q}$$

which concludes the proof. □

3. Aligned affine windows for $J^1(\overline{\mathcal{F}}_q^q)$. Using Lemma 2.2, Lemma 2.8 and Lemma 2.7 we have so far proved the following result.

Lemma 2.9. *Denote by $\mathcal{D}^{(q,k)}$ the product $(2,2)$-window with values in $\mathbb{A}^2$ defined by*

$$(x^{(h)},x^{(v)})\mapsto\left(\mathcal{D}_1^{(q)}(x_1^{(h)},x_1^{(v)}),\mathcal{D}_2^{(q,k)}(x_2^{(h)},x_2^{(v)})\right)$$

for $x^{(h)} = (x_1^{(h)}, x_2^{(h)})$ *and* $x^{(v)} = (x_1^{(v)}, x_2^{(v)})$. *Then the window*

$$\mathcal{C}^{(q,k)} = J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q) \circ \mathcal{D}^{(q,k)} = J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)})$$

is aligned with the window $\mathcal{D}^{(q,k+1)}$, *with parameters*

$$\mu(\mathcal{C}^{(q,k)}, \mathcal{D}^{(q,k)}) \leq \overline{\mu}\, q\, d_0^{-\nu(q;\sigma)}, \qquad \chi(\mathcal{C}^{(q,k)}, \mathcal{D}^{(q,k)}) \leq \chi_2 \tag{2.26}$$

where $\overline{\mu}$ *and* χ_2 *were defined in Lemma 2.7.*

2.4.2 Remainders and proof of Lemma 2.5

The next and last lemma provides us with the necessary estimates on the remainders.

Lemma 2.10. *For* $q \in \mathbb{N}$ *and* $k \in \mathbb{Z}$ *we set*

$$\widehat{\mathcal{C}}^{(q,k)} = \mathcal{F}_q^q \circ \mathcal{D}^{(q,k)} - J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)}).$$

There exists $q_1 \in \mathbb{N}$, *a constant* $d_1 \in]0,1[$ *and a neighborhood* V *of* $[-1,1]^2$ *in* $\mathbb{R}^2$ *such that the following inequality*

$$\big\|\widehat{\mathcal{C}}^{(q,k)}\big\|_{C^1(V)} \leq C\, d_1^{2\nu(q;\sigma)}$$

holds true for all $k \in \mathbb{Z}$.

Proof. It is completely elementary and relies on the mean value theorem applied to the second derivative of $\widehat{\mathcal{C}}^{(q,k)}$, for which we will obtain upper bounds using the following formula

$$D^2 g \circ f = (D^2 g \circ Df) \cdot Df^{\otimes 2} + (Dg \circ Df) \cdot D^2 f$$

for the composition of differentiable maps. We first write

$$\widehat{\mathcal{C}}^{(q,k)} = (\mathcal{F}_q^q - \overline{\mathcal{F}}_q^q) \circ \mathcal{D}^{(q,k)} + \Big(\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)} - J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)})\Big).$$

Since the window $\mathcal{D}^{(q,k)}$ is linear with bounded C^1-norm, one immediately gets the following inequality

$$\big\|D^2(\mathcal{F}_q^q - \overline{\mathcal{F}}_q^q) \circ \mathcal{D}^{(q,k)}\big\|_{C^0(V)} \leq c_1\, \delta_0^{2\nu} \tag{2.27}$$

from Lemma 2.6, for a constant $c_1 > 0$ large enough.

As for the second term $\overline{\mathcal{C}}^{(q,k)} = \overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)} - J^1_{\omega^{(q,k)}}(\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)})$, we now choose $d_0 > \delta_0$, and remark that it is enough to find upper bounds for the second derivative of $\overline{\mathcal{F}}_q^q \circ \mathcal{D}^{(q,k)}$. We will make use of the explicit expression $\overline{\mathcal{F}}_q^q = \Phi^{\frac{1}{q} f^{(q)}} \circ \mathcal{F}_*^q \circ \mathcal{D}^{(q,k)}$ and begin with the second derivative of $\mathcal{F}_*^q \circ \mathcal{D}^{(q,k)}$. Since both maps are direct products, their composition is a product too, and its second factor is affine, so one has to consider only the first one:

$$\mathcal{P}^q \circ \mathcal{D}_1^{(q,k)}(x) = a^{(q)} + \Big(d_0^\nu x^{(h)} - T\big(h^{(q)} + \tfrac{d_0^\nu}{T_q'}(x^{(h)} - x^{(v)})\big), \tfrac{d_0^\nu}{T_q'}(x^{(h)} - x^{(v)})\Big).$$

The second derivative contains only terms of the form

$$\frac{d_0^\nu}{(T_q')^2}T''(h^{(q)}+\tfrac{d_0^\nu}{T_q'}x^{(h)}),\qquad \frac{d_0^\nu}{(T_q')^2}T''(h^{(q)}+\tfrac{d_0^\nu}{T_q'}x^{(v)}),$$

and usual estimates analogous to (2.23) yield

$$\frac{1}{(T_q')^2}T''(h^{(q)}+\tfrac{d_0^\nu}{T_q'}x^{(h)})\sim\frac{2\pi(h^{(q)})^2}{(h^{(q)}+\tfrac{d_0^\nu}{T_q'}x^{(h)})^2},$$

so one immediately obtains the inequality $\left\|D^2\mathcal{F}_*^q\circ\mathcal{D}^{(q,k)}\right\|_{C^0(V)}\le c_2\,d_0^{2\nu}$ for c_2 large enough. Now since the perturbative diffeomorphism has bounded derivatives, one gets the final inequality

$$\left\|D^2\overline{\mathcal{C}}^{(q,k)}\right\|_{C^0(V)}\le c_3\,d_0^{2\nu}\tag{2.28}$$

for c_3 large enough. Therefore the conclusion follows from the mean value theorem applied twice to (2.28), together with inequality (2.27), choosing $d_1>d_0$ and q_1 large enough.
□

Proof of Lemma 2.5. It is now an immediate consequence of Lemma 2.9, Lemma 2.4 and Lemma 2.10. One simply has to choose the width $\overline{\sigma}$ so as to obtain the inequality $\overline{d}^{\nu(q;\sigma)}\le\frac{1}{q^2}$ for $\overline{d}$ slightly larger than d_1, which is possible thanks to equation (2.8). □

3 Proofs of Theorem A and Theorem B

This section is very similar to the corresponding one in [27]. We "add degrees of freedom" to our family $(\mathcal{F}_q)$, in two different and consecutive steps. The first one is based on the coupling lemma introduced in [28], which applies to discrete systems; it makes it possible to pass from the initially hyperbolic context on $\mathbb{A}^2$ to the initially elliptic one on $\mathbb{A}^n$, $n\ge 3$, and to prove Theorem A. The second step is an analytic suspension to pass from discrete systems on $\mathbb{A}^n$ to continuous Hamiltonian systems on $\mathbb{A}^{n+1}$ and prove Theorem B.

3.1 From initially hyperbolic to initially elliptic perturbations

The base of the construction is the coupling lemma introduced in [28]. This lemma enables us to "embed" the previous family $(\mathcal{F}_q)$, or more precisely a subsequence $\mathcal{F}_{q_j}$, into an initially elliptic sequence of diffeomorphisms (Ψ_j) of $\mathbb{A}^n$ which converges to the elliptic completely integrable diffeomorphism $\Phi^{\frac12(r_1^2+\cdots+r_n^2)}$ when j tends to $+\infty$.

Troughout this section we split the $2n$-dimensional annulus $\mathbb{A}^n=\mathbb{A}^2\times\mathbb{A}^{n-2}$ into two factors and adopt the following notation for the variables:

$$x=(\theta_1,\theta_2,r_1,r_2)\in\mathbb{A}^2,\quad \widehat{x}=(\widehat{\theta},\widehat{r})=(\theta_3,...,\theta_n,r_3,...,r_n)\in\mathbb{A}^{n-2}.$$

3.1.1 The coupling lemma

We refer to [28] for the proof of the following coupling lemma, which was already used in a similar context in [27].

Lemma 3.1. *Consider two diffeomorphisms F and G of $\mathbb{A}^m$ and $\mathbb{A}^{m'}$ respectively, an integer $N \geq 2$, and an N-periodic point $a \in \mathbb{A}^{m'}$ for G. Let $f : \mathbb{A}^m \to \mathbb{R}$ and $g : \mathbb{A}^{m'} \to \mathbb{R}$ be two Hamiltonian functions which generate complete vector fields, and assume furthermore that g satisfies the following synchronization conditions:*

$$g(a) = 1; \quad g(G^k(a)) = 0,\ 1 \leq k \leq N-1; \qquad dg(G^k(a)) = 0,\ 0 \leq k \leq N. \tag{3.29}$$

Then, if $\Psi = \Phi^{f \otimes g} \circ (F \times G)$, the following equality

$$\Psi^N(x, a) = \Big(\Phi^f \circ F^N(x), a\Big) \tag{3.30}$$

holds for $x \in \mathbb{A}^m$.

An immediate consequence is that the submanifold $\mathcal{V} = \mathbb{A} \times \{a\}$ is invariant under Ψ^N, and that, canonically identifying $\mathcal{V}$ with $\mathbb{A}$, the restriction of Ψ^N to $\mathcal{V}$ is given by $\Phi = \Phi^f \circ F^N$. As a consequence, the system $(\mathcal{V}, \Phi)$ may be seen as a subsystem of Ψ^N.

3.1.2 The family Ψ_j

We denote by $(p_j)_{j \geq 0}$ the ordered sequence of prime numbers. As in the previous section, the diffeomorphisms we now construct will be obtained by composing the time-one map of a Hamiltonian function by the time-one map of a small perturbation, but this time the Hamiltonian function is not fixed and converges to the completely integrable Hamiltonian $\frac{1}{2}r^2$.

To be more precise, for $j \geq n-3$ we consider the maps

$$\Psi_j = \Phi^{S^{(j)}} \circ \Phi^{H_j}$$

where

$$H_j = \tfrac{1}{2}(r_1^2 + \cdots + r_n^2) + \tfrac{1}{N_j^2} \cos 2\pi\theta_1 \qquad \text{with} \qquad N_j = p_{j-(n-3)}\, p_{j-(n-4)} \cdots p_j, \tag{3.31}$$

and where $S^{(j)}$ is an analytic function to be defined below, which will depend only on the angles, and the norm of which will satisfy the inequality

$$\big\| S^{(j)} \big\|_{C^0(V_\sigma)} \leq \tfrac{1}{N_j^2} \tag{3.32}$$

where V_σ was defined in Section 1.2, and where σ was introduced in Proposition 2.1.

To obtain the function $S^{(j)}$ we apply the previous coupling lemma to the diffeomorphisms

$$F_j = \Phi^{\frac{1}{2}(r_1^2 + r_2^2) + \frac{1}{N_j^2} \cos 2\pi\theta_1} \quad \text{and} \quad G = \Phi^{\frac{1}{2}(r_3^2 + \cdots + r_n^2)}$$

of $\mathbb{A}^2$ and $\mathbb{A}^{n-2}$ respectively. The role of f is played by the function $\frac{1}{q}f^{(q)}$ defined in the previous section, for a suitable q to be defined below. The characteristic period of the coupling will be the integer N_j defined in (3.31), and we choose the N_j-periodic point

$$a^{(j)} = (0, \widehat{r}^{(j)}) \in \mathbb{A}^{n-2} \quad \text{with} \quad \widehat{r}^{(j)} = (1/p_{j-(n-3)}, \cdots, 1/p_j),$$

for the diffeomorphism G.

We then have to find an analytic function $g^{(j)}$ which satifies the conditions (3.29). We proceed as in [27] and introduce for $p \in \mathbb{N}^*$ the analytic function $\eta_p : \mathbb{T} \to \mathbb{R}$ defined by $\eta_p(\theta) = \left(\frac{1}{p}\sum_{\ell=0}^{p-1}\cos 2\pi\ell\theta\right)^2$, which satisfies $\eta_p(0) = 1$, $\eta_p(k/p) = 0$ for $1 \le k \le p-1$ and $\eta_p'(k/p) = 0$ for $0 \le k \le p$. We set

$$g^{(j)}(\theta_3, \cdots, \theta_n) = g_3^{(j)}(\theta_3)\cdots g_n^{(j)}(\theta_n) \quad \text{with} \quad g_i^{(j)}(\theta_i) = \eta_{p_{j-(n-i)}}(\theta_i) \quad \text{for } 3 \le i \le n.$$

One easily checks that $g^{(j)}$ satisfies the desired conditions (3.29). Note that $g^{(j)}$ is an analytic function, the norm of which is easily estimated from above:

$$\left\|g^{(j)}\right\|_{C^0(V_\sigma)} \le \mathrm{e}^{4\pi\sigma(n-2)p_j}. \tag{3.33}$$

Finally, we set

$$S^{(j)} = \tfrac{1}{q_j}f^{(q_j)} \otimes g^{(j)} \qquad \text{with} \qquad q_j := N_j^4[1 + \mathrm{e}^{8\pi\sigma(n-2)p_j}]. \tag{3.34}$$

So, by equations (2.9) and (3.33), the norm of the function $S^{(j)}$ satisfies inequality (3.32).

Now the application of the coupling lemma 3.1 immediately yields the following result.

