University of Bristol - Explore Bristol Research
General rights

Fast drift and diffusion in a class of isochronous systems with the windows method[*]

Alessandro Fortunati[§]

School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom.

Abstract

The aim of this paper is deal with some peculiar difficulties arising from the use of the geometrical tool known as *windows method* in the context of the well known problem of Arnold's diffusion for isochronous nearly-integrable Hamiltonian systems.

Despite the simple features of the class of systems at hand, it is possible to show how the absence of an anisochrony term leads to several substantial differences in the application of the method, requiring some additional devices, such as non-equally spaced transition chains and variable windows.

As a consequence, we show the existence of a set of unstable orbits, whose drifting time matches, up to a constant, the one obtained via variational methods.

Keywords: Hamiltonian systems, Arnold's Diffusion, Instability, Splitting of invariant manifolds, Transition chains.

2010 MSC: Primary: 37J40. Secondary: 37C29, 37C50, 70H08.

1 Introduction

Arnold's Diffusion [Arn64] is a topological instability phenomenon arising in nearly integrable Hamiltonian systems with more than two degrees of freedom. It consists of the existence of a class of motions whose effect is to produce a drift of order one of the slow variables, over a "very long" time T_d (i.e. tending to infinity as the perturbation size μ approaches to zero, for all sufficiently small μ).

A quite relevant problem concerns the estimate of the minimal time in order to observe this phenomenon. The earlier literature on this topic, such as [CG94], [Bes96] etc., has shown that the attempts to give an estimate of T_d crucially depend on the method used to construct such a class of trajectories, rather than the features of the system at hand, as it would be natural to expect. This question is addressed in [Gal97] and in [Gal99], by taking as a paradigmatic example the following simple *a priori-unstable* isochronous system

$$H(p,q,\underline{A},\underline{\varphi}) = \underline{\omega} \cdot \underline{A} + \frac{p^2}{2} + (\cos q - 1) + \mu F(q,\underline{\varphi}), \tag{1}$$

with $(q,\underline{\varphi},p,\underline{A}) \in \mathcal{M} := T^*(\mathbb{S}^1 \times \mathbb{T}^2)$, where T^* denotes the cotangent bundle. Some relevant subsequent works on this class of systems (with some various generalizations) extended this gap further, as summarized in the (non-exhaustive) Table 1.

In the much more investigated anisochronous case, related to the original *Arnold's example*, see [Arn64], the situation was a little bit more encouraging. In particular, the logarithmic estimate finally given by the variational approach of [BBB03], was also obtained in [CG03] via the so-called Windows Method

---

[§]E-mail: alessandro.fortunati@bristol.ac.uk

Work	T_d	approach
[Gal97]	super-exponential	geometrical
[Gal99]	exponential (i.e. $O(\exp(\mu^{-1}))$)	geometrical
[Cre01]	polynomial (i.e. $O(\mu^{-\gamma})$)	geometrical
[BB02]	logarithmic (i.e. $O(\mu^{-1}\log\mu^{-1})$)	variational

Table 1: γ is a positive $O(1)$ constant.

(WM). This powerful geometrical tool (originally due to Alekseev and Easton, see for instance [Eas81] and references therein), was reconsidered and developed in [Mar96] by analysing a simplified version of Arnold's example. Nowadays, the WM has been deeply improved leading to some advanced results on the problem of Arnold's diffusion, see for instance [GZ04], [GR07] and [Sab15].

The motivation of this work arises from the fact that a treatment of the isochronous case cannot be obtained neither from [Mar96] nor from [CG03] by taking as zero the anisochrony coefficient. The main reason can be briefly described as follows. As done in [Cre03], it is possible to show the existence of a conjugacy between the dynamics in a neighbourhood of a Graff torus[1] and a symbolic dynamics on a suitable alphabet and then, in particular, one can conclude the existence of periodic orbits. Following [CG03], one can construct a chain of (hyperbolic) periodic orbits $\mathcal{O}_k$, surrounding a transition chain[2] of partially hyperbolic tori. Arnold's diffusion is obtained via WM as a shadowing of the chain $\mathcal{O}_k$.

It is clear that the above described argument cannot be applied in the isochronous case as this class of systems does not admit periodic orbits: the (Diophantine) frequencies of the rotators are fixed and then the trajectories are open for all times, densely filling the underlying torus $\mathbb{T}^2$.

The impossibility of adapting the construction described in [CG03], is motivated by a deep obstruction, which arises in the isochronous case: the absence of an anisochrony term inhibits a remarkable geometrical phenomenon pointed out in [CG08] and called "transversality-torsion". The latter is related to a suitable compressing-stretching action of the phase flow on the windows, giving a key condition for a shadowing result (called *correct alignment*) in a quite natural way. This cannot be done in our case and the correct alignment condition requires a more careful treatment.

In order to overcome this difficulty, a suitable sequence of "self deforming" (i.e. non-constant) windows has been considered, referring to this mechanism as *simulated torsion*. Such a sequence has been determined by solving a one dimensional discrete dynamical system over the transition chain.

The more restrictive (technical) estimates arising in the isochronous case, also obstruct the use of an equally spaced transition chain[3]. For this purpose, we have used the *elastic chain* tool, introduced for the first time in [Gal99] and consisting of a step-by-step variation of an equally spaced chain. This tool, has the remarkable feature to allow a suitable variation the hetero/homoclinic point between the whiskers of the tori of the transition chain, see Sec. 4, in such a way to bypass the ergodization time[4].

Despite the use of the elastic chains leads in [Gal99] to a remarkable improvement of the speed of drift with respect to [Gal97], see Table 1, it turns out to be not enough in order to get a drift time as fast as the one obtained in [BB02]. This is due to some peculiar differences between the geometrical method initially used in [CG94] and the WM[5]. In this paper we show how a combination of the elastic chains tool with the WM leads to the "logarithmic" T_d found in [BB02].

---

[1]See cited reference for its definition.

[2]See Def. 2.5.

[3]Namely, a transition chain in which the distance in the actions space between two consecutive tori is constant.

[4]i.e. the time necessary to approach a given point on the flat torus within a prefixed distance.

[5]For a review of the differences between the geometrical approaches of [CG94] and of [Eas81] and how these are able to improve the speed of diffusion, we refer to [For12].

The features of the model (1) are, to date, well known in the literature (see e.g. the works listed in Table 1), especially in presence of the choice (2) for the perturbation. Nevertheless, it exhibits all the relevant difficulties arising in the study of a "fast" diffusion phenomenon by means of geometrical methods. On the other hand, the particular structure of the Hamiltonian allows several technical simplification of some standard (but very specialized) background topics discussed in Sections 2 and 3.

2 Preliminary facts and main result

2.1 System at hand, invariant tori and whiskers intersection

Let us consider the system (1) where the perturbing function is supposed of the form $F(q, \underline{\varphi}) = (1 - \cos q)f(\underline{\varphi})$, with

$$f(\underline{\varphi}) = \sum_{\substack{\underline{k} \in \mathbb{Z}^2 \\ |k_1| + |k_2| \leq \Gamma}} f_{\underline{k}} \cos(\underline{k} \cdot \underline{\varphi}), \tag{2}$$

Γ is a fixed positive constant, and $f_{\underline{k}} = f_{-\underline{k}} > 0$ for all $\underline{k}$. In particular f is an even trigonometric polynomial. We suppose $\underline{\omega}$ to be a $(C, \tau)-$Diophantine vector[6].

First of all note that the unperturbed system possesses a two parameter continuous family of invariant tori

$$\mathcal{T}(\underline{A}) = \{(p, q, \underline{A}, \underline{\varphi}) : p = q = 0, \quad \underline{\varphi} \in \mathbb{T}^2\},$$

with $\underline{A} \in \mathbb{R}^2$. Due to the term $1 - \cos q$ in the perturbing function, the entire family survives also in the perturbed system. This is of course one of the main advantages of this simple example. If one requires that two tori with different actions $\underline{B}, \underline{C}$ lie on the same energy level, these have to satisfy $\underline{\omega} \cdot (\underline{B} - \underline{C}) = 0$. Hence we have

$$\underline{B} = \underline{C} + \delta \underline{\omega}^{\perp}, \tag{3}$$

for some δ. So they form a one parameter continuous family.

These tori are *whiskered* in the sense that each of them admits two invariant manifolds $W^{s,u}(\mathcal{T}(\underline{A}))$ called *whiskers*, such that, by denoting by Φ^t the phase flow given by (1), $\lim_{t \to \infty} \Phi^t(\underline{z}) \in \mathcal{T}(\underline{A})$ for all $\underline{z} \in W^s(\mathcal{T}(\underline{A}))$ and $\lim_{t \to -\infty} \Phi^t(\underline{z}) \in \mathcal{T}(\underline{A})$ for all $\underline{z} \in W^u(\mathcal{T}(\underline{A}))$.

As can be easily seen by examining the equations of motion, the (positive branches of the) whiskers $W^u(\mathcal{T}(\underline{B}))$ and $W^s(\mathcal{T}(\underline{C}))$ possess the following structure in the unperturbed system

$$
\begin{aligned}
W^u(\mathcal{T}(\underline{B})) &= \{(q, \underline{\varphi}, p, \underline{A}) : p_0(q) := \sqrt{2(1 - \cos q)}, \quad \underline{A} = \underline{B}, \quad (q, \underline{\varphi}) \in (0, 2\pi) \times \mathbb{T}^2\}, \\
W^s(\mathcal{T}(\underline{C})) &= \{(q, \underline{\varphi}, p, \underline{A}) : p_0(q) := \sqrt{2(1 - \cos q)}, \quad \underline{A} = \underline{C}, \quad (q, \underline{\varphi}) \in (0, 2\pi) \times \mathbb{T}^2\}.
\end{aligned}
$$

It is clear that these whiskers intersect each other if, and only if, $\delta = 0$ and moreover the intersection is flat, meaning that the whiskers tangent spaces coincide at every point. Note that the whiskers possess the structure (trivial in the $\underline{A}$) of graphs over the angles.

By continuous dependence on the parameters, the whiskers will keep the graph structure also in the perturbed case, for sufficiently small μ. So, for all positive $O(1)$ constants d, the whiskers can be written as

$$
\begin{aligned}
W^u(\mathcal{T}(\underline{B})) &= \{(q, \underline{\varphi}, p_0(q) + \mu p^u(q, \underline{\varphi}, \mu), \quad \underline{B} + \mu \underline{A}^u(q, \underline{\varphi}, \mu)), (q, \underline{\varphi}) \in (0, 2\pi - d) \times \mathbb{T}^2\}, \\
W^s(\mathcal{T}(\underline{C})) &= \{(q, \underline{\varphi}, p_0(q) + \mu p^s(q, \underline{\varphi}, \mu), \quad \underline{C} + \mu \underline{A}^s(q, \underline{\varphi}, \mu)), (q, \underline{\varphi}) \in (d, 2\pi) \times \mathbb{T}^2\},
\end{aligned} \tag{4}
$$

---

[6]There exist constants $C \neq 0$ and $\tau \geq 1$ such that $|\underline{\omega} \cdot \underline{k}| \geq C|\underline{k}|^{-\tau}$.

see e.g. [Gal94], also for their recursive construction to all orders.

We introduce the following notation

$$\Delta p := p^u(q, \underline{\varphi}, \mu) - p^s(q, \underline{\varphi}, \mu), \qquad \Delta \underline{A} := \underline{A}^s(q, \underline{\varphi}, \mu) - \underline{A}^u(q, \underline{\varphi}, \mu),$$

then

$$\underline{F}(q, \underline{\varphi}, \delta) := \delta \underline{\omega}^\perp + \mu \Delta \underline{A}(q, \underline{\varphi}, \mu), \tag{5}$$

whose zeros clearly correspond to homo/heteroclinic intersections ($\delta = 0$ and $\delta > 0$, respectively) of the whiskers $W^u(\mathcal{T}(\underline{B}))$ and $W^s(\mathcal{T}(\underline{C}))$. The standard analysis on $\underline{F}$, see e.g. [GGM99], uses the remarkable relation between the first order term $\Delta \underline{A}(q, \underline{\varphi}, 0)$ and the *Melknikov integral*, by virtue of its dominance (for a wide class of models, including this one) for all sufficiently small μ. The solutions of $\underline{F}(q, \underline{\varphi}, 0) = \underline{0}$ (i.e. homoclinic points) are extended in a suitable neighbourhood of $\delta = 0$ via a standard Implicit Function argument, uniformly in μ.