Lemma 3.2. *Let* $\Phi_j = \Phi^{\frac{1}{q_j}f^{(q_j)}} \circ F_j^{N_j}$. *Then for* $(x_1, x_2) \in \mathbb{A}^2$,

$$\Psi_j^{N_j}\big((x_1, x_2), a^{(j)}\big) = (\Phi_j(x_1, x_2), a^{(j)}). \tag{3.35}$$

The submanifold $\mathcal{V}^{(j)} = \mathbb{A}^2 \times \{a^{(j)}\}$ *is thus invariant under* $\Psi_j^{N_j}$.

Moreover, if $\sigma_{N_j}\big(\theta_1, \theta_2, r_1, r_2\big) = \big(\theta_1, \theta_2, N_j r_1, N_j r_2\big)$, *the conjugacy relation*

$$\Phi_j = (\sigma_{N_j})^{-1} \circ \mathcal{F}_{q_j/N_j} \circ \sigma_{N_j} \tag{3.36}$$

holds true for all $j \in \mathbb{N}$.

Let us introduce the point $u^{(j)} = \sigma_{N_j}^{-1}(\zeta^{(q_j/N_j)})$, where the $\zeta^{(q)}$ were defined in Proposition 2.1. Using equation (3.36) one checks that $N_j(q_j/N_j)^2 = q_j^2/N_j$ iterates of Φ_j make the r_2 action of the point $u^{(j)}$ drift over an interval of length 1. As a consequence, if one sets

$$z^{(j)} = (u^{(j)}, a^{(j)}) \in \mathbb{A}^n \tag{3.37}$$

equation (3.35) shows that q_j^2 iterates of Ψ_j make the r_2 action of $z^{(j)}$ drift over a length 1.

3.1.3 Complex analytic estimatesfor Ψ_j and proof of Theorem A

The following lemma, proved in [27], provides the necessary analytic estimates on the distance between the perturbed diffeomorphism Ψ_j and the elliptic system $\Phi^{\frac{1}{2}r^2}$. We use the notation introduced in Section 1.2 for the complex neighborhoods.

Lemma 3.3. *Let $\varrho = \overline{\sigma}/6$, where $\overline{\sigma}$ were defined in Proposition 2.1 Then there exists $j_0 \in \mathbb{N}$ and $c > 0$ such that for $j \geq j_0$*

$$\left\|\Psi_j - \Phi^{\frac{1}{2}r^2}\right\|_{C^0(U_\varrho)} \leq \frac{c}{N_j^2}. \tag{3.38}$$

Proof of Theorem A. It only remains to gather together Lemma 3.2 and the definition of the drifting point $z^{(j)}$, the estimate (3.32) for the norm of the function $S^{(j)}$ and the estimates of Lemma 3.3. The main point is to determine the relation between the parameter q_j and the size of the perturbation $\varepsilon_j = \left\|\Psi_j - \Phi^{\frac{1}{2}r^2}\right\|_{C^0(U_\varrho)}$. If j_0 is large enough, the Prime Number Theorem yields the inequality $p_{j-(n-3)} \geq \frac{1}{2}p_j$ for $j \geq j_0$. Therefore, since $\varepsilon_j \leq \frac{c}{N_j^2}$ by Lemma 3.3,

$$p_j \leq 2\, N_j^{\frac{1}{n-2}} \leq 2\, c^{\frac{1}{2(n-2)}} \left(\tfrac{1}{\varepsilon_j}\right)^{\frac{1}{2(n-2)}}.$$

On the other hand, by definition

$$q_j = N_j^4[1 + e^{-8\pi\sigma(n-2)p_j}] \leq \exp\left(\overline{\kappa}\left(\tfrac{1}{\varepsilon_j}\right)^{\frac{1}{2(n-2)}}\right)$$

for $\overline{\kappa} > 16\pi c^{\frac{1}{2(n-2)}}\sigma(n-2)$ and $j \geq j_0$ large enough. The proof easily follows. $\square$

3.2 Analytic suspension and proof of Theorem B

We now want to pass from the discrete case to the continuous one. As in [27] we follow the approach of Kuksin and Pöschel.

Theorem ([21]). *Let D be a convex bounded domain in $\mathbb{R}^n$, $n \geq 1$. For $j \in \mathbb{N}$ let $F_j : \mathbb{T}^n \times D \to \mathbb{A}^n$ be an exact-symplectic diffeomorphism, with analytic continuation to a complex neighborhood U_ϱ, for some $\varrho > 0$ independent of j. Let h be the Hamiltonian function defined on $\mathbb{A}^n_c$ by $h(r) = \frac{1}{2}(r_1^2 + \cdots + r_n^2)$ and $\Phi^h : \mathbb{A}^n_c \to \mathbb{A}^n_c$ its time-one map. Let $\varepsilon_j = \left\|F_j - \Phi^h\right\|_{C^0(U_\varrho)}$, and assume that $\varepsilon_j \to 0$ when $j \to \infty$.*

Then there exists $j_0 \in \mathbb{N}$ such that for all $j \geq j_0$ there exists a real analytic 1-periodic time dependent Hamiltonian $\mathcal{H}_j$ defined on $\mathbb{T}^n \times D \times \mathbb{T}$ such that the time-one map $\Phi^{\mathcal{H}_j}$ is well-defined on $\mathbb{T}^n \times D$ and coincides with F_j. Moreover, there exists a constant $\rho < \varrho$ such that each $\mathcal{H}_j$, $j \geq j_0$, is analytic on $\overline{U}_\rho = V_\rho(\mathbb{T}^{n+1}) \times W_\rho(D)$ and satisfies

$$\left\|\mathcal{H}_j - h\right\|_{C^0(\overline{U}_\rho)} \leq C\,\varepsilon_j \tag{3.39}$$

for some constant $C > 0$ independent of j.

We can now pass to the proof of Theorem B. Let $R > 0$ be fixed, and let D be the ball of radius R centered at 0 in $\mathbb{R}^n$. For $j \geq 0$ we consider the restriction $F_j = \Psi_{j|\mathbb{T}^n \times D}$, with values in $\mathbb{T}^n \times \mathbb{R}^n$. As a composition of two Hamiltonian time-one maps, F_j is exact-symplectic.

Since $W_\varrho(D) \subset W_\varrho(\mathbb{R}^n)$, Lemma 3.3 shows that F_j admits an analytic continuation to U_ρ, with the same estimate (3.38). So the previous suspension theorem provides us with a non-autonomous Hamiltonian function $\mathcal{H}_j : \overline{U}_\rho \to \mathbb{C}$ satisfying (3.39). To obtain an autonomous system we simply have to consider the function $\mathfrak{H}_j$ defined on $\overline{U}_\rho \times \times \mathbb{C}$ by

$$\mathfrak{H}_j(\bar{\theta}, \bar{r}) = \mathcal{H}_j(\theta, r, \theta_{n+1}) + r_{n+1}.$$

For each energy $\boldsymbol{e} \in \mathbb{R}$, the surface $\mathfrak{H}_j^{-1}(\boldsymbol{e}) \cap \{\theta_{n+1} = 0\}$ is symplectic, transverse to the flow, and admits (θ, r) as a coordinate system. In this system, the associated return map coincides with Ψ_j. Theorem B immediately follows. □

4 Sternberg's theorem for normally hyperbolic manifolds

In this section we develop a local conjugacy result which we then apply to our construction of drifting orbits. We found it convenient to make this section essentially self-contained, including the notation. The application to our case, which yields the crucial Lemma 2.6 above is detailed in §4.5 below.

4.1 Setup and synopsis

Let f_0, f_1 be two symplectic diffeomorphisms of the symplectic manifold V, which preserve the submanifold $M \subset V$ and are normally hyperbolic along V. All these data, namely V, M, f_0, f_1 are assumed to be analytic. We wish to show that if f_0 and f_1 have a contact of large enough order along M, they are C^ℓ conjugate in a neighborhood of M, for an integer $\ell \geq 1$ which we will compute. We first make the setting both more precise and more restrictive.

1. The maps f_0, f_1 are hyperbolic transversely to M; here we will assume a simple product structure and that their invariant manifolds have been simultaneously straightened. Namely we take $V = M \times E^s \times E^u = M \times E$ with $dim(V) = d$, $dim(M) = m$, $E^s \simeq \mathbb{R}^{n_s}$, $E^u \simeq \mathbb{R}^{n_u}$, $m + n_s + n_u = d$. We will *not* assume that M is symplectic from the start, because it turns out that the proof for a symplectic M actually uses the non symplectic case. If M is indeed symplectic, a case which is of special interest, one has $n_u = n_s = n$ and the vector space $E = E^s \oplus E^u$ is endowed with the standard symplectic structure, E^s and E^u being Lagrangian subvector spaces. The manifold V is provided with the product symplectic structure and we regard M, identified with $M \times \{0\}$, as a symplectic submanifold of V. Finally, as mentioned above, we assume that TM is trivial.

In the applications we have in mind, $M = \mathbb{A}^m \simeq T^*\mathbb{T}^m$ is the m-dimensional infinite ring (cylinder), a symplectic manifold with trivial tangent bundle.

2. Next we assume that f_0 and f_1 are isotopic, that is they can be interpolated by a family f_ε $(0 \le \varepsilon \le 1)$ and we refer once and for all to [2] for detail on the deformation method we will use. More precisely we assume that the family f_ε can be written in the form $f_\varepsilon = \Phi_\varepsilon^{F_\varepsilon} \circ f_0$, where F_ε $(0 \le \varepsilon \le 1)$ is a family of analytic Hamiltonians describing the deformation. In other words f_ε solves the evolution equation in ε:

$$\frac{d}{d\varepsilon} f_\varepsilon = \mathcal{F}_\varepsilon \circ f_\varepsilon, \tag{1}$$

where $\mathcal{F}_\varepsilon$ is the vector field with Hamiltonian F_ε.

3. Concerning regularity, we will work with data f_ε which are analytic in the space variables and we are interested in retrieving a C^ℓ conjugacy g between f_0 and f_1 in a neighborhood of M; ℓ will depend on the data. Regularity in ε is not essential for our purpose, which is to find and study the conjugating map g. It will turn out that continuity in ε will suffice (see below for detail).

4. Let us introduce 'coordinates': We let $x_s \in \mathbb{R}^{n_s}$ (resp. $x_u \in \mathbb{R}^{n_u}$) describe E^s (resp. E^u) and coordinatize the points of M by means of y. Because the latter variety is not necessarily compact we will need estimates that are uniform over M, that is w.r.t. y. We write $z = (y, x) = (y, x_s, x_u)$ for a point in V. We let $|x_s|$, $|x_u|$ and $|x|$ denote the norm (say Euclidean norm) on E^s, E^u and E, with E^s and E^u mutually orthogonal. We will write D_y for y-derivatives, that is derivatives along M and D_x for x-derivatives, that is transverse derivatives; D_s and D_u denote derivative w.r.t. x_s and x_u respectively. In trying to keep a manageable notation, we will (almost always implicitly) use a multiindex notation for the various tensor quantities which appear; the reader should be able to restore a fully detailed expression if need be (which will in principle not be the case).

5. The diffeomorphisms f_ε are always assumed to coincide on M together with their derivatives; the order of contact will in fact be assumed to be much larger. We denote by $A(y) = Df_x(y, 0)$ the common value of this derivative in the transverse direction. We assume that $W^s = M \times E^s$ (resp. $W^u = M \times E^u$) is the stable (resp. unstable) invariant manifold of f_ε (for any $\varepsilon \in (0, 1)$). The matrix $A(y)$ is thus block diagonal and we get a contracting endomorphism $A_s(y)$ of E^s, and a dilating one $A_u(y)$ on E^u. More quantitatively, let $Spec(A_s(y))$ denote the spectrum of $A_s(y)$ (as a finite set) and $|Spec(A_s(y))|$ the list of the norms of its eigenvalues. We assume that:

$$|Spec(A_s(y))| \subset (\mu_s, \lambda_s), \quad \text{with} \quad 0 < \mu_s \le \lambda_s < 1, \tag{2}$$

these bounds being indeed independent of $y \in M$. In the same vein, we assume that

$$|Spec(A_u^{-1}(y))| \subset (\mu_u, \lambda_u) \quad \text{with} \quad 0 < \mu_u \le \lambda_u < 1. \tag{3}$$

If M is symplectic, $A(y)$ is a symplectic operator and $A_u^{-1} = {}^tA_s$, so that $\mu_u = \mu_s$, $\lambda_u = \lambda_s$. Concerning the restriction of the system to M we only assume that the vector field is bounded, that is there exists $\nu > 0$ such that for any $y \in M$:

$$|D_y f(y, 0)| \le \nu, \tag{4}$$

where $f = f_\varepsilon$ and the restrictions of the f_ε coincide on M anyway.

Items 1 through 5 above provide our general setting which will be refined below. We note that although the geometric setting in 1 may seem quite restrictive, one can reduce seemingly much more general situations to it, using formal constructions (see in particular [20] and [12], §3.5). We also insist that the assumptions and the conclusions will be local around M, so that one can actually work on the product $M \times B_\rho$ for some $\rho > 0$, where B_ρ denotes the ball $\{|x| < \rho\} \subset E$. In particular it is enough in practice to analytically straighten local invariant manifolds.

Following [2] we will use the so-called deformation method in order to solve the local conjugacy problem. We refer to the latter article for a concise exposition of the method with references. Here we will confine ourselves to a bare minimum. We wish to conjugate f_0 and f_1 and have connected this pair by a path f_ε $(0 \le \varepsilon \le 1)$. We will try to achieve more, and look for a family g_ε such that $g_\varepsilon^{-1} \circ f_\varepsilon \circ g_\varepsilon = f_0$ for all $\varepsilon \in (0, 1)$, so that $g = g_1$ will answer the initial problem.

We also require that g_ε be regular enough (C^1) in ε so that it will satisfy an evolution equation of the same form as f_ε, say:

$$\frac{d}{d\varepsilon} g_\varepsilon = \mathcal{G}_\varepsilon \circ g_\varepsilon, \tag{5}$$

with the initial condition $g_0 = 1$ (the identity map). Finally we are also looking for a symplectic conjugacy, that is we want $\mathcal{G}_\varepsilon$ to be a Hamiltonian vector field, with Hamiltonian G_ε. In order to derive the equation for $\mathcal{G}_\varepsilon$, hence for G_ε, one simply translates the fact that $g_\varepsilon^{-1} \circ f_\varepsilon \circ g_\varepsilon$ is a constant map, namely f_0, so that the derivative of this quantity vanishes. Formal computations (see [2], §§2,3) lead to the equation satisfied by the 'conjugating Hamiltonian' G_ε, namely:

$$G_\varepsilon - G_\varepsilon \circ f_\varepsilon^{-1} = F_\varepsilon. \tag{E}$$

It is useful to write the equivalent equation obtained by composing each term with f_ε, which also amounts to changing f_ε into its inverse:

$$G_\varepsilon - G_\varepsilon \circ f_\varepsilon = -F_\varepsilon \circ f_\varepsilon. \tag{E'}$$

The problem of conjugating f_0 to f_1 has now been reduced to finding a solution of (E) or equivalently of (E'), and study its regularity in the space variables. As for regularity in ε, it is enough for our purpose to be able to solve equation (5) above. By Cauchy-Lipschitz we should require that $\mathcal{G}_\varepsilon$ be Lipschitz in z and continuous in ε, which is the time-like variable. In turn the vector field $\mathcal{G}_\varepsilon$ is derived from the Hamiltonian G_ε by taking z-derivatives so that $\mathcal{G}_\varepsilon$ and G_ε have the same regularity in ε. Hence we only need G_ε to be continuous in ε, something which will be obvious from the algorithms we use so that we will not need to return to this issue. Regularity in ε could of course be discussed if need be, much as space regularity. Note that the above discussion also applies to $\mathcal{F}_\varepsilon$ and F_ε, that is to equation (1).

We now write down the formal solutions of (E) and (E') obtained by iterating these equations. We get:

$$G_\varepsilon = \sum_{n=0}^{\infty} F_\varepsilon \circ f_\varepsilon^{-n} \tag{FS}$$

and

$$G'_\varepsilon = -\sum_{n=1}^{\infty} F_\varepsilon \circ f_\varepsilon^{n}, \tag{FS'}$$

which will be put to use below.

This completes our description of the setting and the main characters. Let us now move to a brief exposition of the plot, which may sound a little intricate at first reading. The case we are interested in displays several specific features which we will try to accomodate as best as possible or indeed take advantage of. Namely a) the initial data, say f_0 and the deformation F_ε, are analytic, b) there is an invariant manifold M which not only is not reduced to a point but is also possibly not compact and c) we are working in a symplectic setting. This last feature has already been incorparted by reducing the problem to solving the 'homological equation' (E) (or equivalently (E')). Note that these are scalar equations which are moreover linear with respect to the perturbation F.

We adopt the classical overall strategy inaugurated by S.Sternberg in [32]. That is we first treat the contracting case, thus assuming $n_u = 0$, that is $V = W^s = M \times E^s$; clearly the expanding case ($V = W^u$) can be treated in the same way, changing the diffeomorphisms into their inverses. Here we can take full advantage of the analyticity of the initial data. The problem is local near M and analyticity enables one to get an analytic conjugacy in a neighborhood of M which can be explicitly determined. In short analyticity in the contracting case enormously simplifies the problem and yields an effective analytic solution. However, because of b) above, that is the possible non compactness of M, we do have to add an assumption of uniform contraction along M.

One then wishes to reduce the general case to the contracting one, applying the results in the contracting case for the triple (M, V, f) to the triple (W^s, V, f^{-1}) in the general hyperbolic case. Here f stands for f_ε for some fixed $\varepsilon \in (0,1)$; it preserves W^s which is an attracting invariant submanifold for f^{-1}. However to this end one first has to 'prepare' the system so that f_0 and f_1 acquire a contact of high order along W^s (and not merely along M). This can be done by dealing with the jet of the perturbation along W^s, so the traditional 'preparation lemma' deals with a contracting problem which is however more intricate than equation (E). But again, in the case of analytic data, it can be treated very simply and efficiently, confining oneself however to a neighborhood of M which is any case unavoidable because the end result, that is the existence of a germ of conjugacy between f_0 and f_1 is in essence local around M. So by using the analyticity of the data in a neighborhood of M on the stable manifold W^s, we can simplify a large part of the proof and make it more effective, including in the presence of a nontrivial invariant manifold M.

This is however not the end yet because it is (except in very special cases) not possible to prepare the system gobally along W^s and indeed in a neighborhood of that invariant manifold. One in fact has to smoothly truncate using functions whith compact support, thus loosing both analyticity and uniqueness. Indeed in the contracting (not necessarily analytic) case, one gets a uniqueness result for the (germ of) conjugacy which does not hold in the general hyperbolic case, and that conjugacy is analytic if the data are analytic, which also fails in the general hyperbolic case. So after preparing the system along W^s, one has to apply a result in the contracting but not analytic case. This is the reason why we also have to address the latter issue. We will give a direct treatment of that contracting but not analytic case, confining ourselves to the case of equation (E) (or (E')) which is the only one we will have to use. Technically speaking we improve in that case on the results of [2] in a way which will be made precise below. We will also point out in due time a phenomemon having to do with the regularity of the solution along M which seems to have passed unnoticed as it has to do with both analyticity and the presence of a nontrivial invariant manifold.

4.2 The contracting case

In this paragraph we restrict attention to W^s or equivalently assume that $n_u = 0$, $V = M \times E^s$. We consequently simplify the notation, writing x instead of x_s for a point of $E = E^s$; similarly we write $z = (y, x) \in V$. As a rule we will in fact omit the subscript or superscript s altogether in this section. As explained above the dependency on ε will play essentially no role below, so that in order to clarify notation further we will henceforth drop the subscript ε at most places. Everything will take place at a fixed value of $\varepsilon \in (0, 1)$ and as mentioned above continuity in ε, which is ultimately all we need, will be trivially satisfied.

For any $\rho > 0$, we let $B_\rho \subset E$ denote the ball $|x| < \rho$ and everything will take place inside a tubular neighborhood of M of the form $M_\rho = M \times B_\rho$. More precisely we assume that there exists $\rho > 0$ such that f_ε is analytic over M_ρ for all $\varepsilon \in (0, 1)$ or equivalently that f_0 and the deformation Hamiltonian F_ε satisfy this assumption. We also assume that F_ε has a zero of order at least $k > 2$ along M, which can be translated as

$$Sup_{y \in M, 0<|x|<1} \quad |x|^{-k} |F_\varepsilon(y, x)| < \infty. \tag{1}$$

This entails that f_0 and f_1 (or any f_ε, $\varepsilon \in (0, 1)$) have a contact of order at least $k - 1 > 1$ along M and can be expressed in the language of weighted conical norms used in [10] and other papers. This condition already contains a uniformity assumption along M, but we will actually need somewhat more, namely (3) below.

Next we assume of course that the diffeomorphisms are uniformly contracting along M. Because $k > 2$, $A(y)(= A_s(y)) = D_x f_\varepsilon(y, 0)$, that is the derivative of f_ε along $E = E^s$, does not depend on ε and it is assumed to satisfy (2) of §1, that is:

$$|A(y)v| \leq \lambda|v| \qquad \text{for} \qquad y \in M, v \in E, \tag{2}$$

and some $\lambda = \lambda_s < 1$.

Analyticity supplemented by (1) and (2) would suffice in the case of a compact invariant manifold M. Here we will need to reinforce these assumptions in order to make them uniform in a tubular neighborhood of M. We assume first that there is a constant $K_\rho > 0$ such that for $z \in M_\rho$ one has:

$$|F(z)| \leq K_\rho |x|^k. \tag{3}$$

Next let us write $f(z) = (f_y(z), f_x(z)) \in V = M \times E$. We assume that for $z \in M_\rho$ one has:

$$|D_x f_x(z)| \leq \lambda_\rho < 1 \tag{4}$$

for some λ_ρ ($\lambda \leq \lambda_\rho < 1$). This should hold of course for all f_ε, $\varepsilon \in (0,1)$. Under the above assumptions we get the following:

Proposition 4.1 (Analytic contracting case)**.** *Assume that f_0 and the deformation F_ε are defined and analytic (w.r.t z) over $M_\rho = M \times B_\rho$ for some $\rho > 0$; assume moreover that they satisfy (3) and (4) above on that domain. Then there is a unique germ of continuous conjugacy g between f_0 and f_1 which is the identity on M. It is actually defined and analytic over M_ρ and it has a contact of order k with the identity along M.*

Se we not only get existence and uniqueness but we also have an explicit domain over which the conjugacy is defined. Using continuation to the complex domain and the Cauchy formula in a standard way, one can then estimate derivatives. We note again that if M is compact, (1) and (2) imply that (3) and (4) hold true for some $\rho > 0$. The statement above is also purely local, and the existence of the data is actually required only on M_ρ. Proving the above statement is equivalent to showing that (E) and (E') have a unique solution which vanishes on M, that it is actually analytic in M_ρ and that it vanishes on M at order k. We will do just that presently and this is the way in which we will cast the analogous statements in the sequel.

The proof in the present analytic contracting case is quite straightforward. First note that (4) implies that for any positive integer n:

$$|f_x^n(z)| \leq \lambda_\rho^n |x|, \tag{5}$$

where we write $f_x^n = (f^n)_x$ for simplicity. By iterating (E') to order N, we find that any solution G satisfies:

$$G(z) = -\sum_{n=1}^{N} F \circ f^n(z) + G(f^{N+1}(z)). \tag{6}$$

If we require G to vanish on M, we see using (3) and (5) that the last term goes to 0 as N increases to infinity and G thus has to coincide with the formal solution (FS'). The convergence of that series, which we now simply call G, is also obvious since by (3) and (5) the general term is dominated by $K_\rho(\lambda_\rho^n|x|)^k$. We find that there is indeed a unique solution G of (E) vanishing on M, that it is analytic on M_ρ because the convergence is uniform on that domain. We moreover get the estimate:

$$|G(z)| \leq \sum_{n\geq 1} K_\rho(\lambda_\rho^n|x|)^k \leq C_\rho|x|^k, \tag{7}$$

with $C_\rho = \lambda_\rho^k(1-\lambda_\rho^k)^{-1}K_\rho$. This confirms that G vanishes on M at order k. □

The 'preparation lemma' leading to the general hyperbolic case will require solving an equation a little more complicated than (E). Let $Q = Q(z)$ be a square matrix depending on $z \in M_\rho$. We are now interested in solving:

$$G(z) - Q(z)G \circ f(z) = F(z). \tag{8}$$

Here G is now a vector function (of the same size as Q) and we have denoted the vector perturbation simply by F (compare (E') in §1). It turns out that in the analytic category it is essentially as easy to study (8) as (E) or (E'). We have:

Proposition 4.2 (Analytic contracting case with a cocycle)**.** *Assume that f, F and Q are defined and analytic over M_ρ, that f and F satisfy (3) and (4) and that the norm of $Q = Q(z)$ is bounded by $\mu_\rho \geq 0$ on that domain. Then if $\lambda_\rho^k\mu_\rho < 1$ there is a unique solution of (8) which vanishes on M. It is analytic on M_ρ and vanishes on M at order k.*

The proof is essentially the same as above. The candidate formal solution vanishing on M reads:

$$G(z) = \sum_{n\geq 0}\Big(\prod_{m=0}^{m=n-1} Q(f^m(z))\Big)F(f^n(z)). \tag{9}$$

The general term is now dominated on M_ρ by $K_\rho\mu_\rho^n(\lambda_\rho^n|x|)^k$ and one concludes as in the proof of Proposition 4.1. □

Once again the assumptions and the conclusion are local around M. If one assumes that as ρ decreases to 0, λ_ρ tends to $\lambda = \lambda_s$ and μ_ρ tends to a value $\mu = \mu_0$, one finds that provided $\lambda^k\mu < 1$, one gets the conclusion on some neighborhood of M. On the other hand, as soon as $\lambda_\rho < 1$, the conclusion will hold true on M_ρ for k large enough, that is in the original problem if one requires a contact of high enough order between the original diffeomorphisms.

We now turn to the smooth setting and will treat only the case of equation (E). We use a direct method which enables us to get quite precise results but would be substantially more difficult to apply in the case of equation (8). More abstract approaches, using classical fixed point results for contracting maps are naturally less sensitive but also less precise. We refer to [2] for a sketch of the proof in a similar setting (see Lemma 5.5

there; we note that we will implement below the Remark at the end of that statement). So from now on we consider data which are defined and of class C^r ($k < r \leq \infty$) on a tubular neighborhood of M, of the form $M_\rho = M \times B_\rho$ for some ρ. Everything is still local around M and although it is useful to keep r as a free parameter, the reader may think primarily of the smooth case $r = \infty$. We always keep the same letter ρ in the various assumptions but needless to say, it only implies the existence of a value $\rho > 0$ such that the assumption at hand is satisfied.