Hence, define

$$M := \mathcal{D}\underline{F} \equiv \frac{\partial(\Delta p, \Delta \underline{A})}{\partial(q, \underline{\varphi})}.$$

It is an easy consequence of the isoenergy constraint between $\mathcal{T}(\underline{B})$ and $\mathcal{T}(\underline{C})$ that if there exists a point $(q^*, \underline{\varphi}^*) \in (d, 2\pi - d) \times \mathbb{T}^2$ such that $\underline{F}(q^*, \underline{\varphi}^*, \delta) = \underline{0}$ and $\operatorname{rank} M = 2$, then[7] $W^u(\mathcal{T}(\underline{B})) \pitchfork W^s(\mathcal{T}(\underline{C}))$ at $(q^*, \underline{\varphi}^*)$. The condition $\operatorname{rank} M = 2$ is maximal as from the conservation of energy (see also [Val98]), one gets

$$\begin{aligned}
\partial_q \Delta p &= \alpha \underline{\omega} \cdot \partial_q \Delta \underline{A}, \\
\partial_{\varphi_k} \Delta p &= \alpha \underline{\omega} \cdot \partial_{\varphi_k} \Delta \underline{A}, \qquad k = 1, 2,
\end{aligned} \tag{6}$$

with $\alpha := -1/(p_0(q^*) + \mu p^s(q^*, \underline{\varphi}^*, \mu))$, so $\operatorname{rank} M < 3$.

In this case we shall speak, as usual, of a transversal homoclinic point if $\delta = 0$ (i.e. $\underline{B} = \underline{C}$) and of a transversal heteroclinic point otherwise.

Another key ingredient is the symplectic character of the whiskers[8]. Using this property one can state the following

Proposition 2.1. *Let us consider the 2×2 submatrix of M defined by $M_{se} := \partial_{\underline{\varphi}} \Delta \underline{A}$. If $\det M_{se}(q^*, \underline{\varphi}^*) > 0$ holds for some $(q^*, \underline{\varphi}^*)$, then*

$$\partial_q \Delta p \neq 0; \qquad \det \begin{pmatrix} \partial_q \Delta A_1 & \partial_{\varphi_1} \Delta A_1 \\ \partial_q \Delta A_2 & \partial_{\varphi_1} \Delta A_2 \end{pmatrix} (q^*, \underline{\varphi}^*) \neq 0. \tag{7}$$

Proof. (Sketch) Due to their Lagrangian nature, the whiskers are gradient of two respective (family of) generating functions $S^{u,s}(q, \underline{\varphi}, \mu)$, so M is nothing but the Hessian matrix of $S^u(q, \underline{\varphi}, \mu) - S^s(q, \underline{\varphi}, \mu)$, computed at the intersection point. This implies that M is a symmetric matrix. By solving the system given by conditions (6) with respect to the entries of M_{se} and imposing that $\det M_{se} > 0$ the statement easily follows. $\square$

Now we are going to assume the following typical hypothesis

Condition 2.2 (splitting)**.** The system (1) possesses a transversal homoclinic point at[9] $q = \pi$. In other terms there exists a point $(q, \underline{\varphi}) = (\pi, \underline{\varphi}^0)$ such that $\underline{F}(\pi, \underline{\varphi}^0) = \underline{0}$ and $\det M_{se}(\pi, \underline{\varphi}^0) > 0$ with $(M_{se})_{ij} = O(1)$.

---

[7] $\pitchfork$ denotes the transversal intersection.

[8] See [LMS03] for a comprehensive treatment of this topic and a different algorithmical approach for the whiskers building to all orders in μ.

[9] The existence of a transversal homoclinic point is a standard result for this system; see e.g. [GGM99] for the same result in a more general context. The property $\det M_{se} > 0$ (instead of $\det M_{se} \neq 0$) is a feature of this system and is related to the parity of $f(\varphi)$. The latter holds for sufficiently small μ and can be directly checked on the Melnikov integral (first order approximation of $\Delta \underline{A}$), see [LMS03] for its general expression.

4

Remark 2.3. Clearly $(\pi, \underline{\varphi}^0)$ is a particular homoclinic point. As the whiskers intersection takes place along an entire orbit, all the points of the form $(q, \underline{\varphi}^0 + \underline{\omega} T_{q \to \pi})$, where $T_{q \to \pi}$ is the time to evolve from π to q, are homoclinic as well. If q is $O(1)$ bounded away from the origin, M_{se} can be well approximated via Melnikov integral. It can be easily seen that M_{se} does not change (up to $O(\mu^2)$) if different sections for q are chosen.

The section $q = \pi$ is useful in the presence of symmetries of the perturbing function as in this case. It is possible to show (see e.g. [GGM99]) that, by parity, $(q, \underline{\varphi}^0) = (\pi, \underline{0})$ is a transversal homoclinic point to all orders in μ. We denote as $\underline{\varphi}^0 =: (\varphi_1^0, \varphi_2^0)$.

In the current setting, we state the following

Lemma 2.4. *Let us assume the splitting condition, and define the section*

$$\Sigma := \{\underline{\varphi} \in \mathbb{T}^2 : \varphi_2 = \varphi_2^0\}. \tag{8}$$

There exist $\bar{\delta} = O(\mu)$ and two functions q, φ_1 defined on $(-\bar{\delta}, \bar{\delta})$, with $q(0) = \pi$ and $\varphi_1(0) = \varphi_1^0$, such that, for all $\underline{B}$ and sufficiently small μ

$$W^u(\mathcal{T}(\underline{B})) \pitchfork W^s(\mathcal{T}(\underline{B} + \delta\underline{\omega}^\perp)),$$

on Σ at the point $(q(\delta), \varphi_1(\delta))$. In particular

$$\frac{d}{d\delta}\varphi_1(\delta) \neq 0, \qquad \frac{d}{d\delta}\varphi_1(\delta), \frac{d}{d\delta}q(\delta) = O(\mu^{-1}), \tag{9}$$

for all $\delta \in (-\bar{\delta}, \bar{\delta})$.

The just stated version of this classical splitting result is a sort of counterpart of the statement which can be found in [Gal99]. According to the latter, as δ varies, the $\underline{\varphi}$ coordinates of the intersection point move along a curve on $\mathbb{T}^2$, while q is fixed to π. In our case q moves, but φ_2 is fixed, allowing us to use a single section on $\mathbb{T}^2$ for all the values of δ and then to perform an advantageous dimensional reduction.

Proof. By hypothesis $\underline{F}(\pi, \underline{\varphi}^0, 0) = \underline{0}$ and rank $M_{\text{se}} = 2$, so by the Implicit Function Theorem, there exist $\bar{\delta}_1$ and two functions $q(\delta), \varphi_1(\delta)$ such that $(q(0), \varphi_1(0)) = (\pi, \varphi_1^0)$ and

$$\underline{F}(q(\delta), \varphi_1(\delta), \varphi_2^0, \delta) = \underline{0}, \qquad \forall \delta \in (-\bar{\delta}_1, \bar{\delta}_1),$$

so $(q(\delta), \varphi_1(\delta), \varphi_2^0)$ is a transversal hetero/homoclinic point. Note that, as all the entries of M are $O(\mu)$, then $\bar{\delta}$ is $O(\mu)$.

Defining $D := -[\partial_q \Delta A_1 \partial_{\varphi_1} \Delta A_2 - \partial_q \Delta A_2 \partial_{\varphi_1} \Delta A_1]^{-1}$, the Implicit Function Theorem immediately yields

$$\frac{d(q, \varphi_1)}{d\delta}(\delta) = \mu^{-1} \begin{pmatrix} \partial_q \Delta A_1 & \partial_{\varphi_1} \Delta A_1 \\ \partial_q \Delta A_2 & \partial_{\varphi_1} \Delta A_2 \end{pmatrix}^{-1} \begin{pmatrix} \partial_\delta F_1 \\ \partial_\delta F_2 \end{pmatrix} = (\alpha\mu)^{-1} D \begin{pmatrix} -\partial_{\varphi_1} \Delta p \\ \partial_q \Delta p \end{pmatrix},$$

where we have used the second of (7) and formulae (6). The second of (9) is a direct consequence. By the first of (7) and by continuity, there exists $\bar{\delta}_2 > 0$ such that $\partial_q \Delta p \neq 0$ for all $\delta \in (-\bar{\delta}_2, \bar{\delta}_2)$. By choosing $\bar{\delta} := \min\{\bar{\delta}_1, \bar{\delta}_2\}$, the first of (9) immediately follows. $\square$

5

2.2 Transition chains

Definition 2.5. A transition chain of length N for (1) is a (finite) sequence of tori $\{\mathcal{T}_k\}_{k=1,\dots,N}$, where $\mathcal{T}_k := \mathcal{T}(\underline{A}^k)$, such that

$$W^u(\mathcal{T}_k) \pitchfork W^s(\mathcal{T}_{k+1}), \qquad \forall k = 1, \dots, N-1.$$

By (3) the sequence $\{\underline{A}^k\}$ is necessarily of the form $\underline{A}(y_k)$, $y_k \in \mathbb{R}$, where

$$\underline{A}(y) := \underline{A}^0 + \underline{\omega}^{\perp} y. \tag{10}$$

Set $\delta^k := y^{k+1} - y^k$, by Lemma 2.4 every sequence $\{\mathcal{T}_k\}$ such that $|\delta^k| \leq \bar{\delta}$ for all $k = 1, \dots, N-1$ is a transition chain.

Definition 2.6. We call **equally spaced chain** (ESC) a transition chain $\{\mathcal{T}_k\}$ such that

$$\delta^k \in \{-\hat{\delta}, 0, \hat{\delta}\}$$

for all $k = 1, \dots, N-1$ and for some fixed $\hat{\delta} \leq (3/4)\bar{\delta}$.

Now consider the sequence of intervals

$$E_k := (y^k - C_1 \mu, y^k + C_1 \mu),$$

with C_1 suitably chosen in a way that $\mathcal{T}(\underline{A}(\tilde{y}^k))$ is still a transition chain[10] for all $\tilde{y}^k \in E_k$ and for all k.

Definition 2.7. Let $\mathcal{T}_k^*$ be an ESC and let E_k^* be its associated sequence of intervals. An **equivalent elastic chain** (EEC) with the given ESC, is a sequence $\mathcal{T}_k$ such that $y_k \in E_k^*$ for all k.

In this setting we have the following

Theorem 2.8. *Let $N = O(\mu^{-1})$. For all ESC of length N and sufficiently small μ, there exists an EEC of the same length and a set of initial data $\mathcal{U}$ such that the solution of (1) starting in $\mathcal{U}$ shadows[11] the EEC with a time*

$$T = O\left(\frac{1}{\mu} \ln \frac{1}{\mu}\right). \tag{11}$$

As an immediate consequence we get

Corollary 2.9 (Arnold's Diffusion)**.** *The Arnold's Diffusion takes place in the system (1) on a set of initial data $\mathcal{U}_d \subset \mathcal{U}$ with a time given by (11)*

Proof. Choose the EEC given by $\delta^k := \bar{\delta}$ (or $-\bar{\delta}$) for all k and apply Theorem 2.8. As $\delta_k = O(\mu)$, an $O(\mu^{-1})$ number of transitions leads to a drift of $O(1)$. $\qquad\square$

Remark 2.10. As the set of all ESC of length N has cardinality equal to 3^N, the drift Theorem 2.8 shows that the system (1) possesses a rich behaviour. In particular, every step on the transition chain is conjugate to a random walk on a ternary tree. Arnold's diffusion is simply associated to the two lateral branches of the entire tree.