Until now we did not have to mention the nature of the diffeomorphisms when restricted to the invariant manifold M. This is quite remarkable in fact, because that spectrum (Lyapunov exponents) will indeed play an important role in the sequel. This is the phenomenon we alluded to at the end of §1. In the analytic case one gets an analytic solution in all variables; but in the smooth case, transverse regularity and regularity along M will appear to be quite different questions and we will discover the anisotropic character of the problem. Recall the pieces of notation D_x and D_y, as well as condition (4) in §1, stating that the Lyapunov exponents of the restricted diffeomorphisms are bounded by ν on M (for all $\varepsilon \in (0,1)$). Let us reinforce it in the usual way, assuming that in fact, possibly at the expense of shrinking M_ρ, one has:

$$|D_y f(z)| \leq \nu_\rho, \tag{10}$$

for $z \in M_\rho$ and some $\nu_\rho \geq \nu \geq 1$. The last inequality is by convention; we may increase ν and replace it by $\max(1, \nu)$, which we do for convenience.

We will start with a sample statement which will subsequently be generalized. We include it for illustrative purposes as it displays the seeds of the main phenomena. By (1) we have that $D_x^j F = 0$ for $j < k$; we can take the derivative in y, permute the derivatives and find that $D_x^j(D_y F) = 0$ for $j < k$. Note that here we are using of course the standard multiindex notation and $j < k$ actually means that the length of j is strictly smaller than the integer k. We hope that this simplified notation, which we also apply to tensor quantities will help clarify the text without causing misinterpretations. Let us stick for the time being to the case $j = 1$. We now assume the analog of (3) at order 1, namely that there exists a constant $K_\rho^{(1)}$ such that for $z \in M_\rho$ one has:

$$|F(z)| \leq K_\rho^{(1)} |x|^k, \quad |D_x F(z)| \leq K_\rho^{(1)} |x|^{k-1}, \quad |D_y F(z)| \leq K_\rho^{(1)} |x|^k. \tag{11}$$

It is as usual understood that this holds for all F_ε; moreover this assumption will automatically be fulfilled if M is compact. Note the fact that the x derivative vanishes on M at order $k-1$, whereas the y derivatives vanishes at order k.

Under these assumptions we have the following:

Proposition 4.3. *Assume that f and F are defined over $M_\rho = M \times B_\rho$ for some $\rho > 0$ and that they are of class C^r (w.r.t. z) on that domain, for some integer r with $0 < k \leq r \leq \infty$. Assume that the data satisfy (4), (10) and (11) on M_ρ. Then there is a unique continuous solution G of (E) vanishing on M; it can be extended to a function on M_ρ which is transversely (i.e. with respect to x) of class C^1. It is of class C^1 along M (i.e.*

with respect to y) provided the inequality $\lambda_\rho^k \nu_\rho < 1$ obtains (in which case G is of class C^1 on M_ρ) .

The proof begins as that of Proposition 4.1. Following the latter proof quite literally, it ensures in our case the uniqueness of the solution G and that it exists and is continuous over M_ρ. It remains to investigate its regularity. To start with, let us differentiate the defining formula (*cf.* (FS') in §1): $G = -\sum_{n\geq 1} F \circ f^n$. We get:

$$DG(z) = -\sum_{n\geq 1} DF(f^n(z))Df(f^{n-1}(z))\dots Df(f(z))Df(z), \tag{12}$$

where D stands for either D_x or D_y. We immediately encounter an important difference between x and y differentiation, that is between transverse and longitudinal regularity. For $D = D_x$, the n-th term of the sum in (12) can be estimated on M_ρ in the operator norm using, (5) and (11):

$$|D_xF(f^n(z))D_xf(f^{n-1}(z))\dots D_xf(f(z))D_xf(z)| \leq K_\rho^{(1)}(\lambda_\rho^n|x|)^{k-1}\lambda_\rho^n. \tag{13}$$

This in turn yields:

$$|D_xG(z)| \leq \sum_{n\geq 1} K_\rho^{(1)}|x|^{k-1}\lambda_\rho^{kn} = c|x|^{k-1}, \tag{14}$$

where the constant c is easily computable (we will use the letter c for the 'generic' constants, possibly making comments on their nature). This shows that under the above assumptions G is C^1 w.r.t. x with an explicit estimate on M_ρ; moreover D_xG vanishes at order $k-1$ on M. Anticipating a little, it will turn out that in fact there is no loss in x regularity with respect to the data. But regularity along M, that is w.r.t. y, rests on a different mechanism. In the n-th term of (12) with $D = D_y$, we get a product of n terms of the form $D_yf(f^j(z))$ which is asymptotically governed by the Lyapunov exponents of the restriction of f to M, namely:

$$|D_yF(f^n(z))D_yf(f^{n-1}(z))\dots D_yf(f(z))D_yf(z)| \leq K_\rho^{(1)}(\lambda_\rho^n|x|)^k\nu_\rho^n, \tag{15}$$

and from there:

$$|D_yG(z)| \leq \sum_{n\geq 1} K_\rho^{(1)}|x|^k(\lambda_\rho^k\nu_\rho)^n = c|x|^k, \tag{16}$$

with again an easily computable constant c. This finishes the proof of the Proposition and shows that D_xG (resp. D_yG) vanishes at order $k-1$ (resp. k) on M. Note that of course this vanishing property does not presuppose the existence of higher derivatives. □

We have just seen the first manifestation of the fact that when working along an invariant submanifold rather than in the neighborhood of a point, and whatever the characteristics of the induced flow, they can be compensated for by requiring a higher order contact of the initial diffeomorphisms f_0, f_1, that is by increasing k. The instructions for use of such statements of course depend on which parameters are considered free. One can for instance look at what happens on M and its normal bundle, that is determine ν and λ, pick k such that $\lambda^k\nu < 1$ and find ρ small enough so that one still has $\lambda_\rho^k\nu_\rho < 1$, assuming of course that ν_ρ and λ_ρ are continuous functions of ρ ($\nu_\rho \geq \nu \geq 1$, $\lambda \leq \lambda_\rho < 1$). Again if M happens to be compact, global uniformity along M prevails and assumptions over a tubular neighborhood follow automatically, *e.g.* (11) is a consequence of (1). We now explore higher regularity.

Existence and uniqueness of a solution G of (E) and (E') vanishing on M are proved as above, and in fact as in Proposition 4.1, together with the fact that it coincides with the formal series (FS') of §1, and we will use direct differentiation of that expression. We start from the obvious:

$$|D^\ell G| \leq \sum_{n\geq 1} |D^\ell(F\circ f^n)|. \tag{17}$$

Here ℓ is any multiindex and for the time being we do not distinguish between the transverse (D_x) and longitudinal (D_y) factors. We will prove as usual the existence of $D^\ell G$ by showing that the series converges locally uniformly and indeed uniformly over M_ρ under certain assumptions. We will in fact also obtain fairly explicit estimates for that quantity. Let us first recall (see *e.g.* [10], Appendice 1) the formula for the successive derivative of a composition of maps, namely:

$$D^\ell(F\circ h) = \sum_{q=1}^{\ell}\sum_{\underline{m}} \sigma_{\underline{m}}(D^qF\circ h)D^{m_1}h\ldots D^{m_q}h. \tag{18}$$

This is the so-called 'Faa-di Bruno formula', whose proof is formal and which is valid for any composition of maps (here denoted F and h) of Banach spaces. The second sum runs over the set of indices $\underline{m} = (m_1,\ldots,m_q)$ such that $m_j \geq 1$ for any j $(1\leq j\leq q)$ and $m_1+\ldots m_q = \ell$. The σ's are integers which can be defined recursively (see [10] or [3]). We will only retain the obvious fact that they can be bounded with a bound which depends on ℓ only and so can the number of terms in the sum (18).

Our task now consists in estimating (17) using (18) with $h = f^n$; we are however in an anisotropic setting, so that we need to distinguish between the x and y derivatives and study a kind of weighted form of (18). Very roughly speaking, any factor D_xf contributes a converging factor λ, whereas D_yf contributes a possibly diverging factor $\nu \geq 1$. Let $\ell \leq k$; we wish to study the existence and continuity of $D^\ell G$, to which end it is enough to investigate derivatives of the form $D_x^{\ell'}D_y^{\ell''}G$, with $\ell'+\ell'' = \ell$. All multiindices will now be split according to their x and y content, using primes for the first and double primes for the second set.

For $q \leq k$, we have from (1) that $|x|^{q'-k}D^qF$ is bounded as $|x|$ goes to 0. Here again only the number q' of x derivatives comes in: Taking y derivatives does not let the order of contact decrease. So for given $\ell \leq k$ it is sensible to require the higher order analog of (11) over M_ρ; it is once again a consequence of (1) for compact M. So we assume the existence of a constant $K_\rho^{(\ell)}$ such that for $z \in M_\rho$ and $q \leq \ell$ one has the following bound:

$$|D^qF(z)| \leq K_\rho^{(\ell)}|x|^{k-q'}. \tag{19}$$

We also have to assume that the successive derivatives of f are bounded over M_ρ, *i.e.* there exists a constant $D_\rho^{(\ell)}$ such that for $z \in M_\rho$ and $q \leq \ell$:

$$|D^q f| \leq D_\rho^{(\ell)}. \tag{20}$$

Under these assumptions, our goal is now to prove the following:

Proposition 4.4 (Smooth contracting case). *Assume that f and F are defined over $M_\rho = M \times B_\rho$ for some $\rho > 0$ and that they are of class C^r (w.r.t. z) on that domain for some integer r with $0 < k \leq r \leq \infty$. Assume that the data satisfy (4), (10) and (19) on M_ρ where $\ell \leq k$ and the inequality $\lambda_\rho^k \nu_\rho^\ell < 1$ is satisfied. Then:*
i) If the derivatives of f are bounded on M_ρ to order ℓ, that is if (20) holds, there is a unique solution G of (E) vanishing on M and it can be extended to a function of class C^ℓ on M_ρ;
ii) If the derivatives of f of order $\leq r$ are bounded on M_ρ, the function G is transversely (i.e. with respect to x) of class C^r on M_ρ.

If $r = \infty$ no uniformity with respect to the length of the multiindex is required in the last boundedness assumption on the derivatives of f. So if M is compact this assumption is automatically fulfilled (as well as (20) *a fortiori*). As usual one can then replace (4), (10) and (19) by the corresponding infinitesimal assumptions: (4) is implied by (2), (10) follows from (4) in §1 and (19) is a consequence of (1).

We need only prove the regularity assertions. We will show i) and ii) at one go but one could give a simpler direct proof of ii). It may be useful to briefly explain why. The point is that when taking transverse derivatives there are two sources of convergence, one being the contact along M (order of vanishing of F), and the other being contraction. This is enough to ensure that the solution G is transversely as smooth as the data, provided the necessary derivatives are bounded, as recorded in ii). By contrast, as already illustrated in Proposition 4.3, in the case of i) a high order of contact has to compensate for the possible divergence originating from the possibly large Lyapunov exponents (ν) of the flow on M.

First putting (19) and (5) together, we find that:

$$|D^q F(f^n(z))| \leq K_\rho^{(\ell)} (\lambda_\rho^n |x|)^{k-q'}, \tag{21}$$

still for $q \leq \ell$ and $z \in M_\rho$. Looking back at (18) with $h = f^n$, we see that the first term in each factor provides a converging factor $\lambda_\rho^{(k-q')n}$. It remains to investigate the other factors, of the form $D^m f^n$ for $m \leq \ell$. We may and do assume that $m = m' + m''$ and that in fact $D^m = D_x^{m'} D_y^{m''}$. The necessary technical but elementary properties are contained in the following

Proposition 4.5. *For $m \geq 1$ and $n \geq 1$, $D^m f^n = D_x^{m'} D_y^{m''} f^n$ satisfies the following properties:*
i) It is a sum of at most $(m-1)! n^{m-1}$ terms;

ii) Each term is a product of at most mn factors of the form $D^i f \circ f^j$*, with* $1 \leq i \leq m$*,* $0 \leq j \leq n-1$*;*
iii) In each term there enter at least $n-m$ *factors with* $i = 1$ *and at most* m *with* $i > 1$*;*
iv) If $m' \geq 1$ *and* $m'' \geq 1$*, there enter in each term at least* $n-m$ *and at most* $m'n$ *(resp.* $m''n$*) factors of the form* $D_x f \circ f^j$ *(resp.* $D_y f \circ f^j$*).*

The first three items are isotropic and constitute Lemma 5.4 of [3]. The first two are proved by a straightforward induction, using the product and chain rules in order to bound respectively the number of terms and the number of factors in each term (we corrected a typo in [3]: The number of terms is indeed bounded by $(m-1)!n^{m-1}$, not just $m!n^m$; in particular it is 1 for $m = 1$. This does not play any role in the sequel). The third assertion is proved by inspection and we prove iv) much in the same way. The statement is symmetric in x and y, and because it is formal we may swap the x and y derivatives in the proof. In other words it is enough to prove it for $D_x^{m'} D_y^{m''} f^n$ (in this order), and the y derivatives.