---

[10]It is sufficient to choose $C_1 \leq \bar{\delta}/(16\mu)$.

[11]We do not need here a precise definition of the concept of shadowing as it will be intrinsically given by using windows.

3 Flow approximation and windows

This section is devoted to fix the background material in order to prove Theorem 2.8. First of all, we are going to construct a suitable approximation (in the "linear map + remainder" form) of the Hamiltonian flow close to a generic sequence of hetero/homoclinic points associated with a given transition chain. This is obtained in a standard way, that is, by a composition of two maps: the first one describes the "inner" motion close to the invariant tori and the second one the "outer" motion from one torus to another. As a difference from [Mar96] and related works, we use Gallavotti's normal form instead of the Graff-Treshev one, giving a different criterion to determine the Jacobian matrix of the outer map. We also recall some basic notion of windows and their main properties for the readers convenience. For a more efficient comparison with our result, we shall follow the notations of [Mar96] and of [CG03] very closely from now on.

3.1 Inner dynamics: normal form

Let $\mathcal{T}(\underline{A}^*)$ be an invariant torus for the system (1). Fixed $\tilde{R}, \tilde{\kappa}$, let us consider the phase space region defined as

$$\tilde{U}_{\tilde{R}, \tilde{\kappa}}(\underline{A}^*) := \{(q, \underline{\varphi}, p, \underline{A}) : |p|, |q| < \tilde{R}, \quad |\underline{A} - \underline{A}^*| < \mu\tilde{\kappa}, \quad \underline{\varphi} \in \mathbb{T}^2\},$$

that can be regarded as a "neighbourhood" of $\mathcal{T}(\underline{A}^*)$.

Theorem 3.1 (Gallavotti, 1997). *For all $\underline{A}^*$ and μ small enough, there exist $\tilde{R}, \tilde{\kappa}, R, \kappa$ independent on μ and a canonical map*

$$\mathcal{C} : (q, \underline{\varphi}, p, \underline{A}) \in \tilde{U}_{\tilde{R}, \tilde{\kappa}}(\underline{A}^*) \to (Q, \underline{\psi}, P, \underline{I}) \in U_{R, \kappa}(\underline{I}^*),$$

with

$$U_{R, \kappa}(\underline{I}^*) := \{(Q, \underline{\psi}, P, \underline{I}) : |P|, |Q| < R, \quad |\underline{I} - \underline{I}^*| < \mu\kappa, \quad \underline{\psi} \in \mathbb{T}^2\},$$

and $(0, \underline{\psi}, 0, \underline{I}^*) = \mathcal{C}(0, \underline{\varphi}, 0, \underline{A}^*)$, casting the Hamiltonian (1) in the following form[12]

$$H(Q, \underline{\psi}, P, \underline{I}) = \underline{\omega} \cdot \underline{I} + J_\mu(\eta), \tag{12}$$

with $\eta := PQ$ and where $J_\mu(\eta)$ is an analytic function on $U_{R, \kappa} \times (0, \mu_0)$ for some $\mu_0 > 0$.

In the new set of variables, the system is clearly integrable and its evolution is given by

$$(Q(t), P(t), \underline{\psi}(t), \underline{I}(t)) = (Q(t_0)e^{g(\eta)t}, P(t_0)e^{-g(\eta)t}, \underline{\psi}(t_0) + \underline{\omega}(t - t_0), \underline{I}(t_0)), \tag{13}$$

where $g(\eta) := \frac{d}{d\eta} J_\mu(\eta)$ and $g(0) =: g$.
The proof of this result can be found in [Gal97], and goes along the lines of the (more general) proof for the anisochronous case contained in [CG94]. In order to clarify some considerations we shall need later, let us recall that the canonical map $\mathcal{C}$ is the composition of two maps: the Jacobi map $\mathcal{C}_J$

$$\begin{cases} q_J &= S(q, p) \\ p_J &= R(q, p) \end{cases} \tag{14}$$

(straightening the invariant manifolds of the unperturbed pendulum) and a " KAM " map $\mathcal{C}_K$ (μ close to the identity)

$$\begin{cases} q_J &= Q + \mu L(Q, \underline{\psi}, P, \mu) \\ \underline{\varphi} &= \underline{\psi} \\ p_J &= P + \mu M(Q, \underline{\psi}, P, \mu) \\ \underline{A} &= \underline{I} + \mu \underline{N}(Q, \underline{\psi}, P, \mu) \end{cases} \tag{15}$$

---

[12]This result holds with a more general perturbing function, in particular, it does not require the presence of the term $1 - \cos q$.

whose effect is to remove the perturbation to all orders in μ.

3.2   Reduced system of coordinates

In the neighbourhood $U_{R,\kappa}(\underline{I}^*)$, it is useful to define a reduced system of coordinates. This standard approach is systematically used in such cases (see, e.g. in [Eas81] or [Mar96]) and it can be obtained as follows

1. Fix a Poincaré section on $\mathbb{T}^2$: due to isochrony, every straight line on $\mathbb{T}^2$ whose direction is not aligned with $\underline{\omega}$, is a transversal section. The simplest choice is

$$\Sigma_N := \{\underline{\psi} \in \mathbb{T}^2 : \quad \psi_2 = \varphi_2^0\},$$

for a complete interface with the "external" section Σ defined in (8). In such a way we get a discretization of the solutions, and the sampling time t^* is constant because of isochrony. Without loss of generality, we shall suppose t^* to be equal to 1.
As $\underline{\omega}$ is a Diophantine vector, the recurrence frequency on Σ_N, denoted as ν, will be a $(C,\tau)-$ Diophantine number.

2. Set an energy level $\mathcal{M}(h) := \{(P,Q,\underline{\psi},\underline{I}) : H = h\}$ and suppress the action I_2 by restricting the flow to $\mathcal{M}(h)$.

By defining $\theta := \varphi_1$ and $\rho := I_1$, the coordinate system of the restriction to $V^h(\rho^*) := U_{R,\kappa}(\underline{I}^*) \cap \Sigma_N \cap \mathcal{M}(h)$ is simply (Q, θ, P, ρ). By taking into account (13), the images of the section map $f : V^h(\rho^*) \to V^h(\rho^*)$, can be easily written as

$$\underline{f}^n(Q_0, \theta_0, P_0, \rho_0) := (Q_0 e^{ng(\eta)}, \theta_0 + n\nu, P_0 e^{-ng(\eta)}, \rho_0), \tag{16}$$

where $\eta = Q_0 P_0$.

3.3   Sequences of hetero/homoclinic points associated to ESCs and EECs

Let us fix $\underline{A}^0 \in \mathbb{R}^2$, and denote as $\mathcal{M}(h) = \{H = h\}$ the hypersurface given by the energy level $h = \underline{\omega} \cdot \underline{A}^0$. Given $N \in \mathbb{N}$ and a sequence $\{y_k\}_{k=1,\dots,N} \in \mathbb{R}$, by (10), $\mathcal{T}_k := \mathcal{T}(\underline{A}(y_k))$ is a sequence of invariant tori on $\mathcal{M}(h)$ for the system (1).
By Lemma 2.4 we can associate with an ESC $\mathcal{T}_k$, a sequence of hetero/homoclinic points $(q^k, \varphi_1^k) := (q(\delta_k), \varphi_1(\delta_k))$ on Σ. We denote by Φ^l, $l \in \mathbb{Z}$, the sampling of the phase flow on Σ.
Let k be arbitrarily chosen and let

$$\underline{z}_k := (q^k, \varphi_1^k, \varphi_2^0, p_0(q^k) + \mu p^u(q^k, \varphi_1^k, \varphi_2^0, \mu), \underline{A}(y_k) + \mu \underline{A}^u(q^k, \varphi_1^k, \varphi_2^0, \mu)).$$

Recalling the notations introduced in section 3.1, by Theorem 3.1 a "neighbourhood" of each torus $\mathcal{T}_k$ can be endowed by a normal coordinates system. Hence, for all k, it is possible to define $\tilde{U}_k := \tilde{U}_{\tilde{R},\tilde{\kappa}}(\underline{A}(y_k))$ and $U_k := U_{R,\kappa}(\underline{I}_k)$, with $\underline{I}_k := \underline{A}(y_k)$, such that for all k, $\mathcal{C} : \tilde{U}_k \to U_k$.
Subsequently, as described in the previous section, we can construct a reduced system of coordinates (Q, θ, P, ρ) in $V_k := U_k \cap \Sigma_N \cap \mathcal{M}(h)$.
As $\lim_{l \to -\infty} \Phi^l(\underline{z}_k) = \mathcal{T}_k$ and $\lim_{l \to +\infty} \Phi^l(\underline{z}_k) = \mathcal{T}_{k+1}$, for a sufficiently large $|l|$ we get $\Phi^{-l}(\underline{z}_k) \in \tilde{U}_k$ and $\Phi^l(\underline{z}_k) \in \tilde{U}_{k+1}$. In this way, these points can be written in the respective normal coordinates of the tori. In particular, there exist $l_k^{\pm} \in \mathbb{N}$ such that[13]

$$(\mathcal{C} \circ \Phi^{-l_k^-}(\underline{z}_k))_Q < R/2, \qquad (\mathcal{C} \circ \Phi^{l_k^+}(\underline{z}_k))_P < R/2.$$

---

[13]We denote e.g. with $(\underline{v})_{v_1}$ the v_1 component of $\underline{v}$.

In the reduced system these points reads as

$$\begin{aligned}
\underline{X}_k = (Q_k, \theta_k, 0, \rho_k) &:= (\mathcal{C} \circ \Phi^{l_k^-}(\underline{z}_k))|_{V_k} \\
\underline{X}'_k = (0, \theta'_k, P'_k, \rho_{k+1}) &:= (\mathcal{C} \circ \Phi^{l_k^+}(\underline{z}_k))|_{V_{k+1}}
\end{aligned} . \tag{17}$$

As for the map (16), note that if we write

$$e^{\pm ng(\eta)} = e^{\pm ng(0)}\left[1 + \left(\pm n\eta\tilde{g}(x) + \frac{1}{2}(n\eta\tilde{g}(\eta))^2 \pm \dots\right)\right] =: L^{\pm n}[1 + F_{\pm}(\eta)],$$

the map (16) can be expanded in a neighbourhood of $\underline{X}_k$ in the following way

$$\underline{f}^n(\underline{X}_k + \underline{\xi}) = (L^n(Q_k + \xi_1), \theta_k + \xi_2 + n\nu, L^{-n}\xi_3, \rho_k + \xi_4) + \underline{R}(\underline{X}_k, \underline{\xi}), \tag{18}$$

with

$$\underline{R}(\underline{X}_k, \underline{\xi}) = (L^n(Q_k + \xi_1)F_-(\eta), 0, L^{-n}\xi_3 F_+(\eta), 0), \tag{19}$$

where in this case $\eta = (Q_k + \xi_1)\xi_3$. We also denote by $\hat{\underline{f}}^n$ the linear part $\underline{f}^n - \underline{R}$.

Let us now consider an ESC. For all the EECs associated with the given ESC, by following the previous construction we get a double sequence of hetero/homoclinic points $\{X_k(\tilde{y}_k), X'_k(\tilde{y}_k)\}_{k=1,\dots,N}$ for all $\tilde{y}_k \in E_k$. If necessary, we could reduce the range of variation $\bar{\delta}$ (i.e. C_1) in such a way $Q_k(y), P'_k(y) < (3/4)R$ for all $y \in E_k$.

The key point is that, due to isochrony, by varying y in E_k (by an $O(\mu)$), we can move the coordinates of the hetero/homoclinic point (of an $O(1)$, see the second of (9)) on $\mathbb{T}^2$ without ever falling out of Σ_N, i.e. by some functions

$$\theta_k(y) := \varphi_1(\delta) - l_k^- \nu, \qquad \theta'_k(y) := \varphi_1(\delta) + l_k^+ \nu, \tag{20}$$

where $\delta := y_{k+1} - y$ with y_{k+1} fixed in E_{k+1}.