By ii), there enter at least $n - m''$ terms of the form $D_y f \circ f^j$ in the expression of each term of $D_y^{m''} f^n$, and at most $m''n$ terms that are not of this form, that is involve higher derivatives. One then applies the operator $D_x^{m'}$ to this expression. The number of terms of the form $D_y f \circ f^j$ cannot increase, and in fact by the product rule, an application of D_x to any factor lets the number of such terms decrease by at most 1. So the number of terms $D_y f \circ f^j$ in the end result is at least $n - m' - m'' = n - m$ per factor and is also at most $m''n$. This finishes the proof of iv) and thus of the lemma. □

Returning to the proof of i) in Proposition 4.4 we first note that by iii) of the lemma, the number of terms in each factor involving higher derivatives is at most m. Now for $q > 1$ we know nothing about $D^q f$ and can only use the *a priori* estimate (20). We did not try to distinguish there between x and y derivatives as it does not seem useful since we have in general no information in either direction. We can now estimate $D^m f^n$ over M_ρ by:

$$|D^m f^n| \leq c_\ell n^{m-1} (D_\rho^{(\ell)})^m \lambda_\rho^{n-m} \nu_\rho^{nm''}, \tag{22}$$

an estimate which is valid for any $m \geq 1$ (including if $m'' = 0$). Here we used of course Lemma 5 and have absorbed the factor $(m-1)!$ appearing in i) of that lemma in the combinatorial constant c_ℓ; the term n^{m-1} actually plays no role either, being polynomial in n. The crux of the matter is that $D_\rho^{(\ell)}$ is raised to the power $m \leq \ell$ (independently of n) and that the divergence originating from ν has been controlled in an essentially optimal way. We note that we have *de facto* implemented the Remark following Lemma 5.5 in [2], leading to the perhaps optimal exponents appearing in the inequality connecting k (the order of contact) and ℓ (the regularity of the conjugacy), which here takes the form of the condition $\lambda_\rho^k \nu_\rho^\ell < 1$ occuring in the statement of the proposition (for comparison k here should be shifted to $k+1$ in the notation of [2]).

There remains to return to (17) and (18) and collect estimates. Let us compute the powers of λ_ρ and ν_ρ appearing in the estimate for each term of (18), with $h = f^n$. By

(19) the first factor yields a power $\lambda_\rho^{n(k-q')}$, which originally comes from the high order contact of f_0 and f_1 along M and the contracting character of the maps: This is in fact our only source of convergence here. Then from the product of terms of the form $D^{m_i}f^n$ in (18) one gets a factor $\lambda_\rho^a\nu_\rho^b$, where using (22) we can take: $a=\sum_i(n-m_i)=qn-\ell$ and $b=\sum_i nm_i''=nm''$. Since $q\geq q'$ and using some obvious inequalities we can combine the above into:

$$|D^\ell(F\circ f^n)(z)|\leq c_\ell K_\rho^{(\ell)}|x|^{k-\ell'}(D_\rho^{(\ell)})^\ell n^\ell\lambda_\rho^{-\ell}(\lambda_\rho^k\nu_\rho^{\ell''})^n, \tag{23}$$

which is again valid for $z\in M_\rho$. In order to evaluate $D^\ell G$ (or of course any D^qG with $q\leq\ell$) and show the local uniform convergence of its formal expression, it simply remains to sum over $n\geq 1$. By (23) one gets a geometrically convergent series provided $\lambda_\rho^k\nu_\rho^{\ell''}<1$ and one finds that convergence is determined by a factor involving $\ell''\leq\ell$, that is the number of longitudinal derivatives, which proves both i) and ii) (in the latter case $\ell''=0$). It actually yields somewhat more: In particular the derivative D^qG ($q\leq\ell$) actually vanishes to order $k-q'$ on M, where q' is the number of transverse derivatives. It is also plain from the above that one could devise variants of Proposition 4.4 mixing assertions i) and ii) but we will not go into that. □

This completes our study of the contracting case, in particular of the conjugacy problem on $W=W^s$. Obviously the case of W^u is dealt with by changing f into its inverse, and this will be put to use in order to treat the general hyperbolic case in section 4 below.

4.3 The preparation lemma

A key remark due to S.Sternberg in his original paper ([32]) is that the general hyperbolic case can be reduced to the contracting case. In order to achieve this, one has to replace M by $W=W^s$, viewing the latter as a repulsive invariant submanifold. That is one would like to apply the results of section 2 to the triple (W,V,f_ε^{-1}) instead of (M,V,f_ε) (clearly the roles of W^s and W^u could be switched all along, replacing as usual f by f^{-1}). Note that if M is symplectic W is not (being then actually Lagrangian) which is one of the reasons not to confine oneself to a symplectic M in the contracting case. Now in order to apply the results of the contracting case, one needs the f_ε to have a contact of high order along W, not just along M. In order to achieve this one 'prepares' the original system, that is in our case, performs a preliminary conjugating transform which will ensure a contact of high order of the transformed diffeomorphisms along W. This in turn cannot in general be done globally around W in the analytic setting. But a key point in our case is that we can first take advantage of the fact that the data are analytic and only then perform a cutoff which destroys analyticity. Let us turn to more specific matters; we will be a little less detailed than in section 2 and will leave some routine operations or translations to the good will of the reader. We insist however that no new technical estimates are needed here.

So starting from the assumptions of §1 that the f_ε have a contact of high order ($= k-1$) along M, that their stable manifolds coincide and have moreover been straightened (at least in a neighborhood of M), we should 'prepare' the system further in order to ensure a contact of high enough order along W. The results of the last section applied to $W = W^s$ provide a contact of order 0. This means that if one considers $G = G_\varepsilon$ as in section 2 and the associated vector field $\mathcal{G}_\varepsilon$, then solves equation (5) in §1 and conjugates f_ε by the solution g_ε on W^s, one gets a family which is constant over $M \times B^s_\rho = W_\rho \subset W$, that is over a neighborhood of M in W; here B^s_ρ denotes the ball $\{|x_s| < \rho\}$. We now have to examine the behaviour of the jets transverse to W and the contracting case can be seen as the 0-th order of that procedure. We write as in §1 $z \in V$ with $z = (y, x_s, x_u)$ and introduce $w_s = (y, x_s)$, parametrizing the points of the stable manifold W^s, so that a point of V can also appear as $z = (w_s, x_u)$. Everything here is again local around M and in fact takes place on an infinitesimal neighborhood of W_ρ where $\rho > 0$ will be made precise later and is related to the assumptions made in §2.

Recall that the ultimate goal, to be achieved in the next section, is to solve (E) in a neighborhood of M. Here we will solve it in an infinitesimal neighborhood of W_ρ, actually only to a finite order. Let G be a putative solution and expand it formally around W, writing:

$$G(w_s, x_u) = \sum_{i \geq 0} G_i(w_s) x_u^i. \tag{1}$$

We adopt as usual a simplified system of notation; for instance i is a multiindex, we make no notational distinction between multiindices and their lengths, we could write $x_u^{\otimes i}$ instead of x_u^i etc. As for the diffeomorphisms f_ε we leave out the index ε as usual (and ditto for F, G etc.) and write $f = (f^s, f^u)$ with $f^s \in W^s = W$ and $f^u \in E^u$. We expand these components around W as:

$$f^{s,u}(w_s, x_u) = \sum_{i \geq 0} f_i^{s,u}(w_s) x_u^i. \tag{2}$$

We know that $f_0^u = 0$, that is $f^u(w_s, 0) = 0$ for any $w_s \in W_\rho$ simply because W is invariant under f. We write $f_1^u(w_s) = D_u f^u(w_s, 0) = A_u(w_s)$ where D_u denotes of course the derivatives w.r.t. x_u. This notation extends the one in §1 (*cf.* item 5 there) which was introduced for $z \in M$, that is $x_s = 0$ ($w_s = (y, 0)$).

We write out equation (E'), expanding both sides around W ($x_u = 0$). So we need to expand the composition $G \circ f$ (as well as $F \circ f$), which is no more and no less than the Taylor expansion of a composition of maps. One can of course spell out explicit expressions but we will not actually make use of them. We simply write:

$$G \circ f(z) = \sum_i G_i(f^s(w_s, x_u))(f^u(w_s, x_u))^i = \sum_{i \geq 0} H_i(w_s) x^i. \tag{3}$$

Let us say a word again about notation which may become a little misleading. What we actually want to do is simply solve (E') recursively, order by order. To that end we regroup the terms of the same order, corresponding to multiindices of the same length.

From now on G_i (resp. H_i) with integer i will accordingly denote the operators which correspond to the terms of order i, that is to the multilinear map $D_u^i G$ (resp. $D_u^i(G \circ f)$) (this same shift of notation occurs implicitly in [2], Lemma 5.4). The first two terms read:

$$H_0 = G_0 \circ f_0^s, \qquad H_1 = G_1 \circ f_0^s \cdot f_1^u + ((D_w G_0) \circ f_0^s) \cdot f_1^s, \tag{4}$$

where the argument is w_s and if a quantity ϕ is defined near W we write $\phi(w_s)$ for the restriction $\phi(w_s, 0)$. In particular, in the formula above $f_1^u = A_u$. We expand the perturbation $F \circ f$ in a similar way and for integer $i \geq 0$ we denote by $E^i(w_s)$ the i-th order term of the expansion. A little contemplation yields the following two pieces of information about the H_i and E_i, which are defined in a neighborhood W_ρ of M inside W:

i) H_i can be decomposed as $H_i = G_i \circ f_0^s \cdot A_u^i + H_i'$, where H_i' involves only the G_j for $j < i$;

ii) E_i is analytic and vanishes on M at order $k - i$.

We can now rewrite (E') along W as an infinite system:

$$G_i(w_s) - G_i \circ f_0^s(w_s) \cdot A_u^i = -E_i - H_i' \tag{5}$$

which we wish to solve to a finite order $\ell \leq k$, to be determined below. The important point is that we are now in a position to apply Proposition 4.2 in §1 recursively, with $Q = A_u^i$. We apologize at this point that the matrix Q in that proposition stands to the left of the unknown whereas here it stands to the right. It made the writing easier in both cases but it should be plain that this does not actually alter the statement or proof. Here we fully benefit from the analyticity of the data, which enables us to give a much shorter proof and get a much more effective statement than in the smooth case. The latter does not seem to have been treated in the presence of a nontrivial invariant manifold. Again we encounter the fact that the dynamics on M will not play any role in our present analytic setting as it certainly would in the smooth case, via its Lyapunov exponents. In other words no quantity of the type ν or ν_ρ occurs in this section, as it did in Propositions 3,4 and will again in the next section.

Let us make sure that the assumptions in Proposition 4.2 can be met and that it applies recursively to yield the G_i's for $0 \leq i \leq \ell$. Looking back at the statement we find that we now have $f = f_0^s$ describing the dynamics on the stable manifold near M. We write $D_s f_0^s(w_s) = D_s f^s(w_s, 0) = A_s(w_s)$ (with D_s the derivative w.r.t. x_s) thus again extending the notation $A_s(y)$ of §1 to a neighborhood of M inside W^s ($A_s(y) = A_s(y, 0)$). By assumptions (2) and (3) of §1, $A_s(y, x_s) = D_s f^s(y, x_s, 0)$ and $A_u(y, x_s) = D_u f^u(y, x_s, 0)$ are bounded on M, that is for $x_s = 0$, by $\lambda_s < 1$ and $\mu_u^{-1} \geq 1$ respectively. We can now strengthen these assumptions as in the statement of Proposition 4.2, assuming the existence of $\rho > 0$ such that:

$$|A_s(y, x_s)| \leq \lambda_{s,\rho} < 1, \qquad |A_u(y, x_s)| \leq \mu_{u,\rho}^{-1}, \tag{6}$$

for $w_s = (y, x_s) \in W_\rho^s$, that is simply for $|x_s| < \rho$. Here of course $\mu_{u,\rho}^{-1} \geq \mu_u^{-1} \geq 1$ and $\lambda_s \leq \lambda_{s,\rho} < 1$ and there always exists such a $\rho > 0$ if M is compact. We remark that these assumptions can be expressed either in terms of spectral radiuses or of norms of matrices because even in the not necessarily semisimple case one can adapt the norm on the space so that the norm of the relevant matrix is arbitrarily close to its spectral radius. Finally we can leave the assumptions on the perturbation in Proposition 4.2 as is, that is we assume that there exists $K_\rho > 0$ such that on W_ρ:

$$|F(w_s)| \leq K_\rho |x_s|^k, \tag{7}$$

with $F(w_s) = F(w_s, 0)$. We may now state the following:

Proposition 4.6 (Analytic preparation lemma). *Assume that f and F are defined and analytic over $W_\rho = M \times B^s_\rho$ ($B^s_\rho = \{|x_s| < \rho\} \subset E^s$), that they satisfy (6) and (7) and that $\lambda^{k-\ell}_{s,\rho}\mu^{-\ell}_{u,\rho} < 1$ for an integer $\ell \geq 0$.*

Then the homological equation (E) can be solved uniquely for the jet of order ℓ of G on W_ρ. The solution is analytic on W_ρ and vanishes on M at order k.

In other words there exists a unique $G^{(\ell)}(z)$ which is analytic in $w_s \in W_\rho$ and polynomial of degree ℓ in x_u such that:

$$D^j_u\big(G^{(\ell)} - G^{(\ell)} \circ f + F \circ f\big) = 0 \quad \text{for} \quad 0 \leq j \leq \ell. \tag{8}$$

In shorthand one can write $G^{(\ell)}(z) = \sum_{i=0}^{\ell} G_i(w_s)x^i_u$ where i denotes here again a multiindex and the G_i's we have worked and will be working with regroup these for a given length i of the multiindex.

The proof consists indeed in an iterative application of Proposition 4.2 where at the i-th step we apply it with $k - i$ instead of k, f^s_0 (the restriction of f to W^s) $G = G_i$, $Q = A^i_u$ and the right-hand side $D_i = -E_i - H'_i$ which depends on F and the G_j's for $j < i$. Step 0 is just Proposition 4.2 or actually Proposition 4.1 and yields the initial term $G^{(0)} = G_0(w_s)$. The assumptions are readily seen to carry over by induction, namely G_i and E_i are analytic over W_ρ, they vanish at order $k - i$ on M and E_i satisfies the analog of (7). Finally the inequality $\lambda^{k-i}_{s,\rho}\mu^{-i}_{u,\rho} < 1$ holds true since $i \leq \ell$. One uses essentially the fact that E_i is analytic on W_ρ and vanishes on M at order $k - i$ and that Proposition 4.2 provides a solution which is also analytic over W_ρ and vanishes on M at order k (which is replaced by $k - i$ at step i). The fact that H_i also vanishes at order $k - i$ on M is then purely formal.

If M is compact the existence of $\rho > 0$ such that the inequality $\lambda^{k-\ell}_{s,\rho}\mu^{-\ell}_{u,\rho} < 1$ is satisfied results from the corresponding condition on M itself, that is $\lambda^{k-\ell}_s\mu^{-\ell}_u < 1$. If moreover M is symplectic, one has $\lambda_s = \lambda_u = \lambda$, $\mu_s = \mu_u = \mu$ (*cf.* §1) and this can be rewritten as $\lambda^{k-\ell} < \mu^\ell$. □

Technical Note: We corrected above what seems to be an overly optimistic assertion in [2], Lemma 5.4. One finds there a short sketch of proof in the smooth case with an invariant manifold reduced to a point. However the contact with M at step i is taken to be k and not $k - i$ (see also [13], Lemma 4.1 on that point). This results in an overestimate of ℓ, replacing $k - \ell$ by just k in the defining inequality: $\lambda_{s,\rho}^{k-\ell} \mu_{u,\rho}^{-\ell} < 1$. We also point out that this factor $k - \ell$ is of a quite different nature from the one appearing in the statement of Lemma 5.5, still in [2]. The latter factor can indeed be improved to k as suggested by the remark there, and this is precisely what we did in Proposition 4.4 above. Although these observations may look quite technical they actually reflect rather simple geometric phenomena.

4.4 The general analytic normally hyperbolic case

We now explain how to use the above 'preparation lemma' in order to bring back the general hyperbolic case to the contracting case. This is where we will lose both analyticity and uniqueness by using arbitrary smooth cut-off functions. This cannot be avoided: In the hyperbolic but non contracting case there is in general no (germ of) analytic conjugacy for analytic data. So there is no solution to the conjugacy problem in the analytic category. By contrast one can find solutions which are differentiable to a high order (as we will proceed to demonstrate presently) but they are far from being unique: Germs of hyperbolic analytic diffeomorphisms usually have huge centralizers (*cf.* [10]). Note finally that linearization results requiring diophantine arithmetic conditions in the style of the celebrated Siegel theorem and its variants are not really relevant in our setting, if only because of the presence of an invariant manifold.

In this section we will again favor readability in the sense that we will prove a fairly simple istotropic statement without insisting on explicit estimates. The reader who follows the by now simple proof will immediately perceive that we show actually slightly more than what is stated and that most steps can be made fairly explicit. In this field statements cannot usually be applied as stated and the potential user may find it easier to modify the statement rather than the proof.

The general setting is as in Section 4.1. We start from a family (f_ε) $(0 \leq \varepsilon \leq 1)$ of analytic diffeomorphisms which are obtained from f_0 by a deformation using the Hamiltonian F_ε. The f_ε have a contact of large order $(= k - 1)$ along M; equivalently F_ε vanishes on M at order k. We are looking for a germ of conjugacy g_ε between f_0 and f_ε; g_ε will be defined and differentiable to some order m in a tubular neighborhood of M. Equivalently we are looking for a solution G_ε of the homological equation (E) of class C^m in a neighborhood of M. The diffeomorphism g_1, obtained by integrating (5) of Section 4.1 gives the answer to the original conjugacy problem between f_0 and f_1. From now on we drop again the index ε for the most part; as usual everything will be continuous (and in fact much more) in ε which suffices for our needs. Note also that ε varies over the closed unit interval so that all estimates are *de facto* uniform in ε.

We take up the notation of the last section. For $\rho > 0$ we write $B_\rho^s = \{|x_s| < \rho\} \subset E^s$, $B_\rho^u = \{|x_u| < \rho\} \subset E^u$. For $x = (x_s, x_u) \in E$ we use the norm $|x| = \max(|x_s|, |x_u|)$ for convenience. In particular $B_\rho = \{|x| < \rho\} = B_\rho^s \times B_\rho^u \subset E$. We denote by $W_\rho^s = M \times B_\rho^s \subset W^s$ the local stable manifold and we now write $M_\rho = M \times B_\rho \subset V$ for a tubular neighborhood of M in V. The data and the conclusions are all local near M, that is over M_ρ for some $\rho > 0$. We will not try to really keep track of an explicit value but the reader can check that a (ridiculously small) explicit value could be extracted from the procedure described below.

We write $z = (y, x_s, x_u) = (w_s, x_u)$ and decompose f as $f(z) = (f^i, f^s, f^u) \in M \times E^s \times E^u$. We remark –better late than never– that we do not notationally distinguish strongly stable from stable manifolds (and ditto for unstable) but this should not cause confusion. Here f^i described the M-component, that is the motion along the invariant manifold ($f^i(y, 0, 0) = f(y) \in M$ is the induced dynamics). For $z \in M_\rho$ we define $A_{s,u}(z) = D_{s,u} f^{s,u}(z)$, extending to a neighborhood of M in V the notation of the last section where attention was confined to W^s_ρ. On M ($z = (y, 0, 0)$) this again extends the notation of §1. We reintroduce the quantities $\lambda_{s,u}$ and $\mu_{s,u}$ of §1 and will need to control the spectrum of f^{-1} on M (which is independent of ε). So we introduce ν as in (4) of §1: $|D_y f(y)| < \nu$ and also define $\bar{\nu}$ such that $|D_y f^{-1}(y)| < \bar{\nu}$. Equivalently: $\bar{\nu}^{-1} < |D_y f(y)| < \nu$ for $y \in M$. As usual if M is compact these quantities are known to exist and if not we assume that they do. As usual again we actually assume more, that is that these quantities $\lambda_{s,u}$, $\mu_{s,u}$ and $\nu, \bar{\nu}$ can be continued to a tubular neighborhood M_ρ of M into quantities which we denote as before with an index ρ ($\lambda_{s,\rho}$ etc.) which bound the spectra of $A_{s,u}$ and $D_y f^i$ respectively over M_ρ. Again if M is compact this is not an assumption but just a matter of notation.

Let us now gather together our assumptions on the family f_ε. We assume that there exists $\rho > 0$ such that:

i) All f_ε and F_ε are analytic over M_ρ and indeed can be continued into the complex domain to a strip of constant width in all variables;

ii) The quantities $\lambda_{s,\rho}$, etc. bounding the derivatives of f_ε exist over M_ρ with $\lambda_s, \rho < 1$ and $\lambda_{u,\rho} < 1$;

iii) F_ε satisfies the inequality: $|F_\varepsilon(z)| \leq K_\rho |x|^k$ for $z \in M_\rho$ and a constant $K_\rho > 0$.

We remark that all these assumptions reduce to the existence and analyticity of f_ε and F_ε in a neighborhood of M if the latter is compact. Because of i) Cauchy formula estimates ensure that the higher derivatives of the f_ε and of the perturbation F_ε satisfy the assumptions of Proposition 4.4, that is (19) and (20) of §3. We now state a rough version of our final result:

Theorem C. Assume that f_ε and F_ε satisfy i), ii) and iii) above and that $k > 0$ is large enough. Then the homological equation (E) has a solution G of class C^m in a tubular neighborhood $M_r = M \times B_r$ of the invariant manifold M for some $r > 0$. One can take $m = [ck]$ ($[x]$ is the integral part of x) for some constant $c > 0$; G vanishes at order m on M.

As mentioned above we actually show more than what is stated above and will gather part of that information after the proof. Returning to the original conjugacy problem, the above ensures the existence of a C^m conjugacy between f_0 and f_1 in a tubular neighborhood of M (of constant width), whose jet of order $m - 1$ coincides with that of the identity along M.

As explained at the end of §4.1, because we cannot control what happens globally along the invariant manifolds, we now have to modify the original data using a smooth truncation, thereby giving up analyticity. We first modify the family f_ε so as to make it constant (w.r.t. ε) and linear hyperbolic at infinity. Namely denote by A the dynamics on the tangent bundle of M, that is: $A(z) = (f(y), A_s(y), A_u(y))$ where of course $f(y) = f(y, 0, 0)$ describes the induced dynamics. The diffeomorphism A is analytic and independent of ε as we assume $k > 2$. Let $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ denote a truncating function: $\chi \in C^\infty$ is monotone decreasing, $\chi(r) = 1$ for $r \leq 1$ and $\chi(r) = 0$ for $r \geq 2$ (say). The precise shape of χ is immaterial for our purpose. For $R > 0$, χ_R will denote the rescaled function: $\chi_R(r) = \chi(r/R)$.

We now replace f_ε with $\tilde{f}_\varepsilon$ which interpolates between f_ε and A. Specifically and dropping the subscript ε, we define $\tilde{f} = (\tilde{f}^i, \tilde{f}^s, \tilde{f}^u)$ by:

$$\tilde{f}^i(y, x) = f^i(y, \chi_R(|x|)x); \quad \tilde{f}^{s,u}(z)) = \chi_R(|x|) f^{s,u}(z) + (1 - \chi_R(|x|)) A_{s,u}(y),$$

so that $\tilde{f}_\varepsilon$ coincides with f_ε on M_R and with A outside of M_{2R}. We pick $R < \rho/2$ where ρ is the value occuring in assumptions i), ii), iii) above. This implies that for the new quantities $\tilde{f}$ and $\tilde{F}$ assumptions ii) and iii) still obtain and that indeed the higher derivatives of $\tilde{f}$ and $\tilde{F}$ still satisfy estimates of type (19) and (20) in §3, simply because of their definition and the smoothness of χ. We drop the tildes from now on but remember the crucial fact that the new data coincide with the old ones over M_R, so in particular are analytic there. It should perhaps be noticed at this point that in a practical case one can start from a local situation in some M_ρ and the above can be used in order to extend the situation to the whole of $M \times E$.

We now apply Proposition 4.6 on W^s_R. We note that the values of $\lambda_{s,\rho}$ and $\mu_{s,\rho}$ occuring there refer to W^s_ρ but for simplicity we can *a fortiori* use the values that occur in assumption ii) above (using also that $R < \rho$). This is one of the several places where our present assumptions could be weakened if necessary. So we solve (E) for the jet along W^s_R of order ℓ satisfying:

$$\lambda_{s,\rho}^{k-\ell} \mu_{u,\rho}^{-\ell} < 1, \tag{1}$$

the values of the constants being as in assumption ii) above. We then extend that solution $G^{(\ell)}$ to the whole of $W^s = M \times E^s$ using the functional equation it satisfies, namely (E) (More precisely we are actually extending the $G_i(w_s)$, $i = 0, \ldots, \ell$; *cf.* (1) in §3). Recall that $G^{(\ell)}$ is actually the unique solution vanishing on M; it is also analytic on W^s_R. Finally all f_ε coincide with A outside of M_{2R}, so that F_ε is in fact constant outside that ball and the extension there is completely explicit (this will not be needed in the sequel).

Let us now replace $G^{(\ell)}$ with $G_1(z) = \chi_R(|x_u|) G^{(\ell)}$, that is localize around W^s (this explicit truncation is not really necessary and is more a matter of psychological comfort). The function G_1 solves near W^s an equation of type (E) with a right-hand side F_1 such that the order ℓ jets of F and F_1 coincide on W^s. Consequently the function $G - G_1$ in the strip $W^s \times B^u_R$ solves an equation with right-hand side $F - F_1$ which vanishes at order ℓ on W^s. One can actually say more: Because of the analyticity of the original perturbation and the way it was truncated we are now in a position to apply Proposition 4.4. We should replace there f with f^{-1} (V.Arnold once remarked that the 'stable manifold' derives its name from the fact that it is unstable), M with W^s, k with ℓ. We also set $r = \infty$ and the derivatives of f are bounded to all orders (in other words the assumption of the second statement are satisfied). As for the constants λ_ρ and ν_ρ occuring in the statement of Proposition 4.4, they should now be interpreted as follows. First λ_ρ is to be replaced simply by $\lambda_{u,\rho} < 1$ as in assumption ii) of the Theorem. This can be seen easily, recalling that we patched the original f with its linear part A along M, which has unstable exponents (μ_u, λ_u) in the notation of (3), §1. Second, we should replace ν_ρ with $\mu_{W,s}^{-1} = \max(\mu_{s,\rho}^{-1}, \bar{\nu}_\rho)$. This number indeed controls the expansion for f^{-1} inside the strip $W^s \times B_R^u$. Note that this is the first and last time here where we need to worry about the dynamics on M because we used the analytic preparation lemma, in which it does not appear (nor does it appear in Propositions 1 and 2). The integer m in the statement should thus be small enough so as to satisfy the inequality:

$$\lambda_{u,\rho}^{\ell} \mu_{W,\rho}^{-m} < 1. \tag{2}$$

Under this condition we can apply Proposition 4.4 and find a solution G as in the statement of the theorem. □

Let us finish with some elucidations and complements, first repeating one last time that if M is compact the only assumption becomes the existence and analyticity of the data, say f_0 and the deformation Hamiltonian F_ε in the vicinity of M. Given k, which is defined by the fact that the diffeomorphisms f_ε have a contact of order $k-1$ along M, we get a C^m conjugacy with m controlled by (1) and (2) (concerning the exponent $k-\ell$ in (1), see the technical note at the end of §3). If M is symplectic $\lambda_u = \lambda_s = \lambda$, $\mu_u = \mu_s = \mu \leq \lambda < 1$, $\bar{\nu} = \nu \geq 1$. So the best possible value of m, possibly at the expense of a very small ρ is given by the inequalities: $\lambda^{k-\ell} < \mu^\ell$, $\lambda^\ell < \mu_W^m$. If moreover the dynamics on M is elliptic, that is if $\nu = \bar{\nu} = 1$, one has $\mu_W = \mu$ and one gets the inequalities $\lambda^{k-\ell} < \mu^\ell$, $\lambda^\ell < \mu^m$ (recall that $\mu \leq \lambda < 1$).