This is the advantage in our formulation of Lemma 2.4 with respect to the setting used in [Gal99]. As Σ_N is one dimensional the problem of finding y such that e.g. $\theta_k(y)$ is equal to some given θ^* in the image of $\theta_k(y)$ is achievable in a constructive way[14].

As a direct consequence of the normal form, the manifolds $W^u(\mathcal{T}_k)$ and $W^s(\mathcal{T}_{k+1})$ in V_k and V_{k+1} respectively, are given by

$$W^u(\mathcal{T}_k) = \{(Q, \theta, P, \rho) : P = 0, \ \rho = \rho_k\}, \qquad W^s(\mathcal{T}_{k+1}) = \{(Q, \theta, P, \rho) : Q = 0, \ \rho = \rho_{k+1}\}.$$

As $l_k^- = O(1)$ (see Remark 2.3), by writing down the expression of the whisker $W^s(\mathcal{T}_{k+1})$ in terms of normal coordinates[15], it is possible to recover (up to higher orders) two smooth functions $P_{k+1}(Q, \theta; y)$ and $\rho_{k+1}(Q, \theta; y)$ such that the connected component of $W^s(\mathcal{T}_{k+1})$ intersecting $W^u(\mathcal{T}_k)$ at $\underline{X}_k(y)$ in V_k is parameterized by

$$\underline{Y}_{k+1}(Q, \theta; y) := \{(Q, \theta, P_{k+1}(Q, \theta; y), \rho_{k+1}(Q, \theta; y))\}, \qquad \underline{Y}_{k+1}(Q_k, \theta_k; y) = \underline{X}_k(y). \tag{21}$$

Furthermore, the normal splitting matrix

$$M_N := \begin{pmatrix} \partial_Q P_{k+1} & \partial_\theta P_{k+1} \\ \partial_Q \rho_{k+1} & \partial_\theta \rho_{k+1} \end{pmatrix}(Q_k, \theta_k; y) \tag{22}$$

satisfies[16]

$$(M_N)_{11} \neq 0, \qquad \text{rank } M_N = 2, \tag{23}$$

by virtue of (7) for all $y \in E_k$ (if necessary we could reduce C_1 further). As a geometrical consequence, we have that $W^s(\mathcal{T}_{k+1})$ is still a graph over the angles in the normal (reduced) coordinates system.

---

[14]Compare with comments in [Gal99], pag. 307 and 309].

[15]So backward evolved by a time l_k^-.

[16]An equivalent result holds by fixing the section $Q = R/2$, and can be found in [Val98].

3.4 Outer dynamics: transition map

As we have denoted by $\underline{f}^n$ the evolution in V_k, we want to give, here, an approximation of the Hamiltonian flow from V_k to V_{k+1}. For this purpose, let us denote with $\underline{\Psi} : V_k \to V_{k+1}$ the Hamiltonian flow, implicitly defined by

$$\underline{\Psi}(\underline{X}_k) = \underline{X}'_k. \tag{24}$$

By regular dependence on initial data, there exist neighbourhoods $\mathcal{D}_k$ of $\underline{X}_k$ and $\mathcal{D}'_k$ of $\underline{X}'_k$ respectively, such that $\underline{\Psi}$ maps $\mathcal{D}_k$ into $\mathcal{D}'_k$ and, for all $\underline{X}_k + \underline{\xi} \in \mathcal{D}_k$ it can be written in the form

$$\underline{\Psi}(\underline{X}_k + \underline{\xi}) = \underline{X}'_k + A\underline{\xi} + \underline{\tilde{R}}(\underline{X}_k, \underline{\xi}). \tag{25}$$

We have denoted by $A := \mathcal{D}\underline{\Psi}(\underline{X}_k)$ the Jacobian matrix of the map computed at the hetero/homoclinic point and by $\underline{\tilde{R}}(\underline{X}_k, \underline{\xi})$ the remainder that in terms of components reads as

$$\tilde{R}^i(\underline{X}_k, \underline{\xi}) = \int_0^1 (1-t)\underline{\xi}\mathcal{H}\Psi^i(\underline{X}_k + t\underline{\xi}) \cdot \underline{\xi}^T dt, \qquad i = 1, \ldots, 4, \tag{26}$$

where $\mathcal{H}$ denotes the Hessian matrix operator. We denote by $\underline{\hat{\Psi}}$ the linear part $\underline{\Psi} - \underline{\tilde{R}}$.

The map $\underline{\Psi}$, also known as *separatrix map*, see e.g. [Tre02], is a standard and widely used tool to construct trajectories close to invariant manifolds. The approach we are going to use, in order to describe the structure of A, is of geometrical nature and it seems to be very close to the one used in [Tre02]. The matrix A is essential for a non-trivial approximation of the flow, however it is not difficult to understand that this matrix conceals a complicated structure in our case. For instance, let us denote with Φ'_k the Hamiltonian flow (in global coordinates) mapping $\mathcal{D}_k$ in $\mathcal{D}'_k$ and by $\mathcal{C}_R^k$ and $\mathcal{C}_R^{k+1}$ the canonical maps $\mathcal{C}$ with the coordinates reduction to V_k and V_{k+1}, respectively. So, by construction we have

$$\underline{\Psi}(Q, \theta, P, \rho) = (\mathcal{C}_R^{k+1} \circ \Phi'_k \circ (\mathcal{C}_R^k)^{-1})(Q, \theta, P, \rho).$$

As one can easily deduce by (14) and (15), this computation turns out to be a very difficult task[17], even at the first order in μ, unless one would like to resort to a numerical approach.

We state the following

Proposition 3.2. *The matrix A for the system (1) takes the form*

$$A = \begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 1 & 0 & 0 \\ -1/a & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \mu \begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 0 \end{pmatrix}, \tag{27}$$

with $a \neq 0$, and[18] $a_{ij} := a_{ij}^k(\mu)$. These satisfies, for sufficiently small μ

$$a_{11} \neq 0, \qquad D := a_{11}a_{42} - a_{41}a_{12} \neq 0. \tag{28}$$

Note that the above structure holds for all orders in μ. We denote with A_0 the unperturbed part of A.

---

[17]The main difficulty arises if one write down the perturbing function in terms of Jacobi's coordinates and then attempt to perform a perturbative step.

[18]In spite of the dependence also on k (as A is the Jacobian matrix computed at $\underline{X}_k$), we suppress the index k to avoid a cumbersome notation.

Proof. If $\mu = 0$ the system is integrable, then rotators and pendulum are uncoupled. Moreover, the canonical map $\mathcal{C}$ reduces to the Jacobi map for the pendulum, whose Hamiltonian takes simply the form $J_0(PQ)$.

Let Π denote the projection operator on the (P, Q) plane. By the conservation of energy, for $(P, Q) \in \Pi V_k$ and $(P', Q') \in \Pi V_{k+1}$ we get $J_0(PQ) = J_0(P'Q')$. Now consider a small increment of $(Q_k, 0)$ of the form $(Q_k + \sigma, \delta)$ by denoting by $(f(\sigma, \delta), P'_k + g(\sigma, \delta))$ the corresponding variations in ΠV_{k+1}. It is understood that $f(\sigma, \delta)$ and $g(\sigma, \delta)$ are regular functions by definition of $\underline{\Psi}$ and such that $f(0, 0) = g(0, 0) = 0$. Hence, up to higher orders in (σ, δ), we have $f(\sigma, \delta) = u_{11}\sigma + u_{12}\delta$ and $g(\sigma, \delta) = u_{21}\sigma + u_{22}\delta$ where u_{ij} are the entries of $\frac{\partial(f,g)}{\partial(\sigma,\delta)}(0, 0)$. As a consequence, the conservation of energy yields, up to h.o.t., $\sigma\delta = (u_{11}\sigma + u_{12}\delta)(u_{21}\sigma + u_{22}\delta)$, which is equivalent to

$$u_{11}u_{21} = 0, \quad u_{12}u_{22} = 0, \quad u_{11}u_{22} + u_{21}u_{12} = 1.$$

Now recall that the intersection between ΠW_k^u and ΠW_{k+1}^s is flat, then increments in the direction $(Q, 0)$ in ΠV_k imply displacements in the direction $(0, -P)$ in ΠV_{k+1}, from which it follows that $u_{21} < 0$. So, by the previous equalities we get $u_{11} = 0$, so $u_{12} = 1/u_{21}$ and then $u_{22} = 0$. It is sufficient to define $a := -u_{21}$ and note that the Jacobian matrix in the (θ, ρ) variables is trivially the identity matrix, to get the required form of the unperturbed A.

Now suppose $\mu \neq 0$. By definition, if $\underline{\Psi}$ is restricted to the graphs of the functions $P_{k+1}(Q, \theta), \rho_{k+1}(Q, \theta)$ (parameterizing W_{k+1}^s in V_k), its image has to produce points on W_{k+1}^s in V_{k+1} and then such that $Q = 0$ and $\rho = \rho_{k+1}$. In other terms, for sufficiently small $\bar{r} > 0$, we get for all $(\sigma, \varepsilon) \in \mathcal{B}_{\bar{r}}(\underline{0})$

$$\begin{cases} \Psi_1(Q_k + \sigma, \theta_k + \varepsilon, P_{k+1}(Q_k + \sigma, \theta_k + \varepsilon), \rho_{k+1}(Q_k + \sigma, \theta_k + \varepsilon)) &= 0 \\ \Psi_4(Q_k + \sigma, \theta_k + \varepsilon, P_{k+1}(Q_k + \sigma, \theta_k + \varepsilon), \rho_{k+1}(Q_k + \sigma, \theta_k + \varepsilon)) &= \rho_{k+1} \end{cases}. \tag{29}$$

By expanding the first equation in a neighbourhood of $(\sigma, \varepsilon) = \underline{0}$, one has

$$\begin{cases} a_{11} + a_{13}\frac{\partial P_{k+1}}{\partial Q} + a_{14}\frac{\partial \rho_{k+1}}{\partial Q} &= 0 \\ a_{12} + a_{13}\frac{\partial P_{k+1}}{\partial \theta} + a_{14}\frac{\partial \rho_{k+1}}{\partial \theta} &= 0 \end{cases}, \tag{30}$$

where we have recognized the elements of A. By the unperturbed form of A we know $a_{1j} = \mu a_{1j}^1 + o(\mu)$ for $j = 1, 2, 4$ while $a_{13} = a + \mu a_{13}^1 + o(\mu)$. On the other hand, the derivatives appearing in the previous equation are exactly those of the splitting matrix M_N (22). Then, by denoting by

$$\mu \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} := \begin{pmatrix} \frac{\partial P_{k+1}}{\partial Q} & \frac{\partial P_{k+1}}{\partial \theta} \\ \frac{\partial \rho_{k+1}}{\partial Q} & \frac{\partial \rho_{k+1}}{\partial \theta} \end{pmatrix},$$

and equating the powers of μ in (30) we get $a_{11}^1 = -ah_{11}$ and $a_{12}^1 = -ah_{21}$. First of all we note that $a_{11} \neq 0$ for $\mu \neq 0$ by the first of (23). In a similar way, the second equation of (29) yields $a_{41}^1 = -h_{21}$ and $a_{42}^1 = -h_{22}$. Then, by keeping in mind the definition of D, it follows that

$$D = a(h_{11}h_{22} - h_{12}h_{21}),$$

that is, up to higher orders, the determinant of M_N: the latter is non-zero by the second of (23).

Note that the argument is until now quite general, and can be reasonably extended to more general systems.

Now we attempt to get some additional information by using the very simple form of our system, in order to simplify, as much as possible, the structure of the perturbed part of A.

Let us consider two matrices $B_1, B_2 \in \mathbb{M}(n, m)$ for some $n, m \in \mathbb{N}$. We say that B_1 and B_2 possess the *same structure*, and we denote this property with $B_1 \approx B_2$, if B_1 and B_2 have null entries exactly in the

same position(s).