Concerning the order of contact along M of the conjugating diffeomorphism with the identity, or what amounts to the same the order of vanishing of the solution G of the homological equation on M, one can say more. It is indeed to be expected that G vanishes at order k but we only proved that it is C^m, vanishing on M at order $m < k$. However we note that we first applied the analytic preparation lemma which yields a solution $G^{(\ell)}$ which is analytic near M and remains so after truncation. Moreover it does vanish on M along W^s at order k, according to Proposition 4.6. We then applied Proposition 4.4, which leaves untouched the restriction of G to W^s. According to the second part of Proposition 4.4, which we may apply with $r = \infty$ as noted above, the solution G is C^∞ with respect to x_u and does vanish to order k on M in that direction. In other words we actually showed that the higher order derivatives $D_s^q G$ and $D_u^q G$ exist for all q and do vanish for $q \leq k$. We did not however study the existence and vanishing of the mixed D_s/D_u transverse derivatives along M of orders between m and k.

4.5 Application to the case at hand

We will now check that the above result does apply to the system which is the subject matter of the present paper. Indeed, after some notational translation Theorem D below will be a direct application of Theorem C above. So we are again interested in the family $\mathcal{F}_q$ of symplectic diffeomorphisms of $\mathbb{A}^2$ introduced in §2. We write $\mathcal{F}_*$ for the unperturbed diffeomorphism (that is when $q = \infty$), again as in that paragraph. Explicitly we have:

$$\mathcal{F}_* = \Phi^{\frac{1}{2}(r_1^2+r_2^2)+\cos 2\pi\theta_1} = \Phi^{\frac{1}{2}r_1^2+\cos 2\pi\theta_1} \times \Phi^{\frac{1}{2}r_2^2}$$

The diffeomorphism $\mathcal{F}_q$ is the perturbation of $\mathcal{F}_*$ explicitly given by (2.3). Implicit is the choice of a width σ as in (2.8) which will play no role whatsoever in what follows. Finally recall that $\mathcal{F}_q$ is $\frac{1}{\sqrt{q}}$-close to $\mathcal{F}_*$ for large q (see (2.9)).

As for the conjugacy problem, we proceed as follows: Fixing q large enough, we regard $\mathcal{F}_q$ as given by the (ε) time-one map of the autonomous (i.e. ε-independent) Hamiltonian $\frac{\varepsilon}{q}f^{(q)}$. In other words we deform along a straight line in the space of Hamiltonian functions. It is then easily checked that the analog of Theorem C above is valid uniformly for q large enough. In fact, in view of the expression of $f^{(q)}$ and especially $f_1^{(q)}$ (f_2 is independent of q; *cf.* (2.4)) we find that (writing z for θ_1) the q-dependence enters through the sequence of functions ($z \mapsto z^\nu$) on a small disc near the origin (see (2.5); $\nu = \nu(q;\sigma)$). It is as well-behaved as can be; in particular for any given N it converges to 0 in the C^N-topology as q increases to infinity. The reader can then easily check that all the constructions of the previous sections are uniform with respect to this added parameter.

The annulus $\mathcal{A}$ which is invariant and normally hyperbolic for $\mathcal{F}_q$ will play the role of M. Recall that $\mathcal{A} = \{O\} \times \mathbb{A} \simeq \mathbb{A}$ where O denotes the hyperbolic fixed point of the pendulum P, with coordinates $\theta_1 = r_1 = 0$. Let $\mathcal{B}_\rho$ denote, for $\rho > 0$, the neighborhood of O in $\mathcal{A}$ defined by: $\mathcal{B}_\rho = \{(\theta_1, r_1) \in \mathbb{A}, |\theta_1| + |r_1| < \rho\}$. We let $\mathcal{A}_\rho = \mathcal{B}_\rho \times \mathbb{A}$ denote the corresponding tubular neighborhood of $\mathcal{A}$ in $\mathbb{A}^2$. Our first and main goal is to prove the following local conjugacy statement:

Theorem D (Local conjugacy around the invariant annulus)**.** For q_0 large enough, there exist ρ, ρ', with $0 < \rho' < \rho$ such that for $q \geq q_0$, $\mathcal{F}_q(\mathcal{A}_{2\rho'}) \subset \mathcal{A}_\rho$ and the following holds:

There exists a C^k diffeomorphism ϕ_q ($k \geq 1$) defined on $\mathcal{A}_\rho$ with $\mathcal{A}_{\rho'} \subset \phi_q(\mathcal{A}_{2\rho'}) \subset \mathcal{A}_\rho$ and such that on $\mathcal{A}_{\rho'}$:

$$\phi_q \circ \mathcal{F}_* = \mathcal{F}_q \circ \phi_q.$$

Moreover there is a constant a ($0 < a < 1$) such that:

$$||\phi_q^{\pm 1} - Id|| \leq a^{\nu(q)},$$

for the C^k norm on $\mathcal{A}_{\rho'}$.

Here and below we abbreviate $\nu(q;\sigma)$ (*cf.* (2.5)) to $\nu(q)$. The above estimates are uniform in q in the sense that the constants ρ, ρ' and a are independent of $q \geq q_0$. The regularity index k can in fact be taken to be large as q goes to infinity and indeed one can ensure $k = k(q) = [c\nu(q)]$ (of order $\mathrm{Log}\, q$) for some constant $c > 0$. This fact, that the local conjugacy actually gets smoother as q approaches infinity will not be needed in the sequel.

Let us now see how Theorem C implies Theorem D. The annulus $\mathcal{A}$ is symplectic and the map induced by $\mathcal{F}_q$ is elliptic, actually an integrable twist so that in the notation of the previous sections we have $\nu = \bar{\nu} = 1$. We also have $\lambda_u = \lambda_s = \lambda$, $\mu_u = \mu_s = \mu$ and there is only one transverse exponent, so that $\lambda = \mu$. Finally this number is nothing but the stable exponent of the time-one map of the pendulum at the point O, namely $\lambda = e^{-2\pi}$, the exact value being however irrelevant for our present purpose. Because $\mathcal{A}$ is not compact we will have to check some uniformity property w.r.t. to the variable r_2 which comes readily from the fact that the perturbation depends on the angles only (see below).

Before we do that however we go on with a few simple reductions. The main parameters in the statement are ρ and a, which are independent of q. Once we have found ρ, ρ' is determined by the condition that $\mathcal{F}_q(\mathcal{A}_{\rho'}) \subset \mathcal{A}_\rho$. Roughly speaking we should take $2\rho' \sim \lambda\rho$ ($\lambda = e^{-2\pi}$ as above), so that for q large enough we may actually set $\rho' = \lambda\rho/4$. We also note that, if $||\phi_q - Id|| \leq a^{\nu(q)}$ the same estimate holds for ϕ_q^{-1}, perhaps at the expense of increasing a slightly. Below we will not deal in detail with domain and inversion problems. They involve as usual the effective application of the classical implicit function theorem and in our context this has been detailed in [27] (see especially §5.2 there).

We first need to dispose of a preliminary step, namely the straightening of the stable and unstable manifolds. We are working locally around $\mathcal{A}$ and uniformly in q for q large enough on a domain $\mathcal{A}_\rho$. We wish to symplectically conjugate $\mathcal{F}_q$ to a diffeomorphism $\overline{\mathcal{F}}_q$ near $\mathcal{A}$ so that the local manifolds $W^{\pm}_{loc}(\mathcal{A}, \overline{\mathcal{F}}_q)$ of $\overline{\mathcal{F}}_q$ coincide with the planes of our coordinate system, thus providing the product structure which is required in the setting of §1 above. In order to achieve this, the main ingredient is the result of [27] (§5) which asserts that the local manifolds $W^{\pm}_{loc}(\mathcal{A}, \mathcal{F}_q)$ can be represented as graphs of analytic functions $v_q^{\pm}$ on suitable domains, after having performed a linear transformation on the first factor in order to let the axes coincide with the eigendirections at the hyperbolic point O. We refer to [27] for details and the proof. Given the existence of this analytic graph parametrization, one constructs a local conjugacy h_q between $\mathcal{F}_q$ and $\overline{\mathcal{F}}_q$, as explained in [23], §1.9. Moreover, Proposition 4.5.1 of [27] provides an estimate of the form: $||v_q^{\pm} - v_0|| \leq c^{\nu(q)}$; here $c < 1$ is a constant and the norm is the sup-norm over a complex domain which contains a thickening of $\mathcal{A}_\rho$ for some $\rho > 0$ ($v_0 = v_0^+ = v_0^-$). As a result we get the estimate: $||h_q^{\pm} - h_0|| < c^{\nu(q)}$ after slightly increasing c if necessary.

We wish to apply Theorem C to the family $\overline{\mathcal{F}}_q$; this will provide us with a local conjugacy $\bar{\phi}_q$ between $\overline{\mathcal{F}}_q$ and $\overline{\mathcal{F}}_*$: $\overline{\mathcal{F}}_q = \bar{\phi}_q \circ \overline{\mathcal{F}}_* \circ \bar{\phi}_q^{-1}$. Given that $\mathcal{F}_q = h_q \circ \overline{\mathcal{F}}_q \circ h_q^{-1}$, we will get ϕ_q as: $\phi_q = h_q \circ \bar{\phi}_q \circ h_0^{-1}$. In view of the exponential estimate of the difference $h_q^{\pm} - h_0$ recalled above, this shows that in order to secure the estimate in the statement of Theorem D, it is enough to get one of the same form for $\bar{\phi}_q$.

We are now reduced to applying Theorem C to $\overline{\mathcal{F}}_q$ for fixed large enough q, considering that it is obtained from $\overline{\mathcal{F}}_*$ by a straight line deformation governed by the Hamiltonian $\frac{1}{q}h_q^*(f^{(q)})$. Looking back at conditions i), ii) and iii) before the statement of Theorem C, we find that i) is clearly fullfilled. Turning to ii), we have seen above that *on* $\mathcal{A}$ the only parameters are $\nu = \bar{\nu} = 1$ and $\lambda = \mu < 1$. In order to extend this to $\mathcal{A}_\rho$, for ρ small enough, into estimates of the needed kind, we must check that the deformation is uniform w.r.t. the variable r_2, that is in the noncompact direction along the invariant manifold $\mathcal{A}$. But this is clear, because firstly the initial deformation $f^{(q)}$ is independent of r_2 and secondly the straightening diffeomorphism h_q is uniformly close to h_0, the latter being in turn independent of r_2 because $\mathcal{F}_*$ has a simple product structure. Finally iii) holds with k tending to infinity as q goes to infinity, with a value of ρ which is uniform in q for q large enough. This yields Theorem D and slightly more; namely first as mentioned above the regularity index k can be made to tend to infinity as q approaches infinity and second if one considers the constant a as a function of the radius ρ, one may in fact pick $a = a(\rho) = \rho^c$ for ρ small enough and a (possibly very small) constant $c > 0$. $\square$

Having obtained a local conjugacy in a tubular neighborhood of the annulus $\mathcal{A}$, we now would like to extend it along the invariant manifolds of $\mathcal{A}$, much as in [27], §5.2. We use the pieces of notation $\mathcal{B}_\rho$ and $\mathcal{A}_\rho$ as above and we choose ρ and ρ' as given by Theorem D. Let Q be the midpoint of the separatix of the pendulum, with coordinates $Q = (r_1, \theta_1) = (2, 1/2)$. We let Q_σ denote the ball with center Q and radius σ (i.e. the tubular neighborhood of Q with thickness σ) in the pendulum plane (or annulus) (θ_1, r_1). Let $\tilde{\mathcal{A}}$ denote the annulus $Q \times \mathbb{A}$ where $\mathbb{A}$ denotes as usual the (θ_2, r_2) annulus. Finally let $\tilde{\mathcal{A}}_\sigma = Q_\sigma \times \mathbb{A}$ denote the tubular neighborhood of $\tilde{\mathcal{A}}$ with thickness σ.

If Φ^P is as usual the time-one map associated with the flow of the pendulum, we let m be a positive integer such that $(\Phi^P)^m(Q) \in \mathcal{B}_{\rho'}$, that is the point Q enters $\mathcal{B}_{\rho'}$ after m iterations (or less; we do not require that m be minimal). By symmetry the same property holds if we iterate backward, that is we also get that $(\Phi^P)^{-m}(Q) \in \mathcal{B}_{\rho'}$. Although the numbers ρ (and ρ') as well as m will be kept fixed in what follows we note that for small ρ the integer m is on the order of $\frac{1}{2\pi}\ln(1/\rho)$. We now remark that for fixed ρ and m we can choose σ small enough so that for q large enough $\mathcal{F}_q^{\pm m}(\tilde{\mathcal{A}}_\sigma) \subset \mathcal{A}_{\rho'}$. Indeed for $q = 0$ this comes simply from the continuity of $\mathcal{F}_*^m$ with respect to the initial conditions and we can extend this to large enough q's because the difference $\mathcal{F}_q - \mathcal{F}_*$ vanishes as q tends to infinity, being actually of order $\frac{1}{\sqrt{q}}$. Again σ will be held fixed in what follows (one may think of σ as being on the order of $\lambda^m\rho$).