So, by (14) and (15) the Jacobian matrices of the canonical maps satisfy

$$
D\mathcal{C}_J \equiv \begin{pmatrix} S_q & 0 & 0 & S_p & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ R_q & 0 & 0 & R_p & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad D\mathcal{C}_K \equiv \mathbb{I} + \mu \begin{pmatrix} L_Q & L_{\psi_1} & L_{\psi_2} & L_P & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ M_Q & M_{\psi_1} & M_{\psi_2} & M_P & 0 & 0 \\ N_Q^1 & N_{\psi_1}^1 & N_{\psi_2}^1 & N_P^1 & 0 & 0 \\ N_Q^2 & N_{\psi_1}^2 & N_{\psi_2}^2 & N_P^2 & 0 & 0 \end{pmatrix},
$$

where $\mathbb{I}$ is the identity matrix. Both matrices are computed, by definition, at the hetero/homoclinic point and the entries of the second one are functions of μ.

First of all it is clear that $D\mathcal{C}_J \approx D\mathcal{C}_J^{-1}$. Furthermore, it is easy to check from (15) that $D\mathcal{C}_K \approx D\mathcal{C}_K^{-1}$. Now, simply by looking at the equations of motion, we can see that the pendulum evolution does not depend on the action variables to all orders in μ. This implies that the actions evolution (also due to the independence of the perturbing function on $\underline{A}$) is independent of $\underline{A}$ at $O(\mu^k)$ for all $k \geq 1$. So the Jacobian matrix of the flow satisfies

$$
D\Phi^t \approx \begin{pmatrix} \times & 0 & 0 & \times & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \times & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} + \mu \begin{pmatrix} \times & \times & \times & \times & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 \end{pmatrix},
$$

where the entries of the second matrix are functions of μ.

In conclusion, by definition, we have that $A := D\mathcal{C}_K \cdot D\mathcal{C}_J \cdot D\Phi^t \cdot D\mathcal{C}_J^{-1} \cdot D\mathcal{C}_K^{-1} \approx D\Phi^t$. By reducing the set of coordinates, we get the perturbed part of (27) . $\qquad\square$

Remark 3.3. Clearly, the previous argument does not guarantee that the entries of A marked by $\times$ are non-zero since some of them could vanish due to some symmetry of the product. As we shall see, the only key properties we shall need are given by (28): it does not matter if the other entries vanish or not. In this sense we agree with the topological criterion used in [Mar96, p. 239] to give the structure of his matrix A.

3.5 Windows

Definition 3.4. Let $M \subset \mathbb{R}^d$, and $d_h, d_v \in \mathbb{N}$ such that $d_h + d_v = d$. Let $L_h := [-1, 1]^{d_h}$, $L_v := [-1, 1]^{d_v}$ and $\mathcal{Q} := L_h \times L_v$.
A **window** is a C^1 diffeomorphism

$$
\underline{W} : \mathcal{Q} \to M.
$$

We denote by $\tilde{W}$ the image[19] of $\mathcal{Q}$ through $\underline{W}$.

The *horizontals* of $\underline{W}$ are defined as the images of $\underline{W}(x_h, x_v)$ where $x_v \in L_v$ is kept fixed and x_h varies in L_h. Similarly, the *verticals* of $\underline{W}$ are defined as the images of $\underline{W}(x_h, x_v)$ with $x_h \in L_h$ fixed and variable $x_v \in L_v$.

The point $\underline{W}(\underline{0})$ is called the *centre* of the window. We shall refer to *non-degenerate* windows as windows containing a non-zero volume. Even if a non-necessary condition, it is natural to think of

---

[19]We shall refer as window either the map or its image.

integers d_h and d_v as satisfying $d_h = d_v = d/2$. Actually, in applications, d is the (even) dimension of the (eventually reduced) phase space.

An interesting case is represented by *affine windows*. This class of windows arises when the function $\underline{W}$ is linear, and then can be written in the form

$$\underline{W}^a(\underline{x}) = \underline{c} + W\underline{x}.$$

The point $\underline{c}$ is the centre and W is a $d \times d$ matrix we shall call the *representative matrix* of $\underline{W}^a$. Clearly, a window is non-degenerate if $\det W \neq 0$.

Definition 3.5. We say that a window $\underline{W}_1$ is **correctly aligned on** (c.a.o.) $\underline{W}_2$ if for all $y_v \in L_v$ and for all $y_h \in L_h$ there exists unique $x_h \in (-1, 1)^{d_h}$ and $x_v \in (-1, 1)^{d_v}$ such that

$$\underline{W}_1(x_h, y_v) = \underline{W}_2(y_h, x_v), \tag{31}$$

and the intersection is transversal.

The notion of correct alignment leads to the following key result

Theorem 3.6 (Shadowing, Easton)**.** *Let $\{\mathcal{D}_k\}_{k=1,\dots,N}$ be a collection of open sets of $\mathbb{R}^d$, $\Psi_k : \mathcal{D}_k \to \mathcal{D}_{k+1}$ a family of diffeomorphisms and $\underline{W}_k$ a family of windows $\underline{W}_k : \mathcal{Q} \to \mathcal{D}_k$. Let, for all k, $\underline{W}'_k := \Psi_k \circ \underline{W}_k$.*
If $\underline{W}'_k$ is c.a.o. $\underline{W}_{k+1}$ for all $k = 1, \dots, N-1$ there exists (at least) a point $\underline{x}^*$ of $\tilde{W}_1 \subset \mathcal{D}_1$ such that $\Psi_{N-1} \circ \dots \circ \Psi_1(\underline{x}^*) \in \tilde{W}_N \subset \mathcal{D}_N$.

For the proof, given in a more general context, we refer to [EM79].

3.6   A correct alignment criterion

The following statement, due to J.P. Marco, gives us a set of sufficient conditions in order to establish if a window is correctly aligned on another one. First of all, some notations:

Let $\mathbb{M}(n, m) := GL(\mathbb{R}, n \times m)$ be the linear group of the $n \times m$ matrices with real entries. For all $\underline{v} \in \mathbb{R}^d$ let $\|\underline{v}\|_\infty := \max_{j=1,\dots,d} |v_j|$. Consequently, for all $A \in \mathbb{M}(n, m)$ we denote $\|A\|_\infty := \max_{i=1,\dots,n} \sum_{j=1}^m |a_{ij}|$.

Now, let $\underline{F}(\underline{x}) \in C^k(\mathcal{U}, \mathbb{R}^n)$, $k \geq 1$. The standard C^1 norm over $\mathcal{U}$ is denoted as

$$\|\underline{F}(\underline{x})\|_{C^1(\mathcal{U})} := \max\{\sup_{\underline{x} \in \mathcal{U}} \|\underline{F}(\underline{x})\|_\infty, \sup_{\underline{x} \in \mathcal{U}} \|D\underline{F}(\underline{x})\|_\infty\}. \tag{32}$$

Lemma 3.7 (Marco, 1996)**.** *Let*

$$\underline{W}_1^a = \underline{c}_1 + W_1\underline{x}, \qquad \underline{W}_2^a = \underline{c}_2 + W_2\underline{x},$$

be two affine windows, where $\underline{c}_{1,2} \in \mathbb{R}^d$, and $W_{1,2} \in \mathbb{M}(d, d)$ are of the form

$$W_i = \begin{pmatrix} W_i^1 & W_i^3 \\ W_i^2 & W_i^4 \end{pmatrix},$$

with $W_i^1 \in \mathbb{M}(d_h, d_h)$, $W_i^2 \in \mathbb{M}(d_h, d_v)$, $W_i^3 \in \mathbb{M}(d_v, d_h)$ and $W_i^4 \in \mathbb{M}(d_v, d_v)$.
Now define the "intermediary matrices"

$$M = \begin{pmatrix} W_1^1 & -W_2^3 \\ W_1^2 & -W_2^4 \end{pmatrix}, \qquad N = \begin{pmatrix} -W_2^1 & W_1^3 \\ -W_2^2 & W_1^4 \end{pmatrix},$$

13

and

$$\chi^a := \sup_{\underline{x} \in \mathcal{Q}} \| M^{-1} N \underline{x} + M^{-1}(\underline{c}_2 - \underline{c}_1) \|_\infty. \tag{33}$$

Suppose that M is invertible and $\chi^a < 1$. Then $\underline{W}_1^a$ c.a.o. $\underline{W}_2^a$.
Now consider the following windows

$$\underline{W}_1(\underline{x}) = \underline{W}_1^a(\underline{x}) + \hat{\underline{W}}_1(\underline{x}), \qquad \underline{W}_2(\underline{x}) = \underline{W}_2^a(\underline{x}) + \hat{\underline{W}}_2(\underline{x}), \tag{34}$$

where the maps $\hat{\underline{W}}_i(\underline{x}) \in C^1(\mathcal{U}, \mathbb{R}^d)$, and $\mathcal{U}$ is an open set containing $\mathcal{Q}$. And define

$$\chi^c := \| M^{-1}(\hat{\underline{W}}_2(\underline{x}) - \hat{\underline{W}}_1(\underline{x})) \|_{C^1(\mathcal{U})}.$$

Suppose that

c1. $\chi^a < 1$, *implying that $\underline{W}_1^a$ c.a.o. $\underline{W}_2^a$,*

c2.

$$\chi^c < \frac{1}{4}, \qquad \chi^a + \frac{\chi^c}{1 - \chi^c} < 1. \tag{35}$$

Then $\underline{W}_1$ c.a.o. $\underline{W}_2$.

As in [Mar96] it is natural to choose $\mathcal{U} := (-2, 2)^d$. A detailed proof of Lemma 3.7 can be found in the already mentioned paper or, more closely to the above formulation, in [LM05].

3.7 A suitable class of windows

Our aim is now to construct a prototype of non-degenerate affine windows for the problem at hand. Following [Eas81] we choose a class of windows satisfying the following

Condition 3.8.   1. The horizontals of $\underline{B}(\underline{x})$ are parallel to $T_{\underline{X}_k} W_k^u$,

    2. The verticals of $A\underline{B}(\underline{x})$ are parallel to[20] $T_{\underline{X}_k'} W_{k+1}^s$.

Before restricting ourselves to the particular form of A given by (27), we give the next

Proposition 3.9. *Suppose $A = A_0 + \mu A_1$ where $\{A_1\}_{ij}$ are generic functions of μ satisfying (28). There exist (at least) ∞^4 non-degenerate windows satisfying Condition 3.8. More precisely, let $b_{1,..,4} \in \mathbb{R} \setminus \{0\}$ be such parameters, the simplest[21] representative matrix takes the form*

$$B = \begin{pmatrix} b_1 & 0 & b_3 & 0 \\ 0 & b_2 & 0 & b_4 \\ 0 & 0 & \mu\Sigma_{33}b_3 & \mu\Sigma_{34}b_4 \\ 0 & 0 & \mu\Sigma_{43}b_3 & \mu\Sigma_{44}b_4 \end{pmatrix}, \tag{36}$$

such that $\det B \neq 0$. Σ_{kl} are suitable functions of a_{ij}.

---

[20] This item is an easier to handle condition than "the verticals of $\underline{\Psi} \circ \underline{B}(\underline{x})$ are parallel to $T_{\underline{X}_k'} W_{k+1}^s$", but the geometrical meaning is the same. In this way, since $T_{\underline{X}_k} W_k^u$ and $T_{\underline{X}_k} W_{k+1}^s$ are transversal by virtue of the splitting, horizontals and verticals of the window are transversal too, implying the non-degeneration property.

[21] i.e. with the lowest number of non-vanishing entries.

This structure allows us to simplify as much as possible the computation of the intermediary matrices without any loss of generality.

It will be useful to isolate a scaling factor in the parameters b_i. For this purpose we redefine

$$b_i =: \mu^p b_i. \tag{37}$$

Hence, b_i will be $O(1)$ constants from now on and $p \in \mathbb{N}$ is a control parameter of the window size. As in this circumstance, the particular structure of the perturbed part of A given in (27) does not lead to substantial simplifications, so we prefer to consider for the moment a generic structure. In any case the condition $D \neq 0$ points out a clear relation between splitting and Condition 3.8.

Proof. Consider a generic $B = \{b_{ij}\}$, the above mentioned tangent spaces can be written as

$$T_{\underline{X}_k} W_k^u = \{P = 0, \quad \rho = \rho_k\}, \qquad T_{\underline{X}_k'} W_{k+1}^s = \{Q = 0, \quad \rho = \rho_{k+1}\}.$$

Let us denote by $\underline{u}_i$ the canonical basis vectors of $\mathcal{Q}$ and with $\underline{e}_i$ those of $\mathbb{R}^4$. Then, Condition 3.8 and the non-degeneracy take the form

1. $B\underline{u}_i \cdot \underline{e}_j = 0 \qquad i = 1, 2 \quad j = 3, 4,$

2. $AB\underline{u}_i \cdot \underline{e}_j = 0 \qquad i = 3, 4 \quad j = 1, 4,$

3. B invertible.

The first condition is satisfied if $b_{31} = b_{41} = b_{32} = b_{42} = 0$, while the second one leads to the following linear systems

$$\begin{cases} f_1(\mu)b_{33} + \mu a_{14}b_{43} &= \mu d_1 \\ \mu a_{43}b_{33} + f_2(\mu)b_{43} &= \mu d_3 \end{cases} \quad ; \quad \begin{cases} f_1(\mu)b_{34} + \mu a_{14}b_{44} &= \mu d_2 \\ \mu a_{43}b_{34} + f_2(\mu)b_{44} &= \mu d_4 \end{cases},$$

where

$$\begin{array}{rclcrcl} & & & & d_1 &:=& -(a_{11}b_{13} + a_{12}b_{23}) \\ f_1(\mu) &:=& (a + \mu a_{13}) & & d_3 &:=& -(a_{41}b_{13} + a_{42}b_{23}) \\ f_2(\mu) &:=& 1 + \mu a_{44} & & d_2 &:=& -(a_{11}b_{14} + a_{12}b_{24}) \\ & & & & d_4 &:=& -(a_{41}b_{14} + a_{42}b_{24}) \end{array} \tag{38}$$

By setting $\delta(\mu) = f_1(\mu)f_2(\mu) - \mu^2 a_{14}a_{43} = a + O(\mu)$ that is non-zero for sufficiently small μ, the above systems give

$$\begin{array}{rclcrcl} b_{33} &=& \mu(d_1 f_2(\mu) - \mu a_{14}d_3)/\delta(\mu) & , & b_{34} &=& \mu(d_2 f_2(\mu) - \mu a_{14}d_4)/\delta(\mu) \\ b_{43} &=& \mu(d_3 f_1(\mu) - \mu a_{43}d_1)/\delta(\mu) & & b_{44} &=& \mu(d_4 f_1(\mu) - \mu a_{43}d_2)/\delta(\mu) \end{array} \tag{39}$$

Note that, by the previous solution and by definition of $\delta(\mu)$, we get by (39) then by (38)

$$b_{33}b_{44} - b_{43}b_{34} = \mu^2(d_1 d_4 - d_2 d_3)/\delta(\mu) = \mu^2 D(b_{13}b_{24} - b_{23}b_{14})/\delta(\mu). \tag{40}$$

It is easy to check that the simplest form of B is then achievable by choosing

$$b_{23} = b_{14} = b_{21} = b_{12} = 0, \tag{41}$$

in such a way

$$\det B = \mu^2 b_{11}b_{22}b_{13}b_{24}D/\delta(\mu). \tag{42}$$

Now define $b_1 := b_{11}$, $b_2 := b_{22}$, $b_3 := b_{13}$ and $b_4 := b_{24}$. By hypothesis $D \neq 0$, so B is invertible. In conclusion, by substituting the simplified $d_{1,..,4}$ (by choices (41) in (39)) and defining

$$\begin{array}{rclcrcl} \Sigma_{33} &:=& [-a_{11} + \mu(a_{14}a_{41} - a_{11}a_{44})]/\delta(\mu) & , & \Sigma_{43} &:=& [-aa_{41} + \mu(a_{11}a_{43} - a_{13}a_{41})]/\delta(\mu) \\ \Sigma_{34} &:=& [-a_{12} + \mu(a_{14}a_{42} - a_{12}a_{44})]/\delta(\mu) & & \Sigma_{44} &:=& [-aa_{42} + \mu(a_{12}a_{43} - a_{13}a_{42})]/\delta(\mu) \end{array} \tag{43}$$

we get the required form. $\qquad\square$

4 Proof of the theorem

By using the preparatory material of the previous section, we want to prove Theorem 2.8 via WM. The construction is standard: we consider the evolution through $\underline{\Psi}$ of a generic window $\underline{B}_k$ ("close" to the k−th torus) obtaining a window $\underline{W}_1^k$ and the backward evolution through $\underline{f}^{-n}$ of a window $\underline{B}_{k+1}$ (close to the $k+1$−th torus), getting $\underline{W}_2^k$. Now we wonder if centres and representative matrices of $\underline{B}_{k,k+1}$ can be determined in such way $\underline{W}_1^k$ c.a.o. $\underline{W}_2^k$. The windows $\underline{B}_k$ possess, by construction, a "shape" that makes this property easier to obtain. Nevertheless, the absence of an anisochrony term, implies that the windows, if constructed as in [Mar96], are not "deformed" in a suitable way to get the correct alignment. This will be clear by imposing the condition required by Lemma 3.7, that we use here as a guideline for the proof.

4.1 Affine problem

Let us consider a family of affine windows of the form

$$\underline{B}_k(\underline{x}) = \underline{p}_k + B_k\underline{x},$$

defined in the neighbourhood of $\underline{X}_k$

$$\underline{p}_k = (\sigma_k, 0, \delta_k, 0) + \underline{X}_k, \tag{44}$$

where B_k is defined by (36), taking into account the rescaling (37), once $b_{1,\dots,4}^k$ are chosen: the latter, together with σ_k, δ_k, are meant to be determined in a recursive form.

Now consider the actions on $\underline{B}_k$ and $\underline{B}_{k+1}$ of the linearized maps $\hat{\underline{\Psi}}$ and $\hat{\underline{f}}^{-n}$, respectively. Our purpose is to determine the windows centres and their representative matrices in such a way to satisfy the alignment test. We obtain

$$
\begin{array}{llll}
\underline{c}_1 &=& \underline{X}_k' + A(\sigma_k, 0, \delta_k, 0)^T, & \qquad \underline{c}_2 &=& \underline{v}_n + G(n)\underline{p}_{k+1}, \\
W_1 &=& AB_k, & \qquad W_2 &=& G(n)B_{k+1},
\end{array}
\tag{45}
$$

where $G(n) = \mathrm{diag}(L^{-n}, 1, L^n, 1)$ and $\underline{v}_n = (0, -n\nu, 0, 0)$.

Recalling (27) then using (36) and (43) the product AB_k reads

$$
AB_k = \mu^p
\begin{pmatrix}
\mu a_{11}b_1^k & \mu a_{12}b_2^k & 0 & 0 \\
0 & b_2^k & 0 & b_4^k \\
K_0(\mu)b_1^k & \mu a_{32}b_2^k & K_1(\mu)b_3^k & \mu K_2(\mu)b_4^k \\
\mu a_{41}b_1^k & \mu a_{42}b_2^k & 0 & 0
\end{pmatrix},
\tag{46}
$$

where

$$
\begin{array}{lll}
K_0(\mu) &:=& -1/a + \mu a_{31} \\
K_1(\mu) &:=& -[a + \mu(a_{13} - a_{31}a^2) + \mu^2 a(a_{11}a_{33} - a_{13}a_{31})]/[a(a + \mu a_{13})] \,, \\
K_2(\mu) &:=& [a_{32}a - \mu(a_{12}a_{33} - a_{13}a_{32})]/(a + \mu a_{13})
\end{array}
\tag{47}
$$

and the scaling factor μ^p has been defined in (37). On the other hand

$$
W_2 = \mu^p
\begin{pmatrix}
L^{-n}b_1^k & 0 & L^{-n}b_3^k & 0 \\
0 & b_2^k & 0 & b_4^k \\
0 & 0 & \mu L^n\Sigma_{33}b_3^k & \mu L^n\Sigma_{34}b_4^k \\
0 & 0 & \mu\Sigma_{43}b_3^k & \mu\Sigma_{44}b_4^k
\end{pmatrix}.
\tag{48}
$$

4.1.1 Alignment test part one: intermediary matrices

By using the obtained simplifications, the intermediary matrices read as

$$M_k = \mu^p \begin{pmatrix} \mu a_{11} b_1 & \mu a_{12} b_2 & -c_3 L^{-n} & 0 \\ 0 & b_2 & 0 & -c_4 \\ K_0(\mu) b_1 & \mu a_{32} b_2 & -\mu L^n \Sigma_{33} c_3 & -\mu L^n \Sigma_{34} c_4 \\ \mu a_{41} b_1 & \mu a_{42} b_2 & -\mu \Sigma_{43} c_3 & -\mu \Sigma_{44} c_4 \end{pmatrix},$$

and

$$N_k = \mu^p \begin{pmatrix} -c_1 L^{-n} & 0 & 0 & 0 \\ 0 & -c_2 & 0 & b_4 \\ 0 & 0 & K_1(\mu) b_3 & K_2(\mu) b_4 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where

$$b_i := b_i^k, \qquad c_i := b_i^{k+1}, \tag{49}$$

so we have denoted by c_i the entries of B_{k+1}.

As required by Lemma 3.7 we have to compute the inverse of M_k, however its explicit expression is quite complicated and cumbersome. Rather, we prefer a "perturbative" approach for this computation, by setting $L^n = \mu^{-\beta}$ and suitably choosing β. This leads to the following

Lemma 4.1. *For all $p \in \mathbb{N}$, there exists $\beta_p \in \mathbb{N}$ such that, for all $\beta \geq \beta_p$, the matrix M is invertible and its inverse admits the following expression*

$$M_k^{-1} = \frac{\mu^{-p-1}}{2D\Sigma_{33}} \begin{pmatrix} -\frac{\tilde{D} - a_{42}\Sigma_{33}}{b_1} & \mu \frac{a_{12}\tilde{D}}{b_1} & 0 & -\frac{a_{12}\Sigma_{33}}{b_1} \\ -\frac{a_{41}\Sigma_{33}}{b_2} & -\mu \frac{a_{11}\tilde{D}}{b_2} & 0 & \frac{a_{11}\Sigma_{33}}{b_2} \\ \frac{a_{41}\Sigma_{34}}{c_3} & \mu \frac{\Sigma_{34}D}{c_3} & 0 & \frac{a_{11}\Sigma_{34}}{c_3} \\ -\frac{a_{41}\Sigma_{33}}{c_4} & -\mu \frac{\Sigma_{33}D}{c_4} & 0 & \frac{a_{11}\Sigma_{33}}{c_4} \end{pmatrix} + O(\mu^{\beta-p-1}), \tag{50}$$

where $\tilde{D} := \Sigma_{33}\Sigma_{44} - \Sigma_{34}\Sigma_{43} = D/\delta(\mu)$. Moreover, the explicit expression of the third column is

$$(M_k^{-1})_3 = \frac{\mu^{\beta-p-1}}{2D\Sigma_{33}} \begin{pmatrix} -\frac{a_{12}\Sigma_{43}}{b_1} \\ \frac{a_{11}\Sigma_{43}}{b_2} \\ \frac{D - a_{11}\Sigma_{44}}{c_3} \\ \frac{a_{11}\Sigma_{43}}{c_4} \end{pmatrix} + O(\mu^{2\beta-p-2}). \tag{51}$$

As β does not depend on μ this result holds at the price of a multiplicative $O(1)$ constant of the transition time $n(\beta) := \log_L \mu^{-\beta} = O(\log \mu^{-1})$.