Let $n \geq 2m$ be a positive integer, to be thought of as 'large', actually much larger than m. For a point $\varpi \in \mathbb{A}^2$, let us write the following formal equality, of which we will subsequently make sense for ϖ in a certain domain:

$$\mathcal{F}_q^n(\varpi) = \mathcal{F}_q^m \circ \mathcal{F}_q^{n-2m} \circ \mathcal{F}_q^m(\varpi) = \mathcal{F}_q^m \circ \phi_q \circ \mathcal{F}_*^{n-2m} \circ \phi_q^{-1} \circ \mathcal{F}_q^m(\varpi) = \varphi_q \circ \mathcal{F}_*^n \circ \psi_q^{-1}(\varpi), \quad (1)$$

with $\varphi_q = \mathcal{F}_q^m \circ \phi_q \circ \mathcal{F}_*^{-m}$ and $\psi_q = \mathcal{F}_q^{-m} \circ \phi_q \circ \mathcal{F}_*^m$; here ϕ_q denotes of course the local conjugacy whose existence is asserted by Theorem D. The maps φ_q and ψ_q are well-defined C^k diffeomorphisms on $\tilde{\mathcal{A}}_\sigma$, with k as in Theorem D. They are also $\frac{1}{\sqrt{q}}$-close to the identity map, so that we can find σ', with $0 < \sigma' < \sigma$ such that $\tilde{\mathcal{A}}_{\sigma'} \subset \varphi_q(\tilde{\mathcal{A}}_{2\sigma'}) \subset \mathcal{A}_\sigma$ and ditto for ψ_q (see again §5.2 of [27] for quantitative estimates in a similar setting). The following statement is now within easy reach:

Proposition 4.7. *Let m, ρ, ρ', σ and σ' be as above ($0 < \sigma' < \sigma < \rho' < \rho$). For q large enough (i.e. $q \geq q_0$) there exist two C^k diffeomorphisms φ_q and ψ_q (with k as in Theorem D) which are defined on $\tilde{\mathcal{A}}_\sigma$ and are $\frac{1}{\sqrt{q}}$-close to the identity map as q tends to infinity, such that for any integer $n \geq 2m$ and any point $\varpi \in \tilde{\mathcal{A}}_{\sigma'} \cap \mathcal{F}_*^{-n}(\tilde{\mathcal{A}}_{\sigma'})$ the following intertwining relation holds:*

$$\mathcal{F}_q^n \circ \psi_q(\varpi) = \varphi_q \circ \mathcal{F}_*^n(\varpi).$$

Indeed assume that $\varpi \in \tilde{\mathcal{A}}_{\sigma'} \cap \mathcal{F}_*^{-n}(\tilde{\mathcal{A}}_{\sigma'})$, i.e. it is a point whose first projection is very close to Q and whose unperturbed orbit returns there at time n. Equation (1) now makes good sense for ϖ. The main point is that if $\varpi_m = \mathcal{F}_*^m(\varpi)$, we find that $\varpi_m \in \mathcal{A}_\rho$ and $\mathcal{F}_*^{n-2m}(\varpi_m) \in \mathcal{A}_{\rho'}$, the latter property coming from the fact that $\mathcal{F}_*^{n-m}(\varpi) = \mathcal{F}_*^{n-2m}(\varpi_m) \in \mathcal{F}_*^{-m}(\tilde{\mathcal{A}}_{\sigma'}) \subset \mathcal{A}_{\rho'}$. We also note that since m is kept fixed $\mathcal{F}_q^m(\varpi)$ and $\mathcal{F}_*^m(\varpi)$ are $\frac{1}{\sqrt{q}}$-close and we may pick q large enough so that $\mathcal{F}_*^{n-2m} \circ \mathcal{F}_q^m(\varpi) \in \mathcal{A}_\rho$. □

Proposition 4.7 yieds a nice intertwining relation but the intertwining maps are a priori only $\frac{1}{\sqrt{q}}$-close to the identity map. Our final result, which was applied in §2, provides a slighly less natural intertwining relation, but ensures that the intertwing maps are very close to identity. As will be apparent from the proof it is an immediate consequence of Proposition 4.7 supplemented by an easy observation. We first state:

Theorem E. Let m, ρ, ρ', σ and σ' be as above ($0 < \sigma' < \sigma < \rho' < \rho$) and set

$$\chi_q = \mathcal{F}_* \circ \mathcal{F}_q^{m-1} \circ \phi_q \circ \mathcal{F}_*^{-m}, \quad \psi_q = \mathcal{F}_q^{-m} \circ \phi_q \circ \mathcal{F}_*^m.$$

For q large enough (i.e. $q \geq q_0$) χ_q and ψ_q are two C^k diffeomorphisms (with k as in Theorem D) which are defined on $\tilde{\mathcal{A}}_\sigma$ and are c^ν-close to the identity map as q tends to infinity, for some constant c ($0 < c < 1$).

Moreover for any integer $n \geq 2m$ and any point $\varpi \in \tilde{\mathcal{A}}_{\sigma'} \cap \mathcal{F}_*^{-n}(\tilde{\mathcal{A}}_{\sigma'})$ the following intertwining relation holds:

$$\mathcal{F}_* \circ \mathcal{F}_q^{n-1} \circ \psi_q(\varpi) = \chi_q \circ \mathcal{F}_*^n(\varpi).$$

In order to prove this statement, let us analyze the maps φ_q and ψ_q a little more closely, first recalling their respective definitions, namely $\varphi_q = \mathcal{F}_q^m \circ \phi_q \circ \mathcal{F}_*^{-m}$ and $\psi_q = \mathcal{F}_q^{-m} \circ \phi_q \circ \mathcal{F}_*^m$. By Theorem D we have that ϕ_q is actually c^ν-close to identity, where here and below we use the letter c to denote a generic constant satisfying $0 < c < 1$ and $\nu = \nu(q)$. Moreover the difference $\mathcal{F}_q - \mathcal{F}_*$ is of order $\frac{1}{\sqrt{q}}$ on the whole of $\mathbb{A}^2$, but it is in fact of order c^ν outside of $\tilde{\mathcal{A}}_\sigma$ for any fixed $\sigma > 0$.

Upon examining the definition of φ_q and ψ_q we find that the only place where the perturbed map $\mathcal{F}_q$ is applied in a region where the perturbation is significant occurs in the very last factor of φ_q (reading of course from right to left). So we modify this by simply replacing $\mathcal{F}_q$ with $\mathcal{F}_*$ in that factor. This leads to the definition of χ_q and finishes the proof because now in the definition of χ_q and ψ_q the factors $\mathcal{F}_q$ are all applied outside of $\tilde{\mathcal{A}}_\sigma$, which yields an 'exponential' estimate as in the statement.

□

As a final remark, we emphasize that the threshold of validity q_0 in Theorem E (as well as in Proposition 4.7) is independent of n, an important uniformity feature which is actually used in our application of the result.

References

[1] V.I. Arnold: "Instability of dynamical systems with several degrees of freedom", *Dokl. Akad. Nauk SSSR*, Vol. 156, (1964), pp. 9–12; *Soviet Math. Dokl.*, Vol. 5, (1964), pp. 581–585.

[2] A. Banyaga, R. de la Llave and C.E. Wayne: "Cohomology equations near hyperbolic points and geometric versions of Sternberg linearization theorem", *J. Geom. Anal.*, Vol. 6, (1996), pp. 613–649.

[3] A. Banyaga, R. de la Llave and C.E. Wayne: "Cohomology equations and commutators of germs of contact diffeomorphisms", *Trans. AMS*, Vol. 312, (1989), pp. 755–778.

[4] G. Benettin and G. Gallavotti: "Stability of motions near resonances in quasi-integrable Hamiltonian systems", *J. Phys. Stat.*, Vol. 44, (1986), Vol. 293–338.

[5] G. Benettin, L. Galgani and A. Giorgilli: "A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems", *Celestial Mech.*, Vol. 37, (1985), pp. 1–25.

[6] U. Bessi: "An approach to Arnold's diffusion through the calculus of variations", *Nonlinear Anal. TMA*, Vol. 26, (1996), pp. 1115–1135.

[7] U. Bessi: "Arnold's example with three rotators", *Nonlinearity*, Vol. 10, (1997), pp. 763–781.

[8] J. Bourgain: "Diffusion for Hamiltonian perturbations of integrable systems in high dimensions", (2003), preprint.

[9] J. Bourgain and V. Kaloshin: "Diffusion for Hamiltonian perturbations of integrable systems in high dimensions", (2004), preprint.

[10] M. Chaperon: "Géométrie différentielle et singularités de systèmes dynamiques", *Astérisque*, Vol. 138-139, (1986).

[11] M. Chaperon: "Invariant manifolds revisited", *Tr. mat. inst. Steklova*, Vol. 236, (2002), Differ. Uravn. i Din. Sist., pp. 428–446.

[12] M. Chaperon: "Stable manifolds and the Perron-Irwin method", *Erg. Th. and Dyn. Syst.*, Vol. 24, (2004), pp. 1359–1394.

[13] M. Chaperon and F. Coudray: "Invariant manifolds, conjugacies and blow-up", *Erg. Th. and Dyn. Syst.*, Vol. 17, (1997), pp. 783–791.

[14] B.V. Chirikov: "A universal instability of many-dimensional oscillator systems", *Phys. Reports*, Vol. 52, (1979), pp. 263–379.

[15] R. Douady: "Stabilité ou instabilité des points fixes elliptiques", *Ann. Sc. Éc. Norm. Sup.*, Vol. 21, (1988), pp. 1–46.

[16] R. Easton and R. McGehee: "Homoclinic phenomena for orbits doubly asymptotic to an invariant three-sphere", *Ind. Univ. Math. Journ.*, Vol. 28(2), (1979).

[17] E. Fontich and P. Mart´ in: "Arnold diffusion in perturbations of analytic integrable Hamiltonian systems", *Discrete and Continuous Dyn. Syst.*, Vol. 7(1), (2001), pp. 61–84.

[18] G. Gallavotti: "Quasi-integrable mechanical systems", In: K. Osterwalder and R. Stora (Eds.): *Phénomènes critiques, systèmes aléatoires, théories de jauge*, part II (Les Houches 1984), North-Holland, Amsterdam New York, 1986, pp. 539–624.

[19] M. Herman: *Notes de travail*, December, 1999, manuscript.

[20] M.W. Hirsch, C.C. Pugh and M. Shub: *Invariant Manifolds*, Lecture Notes in Mathematics, Vol. 583, Springer Verlag, 1977.

[21] S. Kuksin and J. Pöschel: "Nekhoroshev estimates for quasi-convex Hamiltonian systems", *Math. Z.*, Vol. 213, (1993), pp. 187–216.

[22] P. Lochak: "Canonical perturbation theory via simultaneous approximation", *Usp. Mat. Nauk.*, Vol. 47, (1992), pp. 59–140; *Russian Math. Surveys*, Vol. 47, (1992), pp. 57–133.

[23] P. Lochak, J.-P. Marco and D. Sauzin: "On the splitting of invariant manifolds in multi-dimensional near-integrable Hamiltonian systems", *Memoirs of the Amer. Math .Soc.*, Vol. 163, (2003).

[24] P. Lochak and A.I. Neishtadt: "Estimates in the theorem of N. N. Nekhorocheff for systems with a quasi-convex Hamiltonian", *Chaos*, Vol. 2, (1992), pp. 495–499.

[25] P. Lochak, A.I. Neishtadt and L. Niederman: "Stability of nearly integrable convex Hamiltonian systems over exponentially long times", In: *Proc. 1991 Euler Institute Conf. on Dynamical Systems*, Birkhaüser, Boston, 1993.

[26] J.-P. Marco: "Transition le long des chaines de tores invariants pour les systèmes hamiltoniens analytiques", *Ann. Inst. H. Poincaré*, Vol. 64(2), (1996), pp. 205–252.

[27] J.-P. Marco: "Uniform lower bounds of the splitting for analytic symplectic systems", *Ann. Inst. Fourier*, submitted to.

[28] J.-P. Marco and D. Sauzin: "Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems", *Publ. Math. I.H.E.S.*, Vol. 96, (2003), pp. 77.

[29] N.N. Nekhoroshev: "An exponential estimate of the time of stability of nearly integrable Hamiltonian systems", *Usp. Mat. Nauk.*, Vol. 32, (1977), pp. 5–66; *Russian Math. Surveys*, Vol. 32, (1977), pp. 1–65.

[30] J. Pöschel: "Nekhoroshev estimates for quasi-convex Hamiltonian systems", *Math. Z.*, Vol. 213, (1993), pp. 187–216.

[31] S. Sternberg: "Local contractions and a theorem of Poincaré", *Amer. J. Math.*, Vol. 79, (1957), pp. 809–824.

[32] S. Sternberg: "On the structure of local homeomorphisms II", *Amer. J. Math.*, Vol. 80, (1958), pp. 623–631.

[33] S. Sternberg: "On the structure of local homeomorphisms III", *Amer. J. Math.* , Vol. 81, (1959), pp. 578–604.

[34] S. Sternberg: "Infinite Lie groups and the formal aspects of dynamical systems", *J. Math. Mech.*, Vol. 10, (1961), pp. 451–474.

[35] D.V. Treshchev: "Evolution of slow variables in near-integrable Hamiltonian systems", *Progress in nonlinear science*, Vol. 1, (2001), pp. 166–169.