Proof. It is a straightforward check. By choosing

$$\beta_b := 4(p+1) \tag{52}$$

one gets, for sufficiently small μ and $\beta \geq \beta_p$,

$$\det M_k \sim -2\mu^{4p+3-\beta} b_1 b_2 c_3 c_4 D\Sigma_{33}. \tag{53}$$

which is non-zero by (28). The adjugate can be computed in the same way, by using the dominance of terms containing μ^β. $\square$

By using the already obtained form of M_k^{-1} and by defining

$$\gamma := a_{12}/a_{11}, \tag{54}$$

the product $M_k^{-1}N_k$ yields, up to $O(\mu^\beta)$

$$M_k^{-1}N_k = \frac{1}{2}\begin{pmatrix} 0 & \gamma\frac{c_2}{b_1} & 0 & -\gamma\frac{b_4}{b_1} \\ 0 & -\frac{c_2}{b_2} & 0 & \frac{b_4}{b_2} \\ 0 & -\gamma\frac{c_2}{c_3} & 0 & \gamma\frac{b_4}{c_3} \\ 0 & \frac{c_2}{c_4} & 0 & -\frac{b_4}{c_4} \end{pmatrix}.$$

Now, recalling (49) and reintroducing the dependencies on k, it is easy to see that if $b_i^k = b_i^{k+1}$ for some k, even in the simplest case $\gamma = 0$, one gets

$$\|M_k^{-1}N_k\|_\infty = \frac{1}{2}\left(1 + \max\left\{\frac{b_4^k}{b_2^k}, \frac{b_2^k}{b_4^k}\right\}\right) \geq 1,$$

and the affine alignment test fails.

The situation is quite different in the anisochronous case, in which the θ component of the map $\hat{\underline{f}}^{-n}$ reads as (see [Mar96])

$$f_2^{-n}(\theta, \rho) = \theta - (\nu + \nu_1\rho)n.$$

In this way, one gets an estimate of the form

$$\|M_k^{-1}N_k\|_\infty \leq \frac{2}{|\alpha\nu_1 n - 2|},$$

(see [Mar96, Pag. 248]), where $\alpha = O(\mu)$ and n is kept sufficiently large.

Due to the presence of the term ν_1, the quantity $\|M_k^{-1}N_k\|_\infty$ tends to zero, as n increases, and the alignment test can be satisfied.

This looks as the consequence of the "transversality-torsion" mechanism pointed out in [CG08]: the joint action of splitting ($\alpha \neq 0$) and torsion ($\nu_1 \neq 0$) transforms the partial hyperbolicity (i.e. in the two variables (P, Q)), in total hyperbolicity (in the four variables). In this way the windows are compressed-stretched also in the (θ, ρ) variables. This feature, already mentioned and used in [Eas81], is clearly stressed in Lemma 4.3 of the paper [Cre03]. For such a result, the eigenvalues of the map $\mathcal{L} : \mathcal{Q} \to \mathcal{Q}$ defined by

$$\mathcal{L} := \underline{B}^{-1} \circ \underline{f}^n \circ \underline{\Phi} \circ \underline{B}$$

are explicitly computed, showing that for sufficiently small μ and large n, these are real and of modulus not equal to one. Consistently if $\alpha\nu_1 = 0$, the modulus of the eigenvalues corresponding to directions (θ, ρ) are equal to one and the phenomenon vanishes.

So the presence of torsion appears to be an advantageous property in the WM that we cannot use. Nevertheless, by a careful choice of the representative matrices, as depending on k, it is possible to obtain a sort of "simulation" of the torsion effect:

Lemma 4.2 (simulated torsion). *Let $N = O(1/\mu)$. There exist an $O(1)$ positive constant K and a sequence $\{B_k\}_{k=1,\ldots,N} \in \mathbb{M}(4, 4)$ of representative matrices of the form (36) such that the resulting[22] $\underline{W}_{1,2}^a$ give rise to intermediary matrices satisfying*

$$\|M_k^{-1}N_k\|_\infty \leq 1 - K\mu^2. \tag{55}$$

for all $k = 1, \ldots, N$.

---

[22] In this computation the centres $\underline{p}_k$ do not appear and then these will be determined later.

As it is evident, this result gives a thin (but necessary) layer in order to satisfy the next estimates.

Proof. Suppose $\gamma \neq 0$, the simpler case $\gamma = 0$ is an easy consequence.

Condition $\|M_k^{-1}N_k\|_\infty < 1$ gives the following system of recursive inequalities

$$\begin{cases} b_2^{k+1} + b_4^k & < & 2b_1^k/|\gamma| \\ b_2^{k+1} + b_4^k & < & 2b_2^k \\ b_2^{k+1} + b_4^k & < & 2b_3^{k+1}/|\gamma| \\ b_2^{k+1} + b_4^k & < & 2b_4^{k+1} \end{cases}, \tag{56}$$

redefining $b_i^j := |b_i^j|$, then the solution makes sense only if $b_i^k > 0$ for all k.

Now note that as neither b_1^{k+1} nor b_3^k appear in the previous system, the choice of b_1^k and of b_3^k is free. Hence, the first and the third equations are trivially satisfied. In this way the previous system is satisfied (by a suitable choice of b_1^k and b_3^k) for all k if the following system

$$\begin{cases} b_2^{k+1} + b_4^k & = & Cb_2^k \\ b_2^{k+1} + b_4^k & = & Cb_4^{k+1} \end{cases} \tag{57}$$

is, with $C < 2$. The previous system yields

$$\begin{cases} b_2^{k+1} & = & Cb_2^k - b_4^k \\ b_4^{k+1} & = & Cb_2^k \end{cases}.$$

Shifting the index by one and setting $x_k := b_2^k$, (57) is equivalent to the following one dimensional discrete initial value problem

$$\begin{cases} x_{k+2} & = & Cx_{k+1} - x_k, & k \geq 1 \\ x_2 & = & b_2^2 \\ x_1 & = & b_2^1 \end{cases}.$$

For fixed N, we are interested in a solution such that $x_k > 0$ for all $k = 1, \ldots, N$.

If $C < 2$ the solution takes the form

$$x_k = K_1 \cos(k-1)\alpha + K_2 \sin(k-1)\alpha, \qquad k \geq 1, \tag{58}$$

where $\alpha = \arctan(C^{-1}\sqrt{4 - C^2})$.

By choosing $x_2 = (C/2)x_1$ we get $K_2 = 0$. Moreover, by setting $C := \sqrt{4 - \mu^2 b^2}$ where b has to be determined, we have, $\alpha = \mu b/2 + O(\mu^3)$. In this way (58) reduces, up to higher orders, to

$$x_k = x_1 \cos\left(k\mu\frac{b}{2}\right) \tag{59}$$

for all $k = 1, \ldots, N$. So $x_k > 0$ for all $k = 1, \ldots, N$ if $k\mu\frac{b}{2} \in (0, \pi/2)$, that is

$$b = \frac{\pi}{\mu(N+1)}.$$

Now come back to the original sequences. Recalling the definition of the sequences b_i^k in terms of x_k we see that the remaining two inequalities of (56) are satisfied by taking for all k

$$b_1^k = \frac{2}{C}\frac{\gamma}{2}(b_2^{k+1} + b_4^k) = \gamma b_2^1, \qquad b_3^{k+1} = b_1^k,$$

19

(we have used that $2/C > 1$) for all k (b_3^1 is arbitrary).

By keeping in mind that $x_1 := b_2^1$ is arbitrary, the remaining initial conditions are determined. The datum b_4^1 can be chosen as $b_4^1 = (C/2)b_2^1$.

Note that $C = 2 - \mu^2 b^2/4 + O(\mu^4)$, so, by defining $K := b^2/4 + O(\mu^2)$, we see that b has to be an $O(1)$ constant, otherwise estimate (55) and approximation (59) do not properly make sense. This implies that $N = O(1/\mu)$, a fully compatible choice with our purposes.

$\square$

Remark 4.3. Note that variable parameters b_2^k and b_4^k are involved in the (θ, ρ) part of the windows, while the entries b_1^k and b_3^k, relative to the hyperbolic part, can be chosen as constant during the evolution. In this sense we simulate the torsion effect.

Remark 4.4. The decay of b_2^k and b_4^k implies that the determinant of B_k, and then those of M_k, tend to zero. Obviously, by hypothesis, zero is never attained but the entries b_2^k and b_4^k reach values which are no longer $O(1)$. The smallest value of these entries (we get it for $k = N$) is immediate by (59)

$$b_2^N = b_2^1 \cos\left(\frac{\pi}{2}\frac{N}{N+1}\right) \sim b_2^1 \frac{\pi}{2N} = O(\mu). \tag{60}$$

In such a way (53) changes as follows

$$\det M_k|_{k=N} \sim -2\mu^{4p+5-\beta} b_1^{N-1} b_2^{N-1} b_3^N b_4^N D\Sigma_{33}. \tag{61}$$

As an uniform estimates is more convenient, we extend condition (52) to

$$\beta_p := 6 + 4p, \tag{62}$$

for all k.

4.1.2 Alignment test part two: estimate of the term $M_k^{-1}(\underline{c}_2 - \underline{c}_1)$

Our aim is now to determine the centres $\underline{p}_k$ in such a way the alignment test is satisfied for all k.

Set $\underline{\Delta} := (\underline{c}_2 - \underline{c}_1)$. By (55), for a successful test we need, for all k

$$\|M_k^{-1}\underline{\Delta}\|_\infty \leq \frac{1}{2}K\mu^2. \tag{63}$$

This means that the centres of windows $\underline{W}_1^a$ and $\underline{W}_2^a$ should coincide "as much as possible". For this purpose we shall use the EEC mechanism in a very advantageous way.

Lemma 4.5. *Let us consider the sequence B_k constructed before. For all ESC there exist an EEC, three sequences $(\sigma_k, \delta_k) \in \mathbb{R}^2$ and $n_{k+1} \in \mathbb{R}$ in such a way the windows $\underline{W}_{1,2}^a$ (obtained by $\underline{B}_k$ and now completed by (44)) are correctly aligned for all k.*
Furthermore, for all $k = 1, \ldots, N-1$, the transition time from $\mathcal{T}_k$ to $\mathcal{T}_{k+1}$ satisfies

$$n_{k+1} = O\left(\ln\frac{1}{\mu}\right). \tag{64}$$

The previous formula, as $N = O(\mu^{-1})$, gives immediately the estimate of the transition time (11).

Proof. We need to choose the parameters δ_k and σ_k in order to satisfy (63) for all k. It will be simpler to keep δ_{k+1} and σ_{k+1} fixed, and then determine δ_k, σ_k.

Keeping in mind (45), the two windows centres take the following form

$$
\begin{aligned}
\underline{c}_1 &= (\mu\sigma_k a_{11} + f_1(\mu)\delta_k, \theta'_k, P'_k + \mu a_{33}\delta_k + K_0(\mu)\sigma_k, \rho_{k+1} + \mu a_{41}\sigma_k + \mu a_{43}\delta_k) \\
\underline{c}_2 &= (\mu^\beta(Q_{k+1} + \sigma_{k+1}), \theta_{k+1} - n\nu, \mu^{-\beta}\delta_{k+1}, \rho_{k+1})
\end{aligned}
$$

In order to satisfy (63), we try to nullify as many components of $\underline{\Delta}$ as possible. This is achievable with the first and fourth component via a suitable choice of δ_k and σ_k. More precisely, if

- $a_{41} \neq 0$, we choose $\sigma_k = -\delta_k a_{43}/a_{41}$ and then $\delta_k = \frac{\mu^\beta(Q_{k+1}+\sigma_{k+1})a_{41}}{f_1(\mu)a_{41}-\mu a_{11}a_{43}}$.

- $a_{41} = 0$, we take $\delta_k = 0$ and then $\sigma_k = \mu^{\beta-1}(Q_{k+1} + \sigma_{k+1})/a_{11}$.

The described property holds for all $k = N-1, \ldots, 1$. As free parameters, we choose $\delta_N = \sigma_N = 0$. Note that δ_k can be either zero or $\mathrm{O}(\mu^\beta)$. This implies that

$$
\Delta_3 = O(1).
$$

Moreover, if β is sufficiently large, the third column of M_k^{-1} is arbitrarily small and the contribution of the product $M_k^{-1}\Delta_3$ is not greater than $O(\mu^{3(p+1)})$ and then negligible for a suitable p.

Now we are going to use the elasticity of the transition chain in order to nullify Δ_2. For this purpose, recall the notational setting of section 3.3 and consider values y_{k+1} and y_k of a given ESC. By keeping $\tilde{y}_{k+1}$ fixed we want to move y_k to a suitable $\tilde{y}_k \in E_k$.

Note that for all $y \in E_k$, the argument we have used to get $\Delta_1 = \Delta_4 = 0$ can be repeated simply by replacing the coordinates of $\underline{X}_k, \underline{X}'_k$ with those of $\underline{X}_k(y), \underline{X}'_k(y)$. This is possible as the elements of the family $k+1$ do not change and the algorithm remains well posed. Now δ_k, σ_k have the functional form $\delta_k(y), \sigma_k(y)$.

Define $2e_k := \operatorname{diam} E_k$ and

$$
\begin{aligned}
\mathcal{O}_k^- &:= \{\theta'_k(y): \quad y \in (y_k - e_k, y_k]\} \\
\mathcal{O}_k^+ &:= \{\theta'_k(y): \quad y \in [y_k, y_k + e_k)\}
\end{aligned}
$$

In the time $n(\beta)$ the section Σ_N is visited by points whose reciprocal distance is not greater than $O((n(\beta))^{-\frac{1}{\tau}})$ (see e.g. [Gal99] and references therein). As $\operatorname{diam}\mathcal{O}_k^\pm = O(1)$ there exist subsequences $n_\pm^m$ with $n_\pm^1$ of the same order of $n(\beta)$, in a way that

$$
\theta_{k+1} - n_\pm^m \nu \in \mathcal{O}_k^\pm.
$$

Suppose that $n_{k+1} := n_+^{\bar{m}}$ for some chosen $\bar{m}$ and define

$$
\mathcal{F}(y) := [\theta'_k(y) - (\theta_{k+1} - n_{k+1}\nu)] \bmod 2\pi \equiv \Delta_2. \tag{65}
$$

From (9) we deduce that $\mathcal{F}(y)$ is continuous and $\frac{d}{dy}\mathcal{F}(y) = O(\mu^{-1})$ on E_k, moreover $\mathcal{F}(y_k)\mathcal{F}(y_k + e_k) < 0$, so there exists a unique[23] $\tilde{y}_k \in E_k$ such that $\mathcal{F}(y_k) = 0$.

In such a way $\|M_k^{-1}\underline{\Delta}\|_\infty \sim 0$ and (63) is trivially satisfied. $\qquad\square$

Remark 4.6. At first, the elasticity of the transition chain seems to be as an unnecessary tool in order to achieve affine alignment. Without using it, we would have obtained

$$
\|M_k^{-1}\underline{\Delta}\|_\infty \leq \mu^{-p}J_k\Delta_2, \tag{66}
$$

---

[23] Note that this point can be constructively determined with an arbitrary precision.

where $J_k = \max\{\gamma/b_1^k, 1/b_2^k, \gamma/b_3^{k+1}, 1/b_4^{k+1}\}$. Clearly, due to the decreasing of the sequences $b_{2,4}^k$, the greatest value of J_k is $J_N = O(\mu^{-1})$. In order to satisfy (63), Δ_2 should be of an order not greater than μ^{3+p}, requiring in this way the following ergodization time

$$n = O(\mu^{-(p+3)}).$$

As it will be clear by looking at the estimates of the next section, during such (long) time the remainder of $\underline{f}^{-n}$ is not suitably bounded and the affine problem is not still a good approximation of the complete one.

Remark 4.7. The use of the EEC allows us to bypass the ergodization time on Σ_N so the drift time is independent on the Diophantine constants.

4.2 Estimate of the remainders

Recall Lemma 3.7. By construction, the functions $\hat{W}_1(\underline{x})$ and $\hat{W}_2(\underline{x})$ are given by the remainders (19) and (26) respectively, evaluated at $\underline{\xi} \leftarrow (\sigma_k, 0, \delta_k, 0) + B_k \underline{x}$. We denote these functions with $\xi(\underline{x})$, $\underline{x} \in (-2, 2)^4$.
Furthermore, set $\gamma_k := \|M_k^{-1}\|_\infty$ and define

$$
\begin{aligned}
\chi_1 &:= \sup_{\underline{x}\in\mathcal{U}}[\|M_k^{-1}\underline{R}_{k+1}(\underline{x})\|_\infty + \gamma_0\|\hat{\underline{R}}_k(\underline{x})\|_\infty], \\
\chi_2 &:= \sup_{\underline{x}\in\mathcal{U}}[\|M_k^{-1}\partial_{\underline{x}}\underline{R}_{k+1}(\underline{x})\|_\infty + \gamma_0\|\partial_{\underline{x}}\hat{\underline{R}}_k(\underline{x})\|_\infty].
\end{aligned}
\tag{67}
$$

In this way

$$\chi^c \leq \max\{\chi_1, \chi_2\}.$$

First of all, it is easy to see from (26) that

$$\|\hat{\underline{R}}_k(\underline{x})\|_\infty \leq C_2\|\xi(\underline{x})\|_\infty^2, \qquad \|\partial_{\underline{x}}\hat{\underline{R}}_k(\underline{x})\|_\infty \leq C_3\|\xi(\underline{x})\|_\infty\|\partial_{\underline{x}}\xi(\underline{x})\|_\infty,$$

where $C_{2,3} < +\infty$ by regular (analytic) dependence on initial data over finite time. Moreover, by a comparison between (50) and (60), we have $\max_k \gamma_k = O(\mu^{-1-p})$. So we get

$$\max_k \gamma_k\|\hat{\underline{R}}_k(\underline{x})\|_\infty \sim \max_k \gamma_k\|\partial_{\underline{x}}\hat{\underline{R}}_k(\underline{x})\|_\infty = O(\mu^{p-2}),$$

uniformly in $\underline{x}$.
By inspecting formula (19), we obtain

$$
\begin{aligned}
\max_k\|M_k^{-1}\underline{R}(\underline{x})\|_\infty &\sim \mu^{-1-p}\left\|\begin{pmatrix} m_{11} & \cdots & m_{13}\mu^{\beta-1} & \cdots \\ m_{21}\mu^{-1} & \cdots & m_{23}\mu^{\beta-2} & \cdots \\ m_{31} & \cdots & m_{33}\mu^{\beta-1} & \cdots \\ m_{41}\mu^{-1} & \cdots & m_{43}\mu^{\beta-2} & \cdots \end{pmatrix}\begin{pmatrix} n\mu^{\beta-p} \\ 0 \\ n\mu^{-\beta+2p} \\ 0 \end{pmatrix}\right\|_\infty \\
&= O(n_{k+1}\mu^{p-3}),
\end{aligned}
$$

recalling (62) and (64). Similarly

$$\max_k\|M_k^{-1}\partial_{\underline{x}}\underline{R}_k(\underline{x})\|_\infty = O(n_{k+1}\mu^{p-3}).$$

Keeping in mind (55), in order to satisfy the second of (35), and then the first one, it is sufficient to choose $p = 6$ (from which $\beta \geq 30$). As $\lim_{\mu\to 0} n_{k+1}\mu = 0$ by definition, we have $\chi_{1,2} = o(\mu^2)$ and the proof is complete.

22

References

[Arn64]   V.I. Arnold. Instability in dynamical systems with several degrees of freedom. *Soviet Math. Dokl.*, 1964.

[BB02]   M. Berti and P. Bolle. A functional analysis approach to Arnold diffusion. *Annales de l'I.H.S. Analyse Non Linèaire*, 19:395–450, 2002.

[BBB03]   M. Berti, L. Biasco, and P. Bolle. Drift in phase space: a new variational mechanism with optimal diffusion time. *J. Math. Pures Appl.*, 82:613–664, 2003.

[Bes96]   U. Bessi. An approach to Arnold's diffusion through the calculus of variations. *Nonlin. An.*, 26(6):1115–1135, 1996.

[CG94]   L. Chierchia and G. Gallavotti. Drift and diffusion in phase space. *Annales de l'I.H.P.*, 60(1):1–144, 1994. section A.

[CG03]   J. Cresson and C. Guillet. Periodic orbits and Arnold diffusion. *Disc. and Cont. Dyn. Syst.*, 9(2):451–470, 2003.

[CG08]   J. Cresson and C. Guillet. Hyperbolicity versus partial-hyperbolicity and the transversality-torsion phenomenon. *Journal of Differential Equations*, 244(9):2123 – 2132, 2008.

[Cre01]   J. Cresson. Temps d'instabilité des systèmes hamiltoniens initialement hyperboliques. *C. R. Acad. Sci. Paris*, (332):831–834, 2001. Sèrie I, Systémes dynamiques.

[Cre03]   J. Cresson. Symbolic dynamics and Arnold diffusion. *J. Differential Equations*, (187):269–292, 2003.

[Eas81]   R.W. Easton. *Classical Mechanics and Dynamical Systems*, chapter Orbit structure near trajectories biasymptotic to invariant tori., pages 55–67. Marcel Dekker, New York, Devaney, R. and Nitecki, Z. edition, 1981.

[EM79]   R.W. Easton and R. McGehee. Homoclinic phenomena for the orbits double asymptotic to an invariant three sphere. *Indiana Univ. Math. J.*, 28:211–240, 1979.

[For12]   A. Fortunati. *Heteroclinic intersections and Arnold's diffusion: comparison and analysis of two methods.* Tesi di Dottorato, University of L'Aquila, 2012.

[Gal94]   G. Gallavotti. Twistless kam tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable hamiltonian systems: A review. *Reviews in Mathematical Physics*, 06(03):343–411, 1994.

[Gal97]   G. Gallavotti. Hamiltonian - Jacobi's equation and Arnold's diffusion near invariant tori in a priori unstable isochronous systems. *Rendiconti del seminario matematico di Torino*, 55(4), 1997.

[Gal99]   G. Gallavotti. Arnold's diffusion in isochronous systems. *Mathematical Physics, Analysis and Geometry*, 1:295–312, 1999.

[GGM99]   G. Gallavotti, G. Gentile, and V. Mastropietro. Mel'nikov's approximation dominance: Some examples. *Reviews in Mathematical Physics*, 11(04):451–461, 1999.

[GR07]   M. Gidea and C. Robinson. Shadowing orbits for transition chains of invariant tori alternating with Birkhoff zones of instability. *Nonlinearity*, 5(20):1115–1143, 2007.

[GZ04]   M. Gidea and P. Zgliczyński. Covering relations for multidimensional dynamical systems I and II. *J. Diff. Eq.*, (1):32–80, 2004.

[LM05]   P. Lochak and J.P. Marco. Diffusion times and stability exponents for nearly integrable analytic systems. *Central European Journal of Mathematics*, 3(3):342–397, 2005.

[LMS03]   P. Lochak, J.P. Marco, and D. Sauzin. *On the Splitting of Invariant Manifolds in Multidimensional Near-integrable Hamiltonian Systems*. Number no. 775 in Memoirs of the American Mathematical Society. American Mathematical Soc., 2003.

[Mar96]   J.P. Marco. Transition le long des chaines de tores invariants pour les systèmes hamiltoniens analyitiques. *Annales de l'I.H.P.*, 64(2):205–252, 1996.

[Sab15]   L. Sabbagh. An inclination lemma for normally hyperbolic manifolds with an application to diffusion. *Ergodic Theory Dynam. Systems*, 35(7):2269–2291, 2015.

[Tre02]   D. Treshev. Trajectories in a neighbourhood of asymptotic surfaces of a priori unstable Hamiltonian systems. *Nonlinearity*, 15:1–18, 2002.

[Val98]   E. Valdinoci. Tori di transizione nella teoria KAM. Tesi di Laurea, Università di Roma 3, 1998.